



# Emotions without language. The case of infants and animals

Marie Michon

## ► To cite this version:

Marie Michon. Emotions without language. The case of infants and animals. pp.48-52, 2018, 978-84-09-06054-2. <hal-04010161>

**HAL Id: hal-04010161**

**<https://hal.science/hal-04010161v1>**

Submitted on 14 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

**Actas del IX Congreso de la Sociedad de  
Lógica, Metodología y Filosofía de la Ciencia en  
España**

*Madrid, 13–16 de noviembre de 2018*



**Proceedings of the IX Conference of the  
Spanish Society of Logic, Methodology and  
Philosophy of Science**

*Madrid, 13–16 November 2018*



**Editado por:**

Cristian Saborido (UNED)

Sergi Oms (U. de Barcelona)

Javier González de Prado (UNED / U. Nova de Lisboa)

Actas del IX Congreso de la Sociedad de  
Lógica, Metodología y Filosofía de la  
Ciencia en España

*Madrid, 13–16 de noviembre de 2018*

---

Proceedings of the IX Conference of the  
Spanish Society of Logic, Methodology  
and Philosophy of Science

*Madrid, 13–16 November 2018*

EDITADO POR:

Cristian Saborido (UNED)

Sergi Oms (U. de Barcelona)

Javier González de Prado (UNED / U. Nova de Lisboa)

**Maquetación:** José Ángel Gascón

**Imagen de portada:** Cristina Zafra

**ISBN:** 978-84-09-06054-2

**Comité organizador**

Cristian Saborido (Chair, UNED)  
Giorgio Airoidi (UNED)  
Clara Bueno (UNED)  
Emilio Cáceres (UNED)  
José Ángel Gascón (UCSC, Chile)  
Javier Gonzalez de Prado (UNED /  
U. Nova de Lisboa)  
María Jimenez Buedo (UNED)  
Lorena Lobo (U. Isabel I)  
Victor Luque (U. de Valencia)  
Laura Menatti (U. Bordeaux)  
Vanessa Triviño (U. Rey Juan Carlos)  
Sergi Oms (U. de Barcelona)  
Cristina Villegas (U. Complutense de  
Madrid)

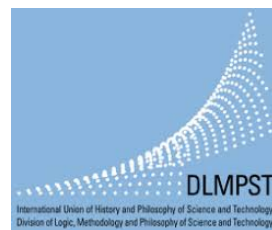
**Comité del programa**

Concepción Martínez Vidal (U. de  
Santiago de Compostela)  
Antonio Blanco (U. Complutense de  
Madrid)  
María Caamaño (U. de Valladolid)  
María Cerezo (U. de Murcia)  
José Martínez (U. de Barcelona)  
Fernando Soler (U. Sevilla)  
Inmaculada Perdomo (U. de la  
Laguna)

**Instituciones patrocinadoras:**



Sociedad de Lógica,  
Metodología y Filosofía  
de la Ciencia en España



# Índice general

|                                                                                                                                                           |          |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|----------|
| <b>Prefacio</b>                                                                                                                                           | XII      |
| <b>Premio para Jóvenes Investigadores</b>                                                                                                                 | XIV      |
| <b>Lullius Lectures 2018: Nancy Cartwright</b><br><i>Mid-level theory: Without it what could anyone do?</i>                                               | 1        |
| <b>Lógica, Historia y Filosofía de la Lógica, y Argumen-<br/>tación</b>                                                                                   | <b>2</b> |
| 1 <b>Baptista Santos, Gonçalo</b><br><i>An intuitionistic understanding of unrestricted quantifiers</i>                                                   | 3        |
| 2 <b>Barés Gómez, Cristina y Fontaine, Matthieu</b><br><i>Razonamiento experimental e inferencia en la práctica médica</i>                                | 6        |
| 3 <b>Berkovski, Sandy</b><br><i>Imperatives and to-do lists: three objections</i>                                                                         | 9        |
| 4 <b>Bertran San Millán, Joan</b><br><i>Frege's adoption of the distinction between concept and object</i>                                                | 16       |
| 5 <b>Carrara, Massimiliano y Lando, Giorgio</b><br><i>Contingent composition as identity</i>                                                              | 19       |
| 6 <b>Domínguez Armas, Álvaro</b><br><i>La argumentación en torno a la desobediencia civil</i>                                                             | 23       |
| 7 <b>Fernandes, Diego</b><br><i>Expressiveness comparisons between logics: an investigation of some intuitions and formal criteria</i>                    | 26       |
| 8 <b>Grimau, Berta</b><br><i>Cat, cats, catses and beyond: In defence of higher-level plural logic</i>                                                    | 33       |
| 9 <b>Herrera González, José Rafael</b><br><i>Las lógicas híbridas como herramienta formal para el desarrollo de sistemas lógicos temporal-epistémicos</i> | 35       |

|    |                                                                                                                                                                            |           |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 10 | <b>Martin, Ben</b><br><i>Identifying logical evidence</i>                                                                                                                  | 42        |
| 11 | <b>Martínez Fabregat, Silvia</b><br><i>Instagram y la pereza argumentativa</i>                                                                                             | 45        |
| 12 | <b>Michon, Marie</b><br><i>Emotions without language. The case of infants and animals</i>                                                                                  | 48        |
| 13 | <b>Moreno Gómez, Manuel Crescencio</b><br><i>Existence as a property of intensions</i>                                                                                     | 53        |
| 14 | <b>Moreno Zurita, Alba</b><br><i>Expressivism: the Frege-Geach problems and new goals</i>                                                                                  | 56        |
| 15 | <b>Orellana Viñambres, David</b><br><i>La lógica del siglo XIX desde la perspectiva de Edmund Husserl y su fenomenología</i>                                               | 59        |
| 16 | <b>Pascual, Carmen</b><br><i>Déficit de conocimientos médicos y no fallos en las estrategias de razonamiento como principales responsables de los errores diagnósticos</i> | 63        |
| 17 | <b>Sánchez Ovcharov, Carmen y Gherab Martín, Karim</b><br><i>Lógica y razonamiento no lingüísticos: el ejemplo de los gráficos existenciales de Ch. S. Peirce</i>          | 66        |
| 18 | <b>Sforza Fogliani, Maria Paola</b><br><i>Revising logic: Anti-exceptionalism and circularity</i>                                                                          | 72        |
| 19 | <b>Viejo, José Manuel</b><br><i>Referencia directa, inocencia semántica y opacidad</i>                                                                                     | 75        |
| 20 | <b>Wójtowicz, Anna y Wójtowicz, Krzysztof</b><br><i>A stochastic process explanation of conditionals</i>                                                                   | 83        |
|    | <b>Filosofía del Lenguaje</b>                                                                                                                                              | <b>87</b> |
| 21 | <b>Agolli, Eno</b><br><i>The Russell, the John, and the Katherine: The-defenders of the-Predicativism</i>                                                                  | 88        |
| 22 | <b>Alhambra Delgado, José</b><br><i>Actos de habla y estatus social. Cómo el lenguaje puede convertirse en un estigma</i>                                                  | 90        |
| 23 | <b>Almagro, Manuel y Torices, José R.</b><br><i>The nature of (covert) dogwhistles</i>                                                                                     | 93        |

|    |                                                                                                                                                                                                   |     |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 24 | <b>Bordonaba, David y Villanueva, Neftalí</b><br><i>Crossed disagreements: A quantitative and qualitative study on the minutes of the sessions of the Spanish Parliament</i>                      | 101 |
| 25 | <b>Campdelacreu, Marta</b><br><i>The difficulty of finding a commonsense answer to the special composition question</i>                                                                           | 109 |
| 26 | <b>Chilovi, Samuele y Wodak, Daniel</b><br><i>The is-ought gap and the grounds of normative facts</i>                                                                                             | 112 |
| 27 | <b>Cortés-Aravena, Juan Esteban</b><br><i>Human deductive reasoning: Conditional perfection and mental models theory</i>                                                                          | 114 |
| 28 | <b>de Pinedo-García, Manuel y Villanueva-Fernández, Neftalí</b><br><i>Power to the people. The indispensable nature of the normative vocabulary and the political turn in analytic philosophy</i> | 122 |
| 29 | <b>Fairhurst, Jordi</b><br><i>Slurs and the semantic-pragmatic divide</i>                                                                                                                         | 125 |
| 30 | <b>Fernández, Ainhoa</b><br><i>Act-based propositions and a posteriori necessity: Linguistic meaning for modes of presentation</i>                                                                | 130 |
| 31 | <b>Fernández Moreno, Luis</b><br><i>Causal theory and reference borrowing</i>                                                                                                                     | 138 |
| 32 | <b>Lanius, David</b><br><i>The value of vagueness</i>                                                                                                                                             | 144 |
| 33 | <b>Macià, Josep</b><br><i>A defense of the common ground view about assertion</i>                                                                                                                 | 148 |
| 34 | <b>Navarro-Laespada, Llanos y Frápolli, María José</b><br><i>Inferentialism, representationalism and moral responsibility</i>                                                                     | 151 |
| 35 | <b>Polo, Nuria</b><br><i>La representación mental del sonido: ¿formalismo lingüístico o realidad cognitiva?</i>                                                                                   | 154 |
| 36 | <b>Rivas Monroy, M<sup>a</sup> Uxía</b><br><i>“La aproximación fácil” a la ontología de Amie L. Thomasson: pragmática y pragmatismo entrelazados</i>                                              | 157 |
| 37 | <b>Soria Ruiz, Andrés</b><br><i>The dynamics of value ascription</i>                                                                                                                              | 163 |
| 38 | <b>Vilanova Arias, Javier</b><br><i>El sentido común y la filosofía del lenguaje informal</i>                                                                                                     | 165 |

|                                                                                                                                                                 |            |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>Filosofía de la Mente y Epistemología</b>                                                                                                                    | <b>171</b> |
| 39 <b>Andrada de Gregorio, Gloria</b><br><i>The extension of cognitive abilities</i>                                                                            | 172        |
| 40 <b>Barba Magdalena, Moisés</b><br><i>El fiabilismo y la justificación de las creencias colectivas</i>                                                        | 175        |
| 41 <b>Blanco, Julia</b><br><i>The bounds of agency: a review of the group person possibility</i>                                                                | 181        |
| 42 <b>Cabrera Miquel, Marta</b><br><i>The role of emotions in philosophical explanations of intentional action</i>                                              | 183        |
| 43 <b>de Donato Rodríguez, Xavier y Falguera, José L.</b><br><i>A neo-Carnapian ontological commitment for science</i>                                          | 185        |
| 44 <b>Duarte Calvo, Antonio</b><br><i>Musement o la actividad de la red neural por defecto</i>                                                                  | 192        |
| 45 <b>Espinosa López, Jorge Alejandro</b><br><i>Hacia una noción de verdad compatible con un pluralismo genuinamente pragmatista o pluralismo radical</i>       | 199        |
| 46 <b>Fernández Prat, Olga y Quesada, Daniel</b><br><i>The experience and processing of temporal features of perception</i>                                     | 207        |
| 47 <b>Filomeno, Aldo</b><br><i>Qué es racional creer sobre las leyes Humeanas a la luz de la objeción de la coincidencia cósmica</i>                            | 214        |
| 48 <b>Garrido Rodríguez, Natividad y Liz Gutiérrez, Manuel</b><br><i>Creación literaria y construcción de espacios de perspectivas</i>                          | 223        |
| 49 <b>Hernández-Conde, José V.</b><br><i>Cuando los conceptos son no-persistentes y no-representacionales: Un examen contextualista-consistente de la mente</i> | 226        |
| 50 <b>Ibarrondo Murguialday, Julen</b><br><i>Ser racional no equivale a responder correctamente a razones: el caso de las intenciones contradictorias</i>       | 230        |
| 51 <b>Liz, Manuel y Vázquez, Margarita</b><br><i>Razonamientos que involucran perspectivas</i>                                                                  | 236        |
| 52 <b>Morales Carbonell, Felipe</b><br><i>On the question of the indispensability of modal thought</i>                                                          | 243        |
| 53 <b>Navarro, Jesús</b><br><i>Fake knowledge-how</i>                                                                                                           | 248        |



|    |                                                                                                                                        |            |
|----|----------------------------------------------------------------------------------------------------------------------------------------|------------|
| 54 | <b>Pérez Otero, Manuel</b>                                                                                                             |            |
|    | <i>Relevancia epistémica de las intuiciones</i>                                                                                        | 251        |
| 55 | <b>Pino, Dani</b>                                                                                                                      |            |
|    | <i>Social virtues upside down: Group epistemic agents as sources for individual epistemic virtues</i>                                  | 256        |
| 56 | <b>Seoane Rodríguez, Pablo</b>                                                                                                         |            |
|    | <i>An hybrid ontology for musical works: Between perceptualism and epistemicism on musical understanding</i>                           | 260        |
| 57 | <b>Serna, Edgar</b>                                                                                                                    |            |
|    | <i>La discusión en torno al estatus epistémico de la doctrina filosófica del falibilismo</i>                                           | 266        |
| 58 | <b>Toribio, Josefa</b>                                                                                                                 |            |
|    | <i>Implicit bias, accessibilism and justification</i>                                                                                  | 269        |
| 59 | <b>Verdejo, Víctor M.</b>                                                                                                              |            |
|    | <i>Thinking and experiencing the embodied self</i>                                                                                     | 272        |
|    | <b>Filosofía y Metodología de la Ciencia</b>                                                                                           | <b>275</b> |
| 60 | <b>Akcin, Haktan</b>                                                                                                                   |            |
|    | <i>What is really wrong with ontic structural realism? On the possibility of reading off ontology from current fundamental science</i> | 276        |
| 61 | <b>Anta, Javier</b>                                                                                                                    |            |
|    | <i>The explanatory power of topo-geometric diagrams in modern physics</i>                                                              | 279        |
| 62 | <b>Atencia, Paula</b>                                                                                                                  |            |
|    | <i>Scientific pluralism and scientific realism: Examining problems and reducing gaps</i>                                               | 287        |
| 63 | <b>Belastegui, Javier</b>                                                                                                              |            |
|    | <i>Kinds, joints and trees: The non-hierarchical structure of natural kinds</i>                                                        | 288        |
| 64 | <b>Bich, Leonardo</b>                                                                                                                  |            |
|    | <i>Operational definitions of life</i>                                                                                                 | 290        |
| 65 | <b>Buckareff, Andrei A.; Andrews, Marc y Brennan, Shane</b>                                                                            |            |
|    | <i>Contra static dispositions</i>                                                                                                      | 292        |
| 66 | <b>Cáceres Vázquez, Emilio</b>                                                                                                         |            |
|    | <i>La autoorganización como propiedad de nivel de sus entidades componentes</i>                                                        | 295        |
| 67 | <b>Camprubí, Lino</b>                                                                                                                  |            |
|    | <i>Varieties of pluralism and their ontological implications: The case of earth system sciences</i>                                    | 302        |

|    |                                                                                                                                                      |     |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 68 | <b>Corbalán, M. Inés y Terzian, Giulia</b><br><i>Ockham's razor in a minimalist guise</i>                                                            | 304 |
| 69 | <b>Corral-Villate, Amaia</b><br><i>Analysis of the infinite gods paradox in classical particle mechanics.<br/>Representation of a causal problem</i> | 308 |
| 70 | <b>Dziurosz-Serafinowicz, Patryk</b><br><i>Justifying Lewis's kinematics of chance</i>                                                               | 311 |
| 71 | <b>Findl, Johannes</b><br><i>Knowledge vs understanding: Epistemic scenarios and scientific practice</i>                                             | 314 |
| 72 | <b>García Lapeña, Alfonso</b><br><i>Truthlikeness for quantitative deterministic laws</i>                                                            | 317 |
| 73 | <b>Henda Gómez de Segura, Mikel</b><br><i>La teoría relacional del tiempo como teoría constructiva y su poder<br/>explicativo</i>                    | 325 |
| 74 | <b>Hermida, Margarida</b><br><i>Functional kinds as universal natural kinds in the life sciences</i>                                                 | 330 |
| 75 | <b>Hernández, Abraham</b><br><i>Más allá de Kitcher: la reformulación del contrato social de la ciencia</i>                                          | 334 |
| 76 | <b>Heylen, Jan y Tump, Lars</b><br><i>The enhanced indispensability argument for Platonism and the circu-<br/>larity problem</i>                     | 337 |
| 77 | <b>Krauss, Alexander</b><br><i>How our mind enables and constrains the theories we develop</i>                                                       | 340 |
| 78 | <b>Kvasz, Ladislav</b><br><i>Patterns of theory change in physics</i>                                                                                | 341 |
| 79 | <b>Loaiza, Juan R.</b><br><i>Can functionalism save emotions from elimination?</i>                                                                   | 348 |
| 80 | <b>Luque, Víctor J.</b><br><i>Is there room for a new foundation for the force interpretation?</i>                                                   | 353 |
| 81 | <b>Martínez-Vidal, Concha y Rivas de Castro, Navia</b><br><i>The enhanced indispensability argument. Where are we?</i>                               | 357 |
| 82 | <b>Militello, Guglielmo</b><br><i>Structural and organisational conditions for machine-likeness</i>                                                  | 360 |
| 83 | <b>Montojo, Carlos y García Valdecasas, Miguel</b><br><i>Los mundos posibles de Lewis son entrópicos</i>                                             | 362 |

|    |                                                                                                                                                    |     |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 84 | <b>Ongay de Felipe, Íñigo</b><br><i>The dappled world of evolutionary biology: Scientific pluralism and the extended evolutionary synthesis</i>    | 365 |
| 85 | <b>Pelacho, Maite</b><br><i>Acciones y motivaciones para la constitución de lo común: Una propuesta desde la ciencia ciudadana</i>                 | 367 |
| 86 | <b>Pinheiro, João</b><br><i>The science of ethics &amp; evolutionary moral realism</i>                                                             | 370 |
| 87 | <b>Pogonowski, Jerzy</b><br><i>Extremal axioms in mathematics</i>                                                                                  | 372 |
| 88 | <b>Rivadulla, Andrés</b><br><i>Theoretical innovation. A new look at creativity in the natural sciences</i>                                        | 376 |
| 89 | <b>Sánchez Ovcharov, Carmen</b><br><i>Diagramatización en física: aproximación semiótica al alcance representativo de los diagramas de Feynman</i> | 384 |
| 90 | <b>Sánchez-Dorado, Julia</b><br><i>The creative similarity account of the means of representation</i>                                              | 388 |
| 91 | <b>Sánchez Gómez, Pedro J.</b><br><i>A propensionist interpretation of Lewis structures</i>                                                        | 393 |
| 92 | <b>San Pedro, Iñaki</b><br><i>Computer simulations, emergence, and scientific representation</i>                                                   | 398 |
| 93 | <b>Soler, Jaime</b><br><i>Usefulness (or not) of defining the concept of life in biology</i>                                                       | 401 |
| 94 | <b>Suárez, Javier</b><br><i>Holobionts as units of selection: A trait-based account of holobiont individuality</i>                                 | 405 |
| 95 | <b>Suñé, Abel</b><br><i>An account of fundamentality for the metaphysics of science</i>                                                            | 408 |
| 96 | <b>Urueña López, Sergio</b><br><i>Scientific realism beyond pessimistic induction: From truth to plausibility</i>                                  | 411 |
| 97 | <b>Vecchi, Davide y Casetta, Elena</b><br><i>Probing geno-centric approaches to conservation biology</i>                                           | 420 |
| 98 | <b>Viñuela Villa, Pedro A.</b><br><i>Kant and Wittgenstein on pictures and proofs in mathematics</i>                                               | 423 |
| 99 | <b>Wójtowicz, Krzysztof</b><br><i>Metaabstract explanations in science</i>                                                                         | 427 |

|                                                                                                                                                            |            |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>Ciencia, Tecnología y Sociedad</b>                                                                                                                      | <b>430</b> |
| 100Bernabeu, Cristina<br><i>Predecir personas: la opresión de género en los perfiles los buyers persona</i>                                                | 431        |
| 101de Cózar Escalante, José Manuel<br><i>Antropoceno y praxeología de la cultura científica</i>                                                            | 436        |
| 102García-Barranquero, Pablo<br><i>A new framework for thinking about life extension</i>                                                                   | 441        |
| 103García Díaz, Laura y Hernández Pérez, Abraham<br><i>El papel de las series de televisión en la democratización de la ciencia</i>                        | 445        |
| 104Perdomo, Inmaculada<br><i>Socially responsible philosophy of science</i>                                                                                | 448        |
| 105Pérez Sedeño, Eulalia<br><i>Biotechnología y maternidad por sustitución: ¿cosificación, mercantilización o altruismo de género?</i>                     | 452        |
| 106Rivero, Thais<br><i>Proyecto Isotipo: Hacia un lenguaje visual unificado</i>                                                                            | 458        |
| 107Torres González, Obdulía<br><i>El desequilibrio de género en filosofía</i>                                                                              | 461        |
| 108Vázquez Campos, Margarita y García Díaz, Laura<br><i>Puntos de vista y referencias temporales en una serie de televisión. El caso de "The affair"</i>   | 468        |
| 109Veit, Walter<br><i>Enhancement technologies and inequality</i>                                                                                          | 471        |
| <b>Historia de la Ciencia</b>                                                                                                                              | <b>477</b> |
| 110Labrador Montero, Daniel<br><i>Telos y especie biológica: la historia de dos conceptos inseparables</i>                                                 | 478        |
| <b>Taller sobre Nancy Cartwright</b>                                                                                                                       | <b>486</b> |
| 111Arruda, Renata<br><i>Causal pluralism and the boundaries of manipulation</i>                                                                            | 487        |
| 112Bujalance, Laura<br><i>In the beginning there were no laws of nature and now what is out there? On Nancy Cartwright's "dappled" causal explanations</i> | 490        |

|      |                                                                                                         |            |
|------|---------------------------------------------------------------------------------------------------------|------------|
| 113  | <b>Helding Vembye, Mikkel</b>                                                                           |            |
|      | <i>Reflecting on Nancy Cartwright's conception of external validity of Randomised Controlled Trials</i> | 494        |
| <br> |                                                                                                         |            |
|      | <b>Simposio:</b>                                                                                        |            |
|      | <b>Issues at the intersection between metaphysics and biology</b>                                       | <b>504</b> |
| 114  | <b>Introduction</b>                                                                                     | 505        |
| 115  | <b>Austin, Christopher J.</b>                                                                           |            |
|      | <i>Form, cause, and explanation in biology: A neo-Aristotelian perspective</i>                          | 507        |
| 116  | <b>Bertolaso, Marta y Cerezo, María</b>                                                                 |            |
|      | <i>Development and cancer: Insights from teleology</i>                                                  | 509        |
| 117  | <b>Triviño, Vanesa</b>                                                                                  |            |
|      | <i>Emergentism in a biological framework: The case of fitness</i>                                       | 512        |
| 118  | <b>Villegas, Cristina y Nuño de la Rosa, Laura</b>                                                      |            |
|      | <i>Chances and propensities in evo-devo</i>                                                             | 514        |
| <br> |                                                                                                         |            |
|      | <b>Simposio:</b>                                                                                        |            |
|      | <b>La filosofía de la argumentación</b>                                                                 | <b>517</b> |
| 119  | <b>Alcolea Banegas, Jesús</b>                                                                           |            |
|      | <i>Maniobras estratégicas y estrategias interpretativas en la argumentación</i>                         | 518        |
| 120  | <b>Galindo Castañeda, Joaquín</b>                                                                       |            |
|      | <i>Las preguntas "por qué" argumentativamente</i>                                                       | 525        |
| 121  | <b>Marraud, Hubert</b>                                                                                  |            |
|      | <i>Usos asertivos, directivos y expresivos de argumentación</i>                                         | 528        |
| 122  | <b>Moldovan, Andrei</b>                                                                                 |            |
|      | <i>Presuposiciones, preguntas y falacias</i>                                                            | 535        |
| 123  | <b>Vega Reñón, Luis</b>                                                                                 |            |
|      | <i>Una invitación a tomar con filosofía la argumentación</i>                                            | 536        |
| <br> |                                                                                                         |            |
|      | <b>Simposio:</b>                                                                                        |            |
|      | <b>La argumentación de la filosofía</b>                                                                 | <b>543</b> |
| 124  | <b>Álvarez Álvarez, J. Francisco</b>                                                                    |            |
|      | <i>Argumentación y vigilancia epistémica en entornos digitales</i>                                      | 544        |
| 125  | <b>Bermejo Luque, Lilian</b>                                                                            |            |
|      | <i>What is wrong with deductivism?</i>                                                                  | 545        |

|     |                                                                                                                                  |            |
|-----|----------------------------------------------------------------------------------------------------------------------------------|------------|
| 126 | <b>Corredor Lanas, Cristina</b>                                                                                                  |            |
|     | <i>Teoría de actos de habla y teoría de la argumentación</i>                                                                     | 550        |
| 127 | <b>Gascón Salvador, José A.</b>                                                                                                  |            |
|     | <i>Cómo argumentar con coherencia</i>                                                                                            | 556        |
| 128 | <b>Olmos Gómez, Paula</b>                                                                                                        |            |
|     | <i>Releyendo a Kuhn desde la teoría de la argumentación</i>                                                                      | 563        |
|     | <b>Simposio:</b>                                                                                                                 |            |
|     | <b>The genesis of geometric knowledge</b>                                                                                        | <b>571</b> |
| 129 | <b>Ferreirós, José</b>                                                                                                           |            |
|     | <i>Introduction</i>                                                                                                              | 572        |
| 130 | <b>de Paz, María</b>                                                                                                             |            |
|     | <i>The cognitive basis of Poincaré's philosophy of geometry</i>                                                                  | 573        |
| 131 | <b>Ferreirós, José</b>                                                                                                           |            |
|     | <i>Mathematics and cognitive science: The case of geometry</i>                                                                   | 575        |
| 132 | <b>García-Pérez, Manuel Jesús</b>                                                                                                |            |
|     | <i>The prehistorical roots of geometry: an overview</i>                                                                          | 578        |
|     | <b>Simposio:</b>                                                                                                                 |            |
|     | <b>Biological inter-identities</b>                                                                                               | <b>581</b> |
| 133 | <b>Etxeberria Agiriano, Arantza y Ruiz-Mirazo, Kepa</b>                                                                          |            |
|     | <i>Introduction</i>                                                                                                              | 582        |
| 134 | <b>Arnellos, Argyris</b>                                                                                                         |            |
|     | <i>Collaboration and biological individuality</i>                                                                                | 584        |
| 135 | <b>Bich, Leonardo y Frick, Ramiro</b>                                                                                            |            |
|     | <i>An organisational framework for the synthetic modeling of biological communication</i>                                        | 586        |
| 136 | <b>Etxeberria, Arantza y Nuño de la Rosa, Laura</b>                                                                              |            |
|     | <i>Biological individuality and reproduction: Intermittent interidentity in the evolution of pregnancy</i>                       | 588        |
| 137 | <b>Ruiz-Mirazo, Kepa; Shirt-Ediss, Ben y Moreno, Álvaro</b>                                                                      |            |
|     | <i>The construction of biological 'inter-identity' as the outcome of a complex process of prebiotic evolutionary development</i> | 590        |
| 138 | <b>Umerez, Jon y Ferreira, M. J.</b>                                                                                             |            |
|     | <i>Biological individuality and reproduction: Intermittent interidentity in the evolution of pregnancy</i>                       | 592        |

|                                                                                                                     |            |
|---------------------------------------------------------------------------------------------------------------------|------------|
| <b>Simposio:</b>                                                                                                    |            |
| <b>Evidencia científica y regulación: alimentos y fármacos</b>                                                      | <b>595</b> |
| 139Introducción                                                                                                     | 596        |
| 140Cámara Hurtado, Montaña                                                                                          |            |
| <i>Health claims. La frontera entre alimento y medicamento</i>                                                      | 597        |
| 141Luján, José Luis y Bengoetxea, Juan Bautista                                                                     |            |
| <i>Estándares de prueba y valores no cognitivos en la ciencia reguladora: el caso de las declaraciones de salud</i> | 598        |
| 142Tempini, Niccolo y Teira, David                                                                                  |            |
| <i>Los registros electrónicos sanitarios como evidencia regulatoria</i>                                             | 599        |
| 143Todt, Oliver y Sanz, Noemí                                                                                       |            |
| <i>La regulación de las declaraciones de salud y el debate público</i>                                              | 600        |
| <br>                                                                                                                |            |
| <b>Simposio:</b>                                                                                                    |            |
| <b>The bodily self</b>                                                                                              | <b>602</b> |
| 144Palmira, Michele y Verdejo, Víctor M.                                                                            |            |
| <i>Introduction</i>                                                                                                 | 603        |
| 145de Vignemont, Frédérique                                                                                         |            |
| <i>The phenomenal contrast of ownership</i>                                                                         | 605        |
| 146Salje, Léa                                                                                                       |            |
| <i>The inside-out binding problem</i>                                                                               | 607        |
| 147Serrahima, Carlota                                                                                               |            |
| <i>Ownership and immunity</i>                                                                                       | 608        |

## Prefacio

En este volumen se recogen las actas de las comunicaciones presentadas al IX Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España, celebrado los días del 13 al 16 de noviembre de 2018 en la Facultad de Filosofía de la sede central de la UNED, en Madrid.

Los congresos de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España se celebran cada tres años en una ciudad diferente y constituyen la reunión científica más longeva, multitudinaria e influyente en estos ámbitos en nuestro país. Estos congresos se llevan a cabo en colaboración entre la SLMFCE (<http://www.solofici.org/>) y la universidad anfitriona de cada edición. Tienen como objetivo el fomento, la divulgación y la promoción de los estudios e investigaciones en Lógica, Metodología de la Ciencia, Filosofía de la Ciencia y sus disciplinas afines, así como la interrelación con los diferentes especialistas y Sociedades e Instituciones científicas, nacionales e internacionales, interesadas en estos campos científicos. El congreso tuvo además una dimensión marcadamente internacional. El formato de las comunicaciones y ponencias fue multilingüe y muchos de los participantes provienen de universidades extranjeras.

En esta ocasión, el congreso celebrado en la UNED consiguió reunir a una gran cantidad de expertos en las áreas de investigación abordadas por la SLMFCE. En total, se recibieron 222 propuestas, entre comunicaciones y simposios, respondiendo a la llamada a la participación lanzada por los organizadores del congreso. Tras un arduo proceso de evaluación ciega dirigido por los prestigiosos miembros del comité científico, se aceptaron finalmente 178 trabajos, entre comunicaciones individuales repartidas en mesas temáticas y las agrupadas en ocho simposios dedicados a cuestiones específicas.

Además, en cada edición del congreso se celebran las conocidas “Conferencias Lullius”, en las cuales un reconocido experto ofrece tres conferencias plenarias. Estas conferencias fueron inauguradas por Philip Kitcher en 2012 en Santiago de Compostela y en 2015 en Barcelona fue Hartry Field el galardonado con este reconocimiento. En esta ocasión, fue la prestigiosa profesora Nancy Cartwright, de la universidad de Durham y la Universidad de California en San Diego (<http://www.profnancycartwright.com/>), quién impartió estas tres conferencias.

Se debe señalar asimismo que hubo una sesión de “Encuentro con los editores”, organizada por la revista *Theoria*. También se celebró una sesión dedicada a la profesora Amparo Gómez, que lamentablemente nos dejó recientemente y a quien sus compañeros pudimos homenajear en el congreso.

El IX Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia tuvo además un decidido compromiso en la promoción del trabajo de los investigadores más jóvenes. Prueba de esto es que se otorgaron dos premios de 500 euros cada uno a las mejores ponencias presentadas por estudiantes o recientes doctores.



Los organizadores de este congreso agradecemos la participación de todos los comunicantes, así como la ayuda de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España, la Facultad de Filosofía y el Departamento de Lógica, Historia y Filosofía de la Ciencia de la UNED, quienes proporcionaron los medios, humanos y materiales, necesarios para que este evento pudiera tener lugar. Además, este congreso estuvo enmarcado en los eventos auspiciados por la *International Union of History and Philosophy of Science and Technology*. Y, por supuesto, no podemos dejar de expresar nuestro reconocimiento a todos los miembros del Comité Científico, el Comité de Programa y el Comité Organizador Local, sin cuya colaboración nada de esto hubiera sido posible.

La primera edición de los congresos de la SLMFCE se celebró precisamente en la Universidad Complutense de Madrid y en la UNED en el año 1994. Hoy, tras 24 años y 8 ediciones, podemos decir con orgullo que este congreso ha vuelto a su primera casa y que lo hace exhibiendo un reconocimiento académico y unos niveles de calidad indiscutibles, tal y como demuestran los excelentes trabajos que se presentan en estas actas.

Cristian Saborido (UNED)  
Javier González de Prado (UNED / U. Nova de Lisboa)  
Sergi Oms (U. de Barcelona)

## Premio para Jóvenes Investigadores

Evalradas, al igual que el resto de las comunicaciones, de forma ciega por parte del comité científico, las ponencias que fueron premiadas con el Premio para Jóvenes Investigadores del IX Congreso de la SLMFCE fueron:

- Alfonso García Lapeña: “Truthlikeness for quantitative deterministic laws”
- Andrés Soria: “The Dynamics of Valued Judgment”

**Lullius Lectures 2018:**  
***Mid-level theory: Without it what  
could anyone do?***

**Nancy Cartwright**  
Durham University & UCSD

Philosophy of science used to be criticized for excessive focus on fundamental theory. Now we do better. We study experimenting, modelling, representing, and a host of other practices as well as the role of values in science. What we don't much study is non-fundamental theory—theory at the mid-level, which is where the heavy lifting is done. These lectures take us there.

Measurement practices are an example of mid-level theory par excellence. Mechanisms too are a big tool at mid-level. During the course of the lectures we will look at both, including mechanisms of 3 different kinds: structural mechanisms (like the ones familiar in philosophy of biology), 'theories of change' of the kind now widely required in policy planning and Jon Elster-type mechanisms, like priming and cognitive dissonance. The theory of the democratic peace, that democracies do not go to war with democracies, will serve as a running example.

Lecture 1. *Markers, mechanisms and measures.*

Lecture 2. *Causal-chain models and theories of change.*

Lecture 3. *Mechanisms and mid-level laws.*

LÓGICA, HISTORIA Y  
FILOSOFÍA DE LA LÓGICA, Y  
ARGUMENTACIÓN

# An intuitionistic understanding of unrestricted quantifiers

**Gonalo Baptista Santos**  
CFCUL, University of Lisbon

## **Abstract**

I discuss three ways of understanding unrestricted quantification. Namely, a contextualist approach according to which quantification is always subject to contextual restrictions, a modal approach according to which unrestricted claims should be interpreted in modal terms and an intuitionistic approach according to which the range of our unrestricted quantifiers should be understood as a generating principle. I also argue that the intuitionistic understanding of unrestricted quantification has significant advantages over the other approaches.

## **Keywords**

Indefinite extensibility, unrestricted quantification, contextualism, modal logic, intuitionism

Context sensitivity occurs in different contexts. For instance, imagine that, before going on a trip, someone says:

1. Everything fits in my hand luggage.

It is evident that, in this context, ‘everything’ refers to a collection of objects that does not contain absolutely everything. Many things cannot fit in hand luggage and yet (1) can be true. Nevertheless, explaining this obviousness is not easy.

As a starting point, we could assume that the content of the sentence is context sensitive. Context seems to contribute to the content of the sentence by somehow identifying a collection of things that includes all and only the objects that are necessary for the trip. When the context somehow distinguishes relevant and irrelevant objects, we can say that there is a contextual restriction of the domain of quantification.

It is only in very particular contexts that our use of ‘everything’ does not appear to assume a distinction between relevant and irrelevant objects. For instance, imagine a philosophical discussion where someone says:

- 2 Everything obeys the laws of physics.

In this context, determining the truth-value of the sentence does not presuppose a distinction between relevant and irrelevant objects. It is the opposite that happens. Given that the truth-value of (2) does not suppose a difference between relevant and irrelevant objects, it would appear that contextual domain

restriction does not occur in this context. We could then say that in (2) our quantifiers are entirely unrestricted. The two previous examples seem to illustrate the following facts. Our use of ‘everything’ in everyday discourse allows us to make claims about all the things that are relevant in a certain context. Our use of ‘everything’ in specific theoretical contexts allows us to make claims about absolutely everything. Contextualists like Michael Glanzberg (2004, 2006) have denied this second fact, arguing that we can never talk about absolutely all things.

One of my goals is to dispute this claim. I will argue that in theoretical contexts where a sentence like (2) might be employed, contextual domain restriction does not occur and ‘everything’ can be said to be unrestricted. My other goal is to argue for an understanding of unrestricted quantification that allows us to make claims about absolutely everything, without assuming a domain of quantification with absolutely everything.

I discuss two strategies for achieving this goal. One involves a modal interpretation of (apparently) unrestricted claims like (2). The general idea is that we should interpret this claim as saying that necessarily everything obeys the laws of physics. There are different formulations of this strategy, but in this work, I focus my discussion on Fine’s (2006) proposal (see Linnebo (2010, 2013) and Studd (2013, 2015, 2017) for alternative formulations).

According to Fine, a quantifier is always associated with a particular interpretation of quantification. A world contains different interpretations of quantification, and different worlds have different collections of interpretations. One of the most distinguishing features of this approach is that while no world contains an interpretation where our variables range over absolutely everything, it is always possible to find a more encompassing interpretation of quantification.

Both the modal approach and the contextualist proposal agree that our variables never range over the elements of a collection that contains absolutely everything. One advantage of the modal proposal is that it manages to defend this claim without assuming a distinction between relevant and irrelevant objects. In other words, the modal approach does not consider the occurrence of contextual domain restriction in theoretical contexts.

The other strategy that I want to discuss involves an intuitionistic interpretation of (apparently) unrestricted claims like (2). According to it, the scope of our unrestricted quantifiers should not be understood as a completed collection but rather as a generating principle. Just as understanding our arithmetical quantifiers intuitionistically allows us to quantify over all numbers but does not commit us to a collection of all numbers, adopting an analogous understanding of our unrestricted quantifiers allows us to quantify over everything but does not commit us to a collection with absolutely all things.

The intuitionistic approach inherits the advantages that the modal approach has over the contextualist approach. Moreover, I argue that the intuitionistic approach has significant advantages over the modal approach to unrestricted quantification. One distinct advantage is that while the modal approach cannot employ a unique unrestricted interpretation of quantification (in fact, depending on how we interpret the relevant modality, there might be infinite unrestricted interpretations), the intuitionistic approach only employs one interpretation of

unrestricted quantification. Another disadvantage of the modal approach is that it commits us to the existence of a modal realm. This commitment is undesirable because, on a natural understanding of this realm, its existence is equivalent to assuming the existence of a collection that contains absolutely everything.

## References

- Kit Fine. “Relatively Unrestricted Quantification”. In: *Absolute Generality*. Eds. Agustín Rayo and Gabriel Uzquiano. Oxford: Oxford University Press, 2006, 20–44.
- Michael Glanzberg. “Quantification and Realism”. *Philosophy and Phenomenological Research* 69(3) (2004):541–572.
- Michael Glanzberg. “Context and unrestricted quantification”. In: *Absolute Generality*. Eds. Agustín Rayo and Gabriel Uzquiano. Oxford: Oxford University Press, 2006, 45–74.
- Øystein Linnebo. “Pluralities and Sets”. *Journal of Philosophy* 107(3) (2010):144–164.
- Øystein Linnebo. “The Potential Hierarchy of Sets”. *Review of Symbolic Logic*, 6(2) (2013):205–228.
- James P. Studd. “The Iterative Conception of Set: a (Bi-)Modal Axiomatization”. *Journal of Philosophical Logic* 42 (5) (2013):1–29.
- James P. Studd. “Semantic Pessimism and Absolute Generality”. In: *Quantifiers, Quantifiers, Quantifiers*. Ed. Alessandro Torza. Dordrecht: Synthese Library (2015).
- James P. Studd. “Generality, Paradox and Extensibility”. *Proceedings of the Aristotelian Society* 117(1) (2017):81–101.

# Razonamiento experimental e inferencia en la práctica médica

Cristina Barés Gómez  
Matthieu Fontaine

## Resumen corto:

La medicina está considerada como una ciencia experimental. Podemos explicar los procesos desencadenadores de lo que no es la situación de bienestar del cuerpo humano mediante una supuesta enfermedad. Esta enfermedad planteada en el diagnóstico médico sería una hipótesis luego contrastada experimentalmente. La contrastación empírica es una de las bases fundamentales de la medicina moderna que nos aporta seguridad y fiabilidad. Sin embargo, ¿se realizan siempre las pruebas médicas para el tratamiento de los pacientes? Si no es el caso, ¿sigue siendo un razonamiento experimental? ¿Cuál es la diferencia en ambos casos a nivel inferencial? En esta comunicación vamos a analizar el diagnóstico médico desde un análisis de la inferencia comparando un diagnóstico con y sin pruebas clínicas. Consideramos que la inferencia es abductiva y la analizamos usando el modelo G-W de Gabbay y Woods (2005,2013) para la abducción comparándola con el ST de Magnani (2001). Seguidamente usaremos la lógica dialógica para destacar los aspectos fundamentales de esta inferencia. La lógica dialógica nos permite aproximarnos a la abducción desde un punto de vista pragmático. Pretendemos analizar sus aspectos más destacados tales como la introducción del agente, la acción y la retractabilidad.

La medicina está considerada como una ciencia experimental. Podemos explicar los procesos desencadenadores de lo que no es la situación de bienestar del cuerpo humano mediante una supuesta enfermedad. Esta enfermedad planteada en el diagnóstico médico sería una hipótesis luego contrastada experimentalmente. La contrastación empírica es una de las bases fundamentales de la medicina moderna que nos aporta seguridad y fiabilidad. Sin embargo, ¿se realizan siempre las pruebas médicas para el tratamiento de los pacientes? Si no es el caso, ¿sigue siendo un razonamiento experimental? En esta comunicación vamos a analizar el diagnóstico médico como expresión de la ciencia médica. Hay dos tipos de diagnósticos, al menos a grandes rasgos. El primero se basa en inducción probabilística. Este diagnóstico, aunque útil, es muy poco utilizado en la práctica médica diaria. El segundo se basa en el razonamiento causal para el diagnóstico etiológico. Los doctores generan un conjunto de diagnósticos hipotéticos y los prueban obteniendo nuevos datos. Sin embargo, el último paso (las pruebas clínicas) no siempre se cumple en la práctica diaria.

Centrándonos en el último diagnóstico basado en razonamiento causal, analizaremos casos concretos. Usaremos casos de médicos generalistas tratando enfer-



medades como malaria, infección de orina o varios virus. Consideramos que la inferencia llevada a cabo en los procesos de diagnóstico médico es una inferencia abductiva, pues postulamos una enfermedad hipotética a partir de sus signos, síntomas. La abducción es una inferencia que no es ni deducción ni inducción, sino que su aspecto principal es tratar con hipótesis. La definición canónica de abducción se la debemos a Peirce [CP 5.189]:

*The surprising fact C is observed.  
But if A were true, C would be a matter of course.  
Hence there is reason to suspect that A is true.*

Debemos destacar entre otros aspectos el hecho sorprendente de la primera premisa. La caracterización de los signos o síntomas como algo fuera del estado normal de bienestar de una persona. El carácter subjuntivo de la segunda. Ello caracteriza la posible relación subjuntiva entre los signos y síntomas y la hipotética enfermedad. La posibilidad del tercer término del silogismo como conjetura de la enfermedad hipotética.

Analizaremos estos casos primero teniendo en cuenta el modelo ST para el diagnóstico médico de Magnani (2001). En este modelo se considera un primer paso abductivo que luego se completa con pruebas clínicas formando lo que llama un ciclo abducción/deducción – inducción. A pesar de que en muchos casos obtenemos pruebas que corroboren el diagnóstico y la inferencia correspondería con este ciclo, consideramos que no siempre se trabaja así en medicina. Generalmente los recursos no son tan ilimitados y el diagnóstico carece de la fiabilidad de las pruebas clínicas. Estos casos son una inferencia desde la mejor explicación, inferencia que preserva la ignorancia puesto que representan una abducción completa. Esta inferencia es abductiva basada en hipótesis y en la que no se llega a comprobar esa conjetura. En estos casos concretos, los médicos actúan sin tener unas pruebas concluyentes y preservando la ignorancia. Analizaremos esos diagnósticos desde el punto de vista de la inferencia, una inferencia abductiva completa usando el modelo G-W para la abducción (2005, 2013).

En este punto nos planteamos las siguientes preguntas. Una vez analizados los casos en los que no existen pruebas clínicas, ¿seguimos estando ante un razonamiento experimental? Desde el punto de vista de la inferencia, ¿es la abducción una inferencia ligada a la acción y la experiencia? Nuestro punto de vista es que sí. A pesar de que no existan pruebas clínicas, el razonamiento médico sigue siendo un razonamiento experimental. Siguiendo algunas ideas de Claude Bernard (1966) explicaremos la posibilidad de la no experimentación directa en el diagnóstico médico como razonamiento experimental. Desde el punto de vista de la inferencia, planteamos la abducción como un razonamiento práctico totalmente ligado al agente y que forma parte de lo que Woods (2013) llama tercera-forma de razonamiento. Para ello nos enfocaremos en la inferencia como tal. Plantearemos una interpretación dialógica de la abducción que pretende precisamente capturar la acción del agente y la retractabilidad de la abducción en general y del diagnóstico médico en particular. Con nuestra interpretación dialógica pretendemos destacar los aspectos del modelo G-W de la abducción tales como triggering, guessing and committing. El aspecto pragmático de la lógica dialógica nos permite acercarnos a la inferencia abductiva desde otro punto de vista planteándonos el significado de estos aspectos en un diálogo.

## Referencias

- BARES GÓMEZ, C. Forthcoming. Abduction in Akkadian Medical Diagnosis. *Journal of Applied Logics – IfCoLoG*.
- BARES GOMEZ, C. & FONTAINE, M. 2017. Argumentation and abduction in dialogical logic. In L. Magnani and T. Bertolotti, editors, *Handbook of Model-Base Science*. Springer.
- GABBAY, D. & WOODS, J. 2005. *The Reach of Abduction – Insight and Trial*. Elsevier. Amsterdam.
- BERNARD, C. 1966. *Introduction à l'étude de la médecine expérimentale*. Garnier Flammarion, Paris.
- MAGNANI, L. 2001. *Abduction, Reason, and Science. Processes of Discovery and Explanation*. Springer.
- PEIRCE, C. 1931-1958. *Collected Papers*. Harvard University Press. Cambridge.
- WOODS, J. 2013. *Errors of Reasoning*. College Publications. London.

# Imperatives and to-do lists: three objections

Sandy Berkovski

## Shorter abstract

Departing from Stalnaker’s theory of common ground, a recent pragmatic theory of imperatives aims to identify a concrete pragmatic mechanism through which the speaker directs (commands, requests, instructs) the hearer to perform a particular action. The utterer of an imperative aims to update the hearer’s To-Do List by including in it the relevant task. I argue that this account relies on a problematic notion of sentential force, is threatened by an infinite regress, and fails to account for different kinds of directive force.

§1. Under Stalnaker’s dynamic analysis of conversation, each participant comes equipped with a set of presuppositions, that is, propositions that he accepts for the purpose of the conversation. The *common ground* of the conversation comprises all and only propositions every participant accepts. The speaker, in asserting a declarative sentence, proposes to *update* the common ground. The pragmatic force of the assertion, what the speaker is trying to achieve by making that assertion, can be interpreted as just such an update.<sup>1</sup>

For each participant, his presuppositions are either true or false, and it is their truth or falsehood that is being accepted for the purpose of conversation. Assertion is also true or false. So we could say that the kind of conversation that Stalnaker is concerned with is a conversation about the state of the world in which various claims about the world are advanced, accepted, and rejected.

Framed this way, this approach appears somewhat limited. Not every conversation people conduct is about the state of the world, and even if it were, not every background assumption people make prior to the conversation is about the state of the world. Shouldn’t these unnamed assumptions have any effect on the conversation? Plenty of conversations contain directive speech acts in which the speakers command, request, and instruct. On many views at least, these utterances do not express truth-conditional content, and so fall by the wayside. And this is regrettable: we know that directives govern conversations where regularities are observed—namely, speakers direct and hearers comply, or refuse to do so. Accounting for such regularities in the dynamic evolution of conversations was precisely the target of Stalnaker’s analysis.

Now, even though directives were examined in speech act theories, the pragmatic mechanism was left obscure.<sup>2</sup> Directives were said to ‘invoke the authority’ of the speaker, to represent an attempt to get the hearer to act, or to give the hearer a

---

<sup>1</sup>For a recent outline, see Stalnaker (2014, 24–25, 46–53).

<sup>2</sup>See, e.g., Searle and Vanderveken (1985) and Bach and Harnish (1979).

‘sufficient reason’ for performance. Little else was added by way of clarification, mostly because any such clarification was thought to fall outside the purview of the philosophy of language. Hence the following problem emerges:

**Representation Problem** What is the pragmatic mechanism through which the speaker directs (commands, requests, instructs) the hearer to perform a particular action?

§2. The lacunae I have described can be filled. On one hand, we could deny that imperatives have no truth value. Thus we could say that they have propositional deontic content: an imperative  $!\phi$  has the propositional content ‘You should  $\phi$ .’<sup>3</sup> If, therefore, they are assertible just like declarative utterances are, then they can be added to the common ground and no special treatment is in order. Let us though not choose the easy route. Let us maintain with the majority that imperatives have no truth value. Whatever they express cannot thus be added to the common ground. But even so, a solution is just around the corner.<sup>4</sup>

We could say that each participant of the conversation comes equipped not only with a body of presuppositions out of which the common ground is extracted, but also with ‘To-Do Lists’. These lists specify the actions to be performed by the owner. We approach most, if not all, conversations with some idea of what needs to be done, what we have to do. So in uttering an imperative sentence the speaker seeks to update the To-Do list of the addressee. But since we have agreed that imperatives cannot be propositions, and in so far as we wish to maintain the analogy with declaratives (or in other words, to give a unified account of imperatives and declaratives), separate semantic rules  $R(S) = I$  have to be provided that take an imperative  $S$  and yield its semantic value  $I$  to be placed as an item on the hearer’s TDL.

It is immaterial for the purposes of my argument how these semantic rules are fixed. For specificity, I adopt Portner’s proposal that semantic values of imperatives are properties to be uniquely satisfied by the addressee. So, e.g., the semantic value of ‘Leave!’ addressed to  $a$  would be a property construct ‘ $a$  leaves’. More precisely:<sup>5</sup>

$$\llbracket !\phi \rrbracket^c = [\lambda w \lambda x : x \text{ is the addressee} . x \text{ is } \phi\text{-ing in } w]. \quad (1)$$

Properties thus defined can be satisfied at a world  $\mathbf{w}$  by an individual  $\mathbf{i}$ , in which case we write  $P(\mathbf{w}, \mathbf{i}) = 1$ .

The utterance of the imperative targets the addition of its semantic value to the TDL of the addressee. Letting  $L(x)$  be the TDL of the addressee  $x$  prior to the utterance of the imperative  $!\phi$  in the context  $c$ , the update function takes this simple form:

$$v(L(x)) = L(x) \cup \{\llbracket !\phi \rrbracket^c\}. \quad (2)$$

---

<sup>3</sup>See, e.g., Kaufmann (2012).

<sup>4</sup>As developed in Portner (2004, 2007) and also in Han (2011). See Portner (2016) for a review.

<sup>5</sup>Following the conventions in Heim and Kratzer (1998).

This account, to which I will now refer as the ‘theory  $\mathfrak{T}$ ’, claims for itself the virtue of solving the Representation Problem. It claims to have identified a concrete pragmatic mechanism of To-Do Lists. It describes the conditions of rationality for the hearer in the situation where a directive is received. We define a partial ordering of worlds as follows:

$$\forall w \forall u (w \leq_i u) \text{ iff: } \{P \mid P \in L(i) \ \& \ P(w, i) = 1\} \subseteq \{P \mid P \in L(i) \ \& \ P(u, i) = 1\}, \quad (3)$$

where  $i$  is the addressee of the imperative, and the worlds  $w, u$  are in the context set. Once he indicates the acceptance of the directive, the addressee would wish to actualise the worlds with the highest rank. That is, he would try to make true as many properties on his TDL as possible.<sup>6</sup>

**§3.** Having laid out the elements of  $\mathfrak{T}$ , I wish now to argue that the Representation Problem has not been solved. I will conduct my discussion under a number of rubrics.

**Sentential force.** The first worry I raise is a broad methodological one. As Portner (along with some other theorists) sees it, TDLs and the associated with them update functions characterise imperative *sentences*—more precisely, the sentential forces of those sentences:

Distinctions among subtypes of imperatives—orders, requests, permissions, etc.—should not be understood at the level of conversational force [i.e. sentential force]. That is, they all share the [sentential] force of Requiring, as they all conventionally add a property to the addressee’s To-Do list. Rather, these differences have to do with the pragmatic or sociolinguistic basis for the speaker’s attempt to add a property to the addressee’s To-Do list. (Portner, 2004, 237–238)

Sentential forces are the fundamental conversational functions with which sentence moods are associated. (Portner, 2018, 122–23)

In other words: There is a minimal pragmatic force of Requiring conventionally associated with imperative mood. Further varieties of that force are attributable to contextual features. But now, suppose that Ian wishes to tell Jon to write a certain letter. What in present terms this wish amounts to is putting a property construct ‘writing the letter’ on Jon’s TDL. If imperative mood have are assigned the force of Requiring, then a sound strategy for Ian would be to utter a sentence in the imperative mood. But this is not so. There is any number of linguistic forms Ian may use to achieve his purpose:

- (i) Write the letter!
- (ii) I order you to write the letter.
- (iii) Would you please write the letter?
- (iv) You should write the letter.

---

<sup>6</sup>See Portner (2004, 243–3).

(v) It is a very good idea for you to write the letter.

(vi) Etc.

All of these utterances have the force of Requiring, but only one sentence there is in the imperative mood. Different contextual elements conspire to yield a particular pragmatic force, and not just the mood. Nor does an imperative sentence always yield the force of Requiring (e.g., in daring the hearer to do something). Thus the theory  $\mathfrak{T}$  is threatened by the same objections as the classical speech act theory. It has to fall back on some version of the ‘literal meaning hypothesis’, but that view, I argue, has largely been discredited.

**Infinite regress.** As already mentioned, in its pragmatic analysis the theory  $\mathfrak{T}$  makes no distinction between different kinds of speech acts, such as commands, orders, requests, or instructions. The TDL-based explanation is supposed to replace all the vague talk about the different kinds of directive force which was a staple notion in the earlier speech-act analysis. But this masks a shortcoming. For suppose we grant that the speaker  $S$  intends to update the hearer  $H$ ’s TDL  $L(H)$ . Yet when  $S$  commands, he should then *command*  $H$  to adjust  $L(H)$ . If  $S$  makes a request for  $H$ , then he should *request*  $H$  to adjust  $L(H)$ . The original directive of the form ‘Do  $A$ !’ induces another directive of the form ‘Place “doing  $A$ ” on your To-Do list!’ Correspondingly, these further directives should receive the same treatment as the original one. Hence we end up with an infinite sequence of directives to perform  $\phi$ -ing, TDLs, directives to update TDLs, meta-TDLs, directives to update meta-TDLs, and so forth. A regress is under way which apparently cannot be terminated.

It is no use to retort that the task of  $\mathfrak{T}$  is merely to chart the causal path from utterances to action. Not only is there any number of such paths (as I have claimed earlier), and not only does  $\mathfrak{T}$  have normative pretensions, but also, crucially, the TDL update is not a physical or psychological event to be recorded and to be part of a causal chain. It is rather a theoretical device for understanding what happens in a linguistic interaction.

Both complaints just sketched target  $\mathfrak{T}$ ’s interpretation of the speaker’s intention. The joint response here may be that the goal of  $\mathfrak{T}$  does not include modelling the speaker’s intention at all. It is rather to model the hearer’s rationality (but not the actual compliance). To model compliance we have no need of TDLs. But we can only interpret rationality through TDLs. Pragmatics asks the question: what is taking place in the interaction between the speaker and the hearer? We concede that the speaker’s intention remains uninterpreted in our model, but insist that the rationality of the hearer has been taken care of (see Table 1). I will now argue that  $\mathfrak{T}$  fails also in that more limited endeavour.

**The disappearance of directive force.** Suppose you receive a request to write a letter. You accept the request, and this, according to  $\mathfrak{T}$ , amounts to changing your preferences. Whereas before your preference for spending tonight was to watch a film, now your preference for tonight becomes writing a letter. So this shift can be represented as an update of your TDL. Suppose now you receive a *command* to write a letter. If you accept, the change in your condition

| Notions                           | Interpretations                                          |
|-----------------------------------|----------------------------------------------------------|
| Speaker's intention               | ?                                                        |
| Hearer's acceptance of $U$        | Updating TDL                                             |
| Hearer's rejection of $U$         | TDL not updated                                          |
| Hearer's rationality of behaviour | Maximising the number of saturated properties on the TDL |

Table 1: Achievements of the theory  $\mathfrak{T}$

is described in exactly the same way. But this is to miss something important. A hearer may refuse to accept the command—i.e. to update TDL—precisely because it is a command ('I refuse to even consider this—you have no right to boss me around!'). On other occasions, the hearer would do the opposite: 'If he simply asked me, I would have refused; but since he commanded, I feel compelled etc.' The complaint I am lodging is that, for the hearer's rationality to be interpreted, we have to take into account the directive force of the utterance. Whether the TDL is updated depends on the illocutionary force of the speaker's utterance.

The same problem emerges when we turn to the actual performance. The hearer's rationality is constrained by maximising the number of saturated properties on the TDL (the partial ordering in (3)). But shouldn't a provision be made for the source of those properties? One reason a hearer can give for saturating the property  $\llbracket!\phi\rrbracket$  as opposed to  $\llbracket!\psi\rrbracket$  is that he was commanded to  $\phi$ , but only requested to  $\psi$ .

Now it is tempting to think that both of these difficulties can be dealt with if we introduced a partial ordering of the elements of TDL. Such an ordering would reflect the relative importance of the elements. But how to cash out this relative importance? Clearly it should reflect their priority in the actual performance. We imagine, simplifying greatly, that the hearer has only to perform one task at a time. Then, other things equal, he will perform the one ranked higher on the list. Thus we could say:

$$\forall P \forall Q (P \preceq Q) \text{ iff: } \forall w \forall u ((\forall R (R \neq P \ \& \ R \neq Q \rightarrow R(w) = R(u)) \ \& \ P(w) = Q(u) = 1 \ \& \ P(u) = Q(w) = 0) \rightarrow w \leq u). \quad (4)$$

A property  $Q$  has a higher directive rank than  $P$  just in case of the two worlds otherwise identical, the one where  $Q$  is saturated is more rational to actualise than the one where  $P$  is saturated. But within the framework of  $\mathfrak{T}$ , this move gets things exactly backwards. We defined the ordering on worlds via the properties of TDL. Now, however, we define the properties of TDL via the ordering of worlds.

Similarly, instead of a simple formula (2), we have to have an update function that inserts new properties into the TDL in accordance with their rank. Conceivably, directive force is a factor determining the rank. But it is also clear that the rank, according to which actions are subsequently performed, is not functionally determined by directive force: e.g., some requests may take priority over

some commands. How this determination is worked out, if it can be worked out, no pragmatic theory can say. We are in the terrain of a decision theory where additional variables should be supplied with the help of notions foreign to the pragmatic theory. In the absence of any such further much more complicated formula, the update function in (2) remains an idle fiction. Secondly, the hearer must demonstrate rationality already at this stage of interaction. We should be able to say already when he acquiesces to the command that the property is inserted with a certain rank. This goes against the ambition of  $\mathfrak{T}$ , where rationality is identified simply with an actualisation of whatever items there are on the list. We are in need of a much more fine-grained description of rationality.

**§4.** None of the difficulties I have raised apply to the dynamic account of declaratives (i.e. to the account of assertion). Contrast the assertion:

Snow is white (5)

with the command:

Believe that snow is white! (6)

What the speaker does with (5) is to assert the truth of a statement. What he does with (6) is to influence, in some way, the beliefs of the interlocutor. A proper reaction to the former would be:

Yes, it is/No, it is not, (7)

whilst the proper reaction to the latter would be:

All right, I will/No, I will not. (8)

There are two different conversations here. In one of them people discuss properties of snow. In the other the subject is the beliefs of the addressee. To assimilate the conversation about snow to the conversation about beliefs is to say that people never describe the world, that the only legitimate subject of conversation is the epistemic state of the interlocutor. In any event, as I see it, Stalnaker's analysis proceeds on the assumption of rejection of that assimilation and applies to the first kind of conversation.

In the conversation where assertions are made the speakers communicate information. To model the mechanism of such a conversation one has to show how information is interpreted and processed, and how this processing affects the next stage in the conversation. Thus Stalnaker's goal in using the two-dimensional apparatus was to show both how speech acts are interpreted in a context, and how, on the other hand, they change the context itself.<sup>7</sup>

When viewed this way, the idea of extending Stalnaker's account to the case of imperatives appears suspect at the outset. A linguistic interaction between the utterer of imperatives and the hearer would only misleadingly be characterised as a conversation where certain information is communicated. The goal of that interaction is to yield a performance. A successful communication of information is often a link in the causal chain towards performance, but it does not

---

<sup>7</sup>For some clear statements see, e.g., Stalnaker (1999, 8–9).



have to be so. To believe otherwise to ignore the gap between two kinds of linguistic interaction. Moreover, modelling that communication must account for the directive force of the utterance—a feature that cannot be accommodated in Stalnaker’s framework.

## References

- Bach, K. and Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. MIT Press.
- Han, C.-H. (2011). Imperatives. In von Stechow, K., Maienborn, C., and Portner, P., editors, *Semantics: An International Handbook of Natural Language Meaning*, volume 2, pages 1785–1804. De Gruyter.
- Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar*. Blackwell.
- Kaufmann, M. (2012). *Interpreting Imperatives*. Springer.
- Portner, P. (2004). The semantics of imperatives within a theory of clause types. *Semantics and Linguistic Theory*, 14:235–252.
- Portner, P. (2007). Imperatives and modals. *Natural Language Semantics*, 15:351–383.
- Portner, P. (2016). Imperatives. In Aloni, M. and Dekker, P., editors, *The Cambridge Handbook of Formal Semantics*, pages 593–628. Cambridge University Press.
- Portner, P. (2018). *Mood*. Oxford University Press.
- Searle, J. R. and Vanderveken, D. (1985). *Foundations of Illocutionary Logic*. Cambridge University Press.
- Stalnaker, R. C. (1999). *Context and Content*. Oxford University Press.
- Stalnaker, R. C. (2014). *Context*. Oxford University Press.

# Frege's adoption of the distinction between concept and object

Joan Bertran San Millán<sup>1</sup>

ABSTRACT. I discuss Frege's philosophical transit from *Begriffsschrift* to *Grundlagen*. I focus on the replacement of function-argument scheme with the concept-object scheme, which has received virtually no attention in historical studies.

*Begriffsschrift*'s reception and Frege's response to criticisms are taken in (Heck; May, 2013) as the main reason for Frege's adoption of the concept-object scheme. I claim that there are two additional causes of Frege's move. First, in the context of the deployment of the logicist project, Frege noticed the limitations of *Begriffsschrift*'s formal system and the need to establish an ontological foundation – based on the concept-object scheme – that allowed him to provide a definition of natural number. Second, the function-argument scheme provided an inadequate analysis of atomic statements.

I defend that the concept-object scheme should not be seen as an extension of the function-argument scheme. What is essential to the notion of concept is its unsaturatedness, which is never attributed to *Begriffsschrift*'s notion of function. From this stance I argue that two features show the virtues of the concept-object scheme: first, Frege's notion of concept sets a foundation for the principle of the priority of judgements over concepts; second, the concept-object scheme allows an adequate semantical analysis of categorical statements.

In *Begriffsschrift* (1879) Frege presented a formal system, the concept-script, which was shaped in terms of the distinction between function and argument. The five years that passed from the publication of *Begriffsschrift* to that of *Grundlagen* (1884) show an intense philosophical development. The abandonment of the distinction between function and argument and the adoption of the distinction between concept and object stand out as one of the key moves in this development. As radical as this may seem, the disappearance of the function-argument scheme in Frege's works – at least, until 1891 – has received virtually no attention in historical studies<sup>2</sup>.

In this talk I discuss Frege's philosophical transit from *Begriffsschrift* to *Grundlagen*. I put the focus on the replacement of the distinction between function

---

<sup>1</sup>The work on this paper was supported by the *Formal Epistemology – the Future Synthesis* grant, in the framework of the *Praemium Academicum* programme of the Czech Academy of Sciences.

<sup>2</sup>In fact, the function-argument scheme is often taken as equivalent to the concept-object scheme. See (Baker, 2001, pp. 537–538), (Weiner, 2004, pp. 76–77), (Textor, 2011, pp. 76–77) or (Kanterian, 2012, p. 139).

and argument with the distinction between concept and object.

Firstly, I clarify the causes for Frege's replacement of the distinction between function and argument with the distinction between concept and object. Frege confirmed in the early 1880s that the nature of the function-argument scheme had been completely misunderstood by the reviewers of *Begriffsschrift*. This alone can be seen as a first reason for the abandonment of this scheme. However, I suggest two additional causes<sup>3</sup>. On the one hand, in the context of the deployment of the logicist project – openly assumed in 1882, in a letter to Stumpf (Frege, 1976, pp. 163–165) – Frege noticed the limitations of *Begriffsschrift*'s formal system and the need to establish an ontological foundation – based on the distinction between concept and object – that allowed him to provide a definition of the notion of natural number. On the other, Frege was aware that the function-argument scheme provided an inadequate analysis of atomic statements, given that it did not capture the semantic structure of these expressions. If ' $\Phi(A)$ ' represents an atomic expression divided into a function (' $\Phi$ ') and an argument (' $A$ ') then, since the original statement can be divided into function and argument in multiple ways, it cannot be determined whether the component represented with ' $\Phi$ ' corresponds to a concept or to a relation.

Secondly, I explain what characterises the concept-object scheme. Preliminarily, I defend that the adoption of this scheme should not be seen as an extension of the application of the function-argument scheme. Both structures have different aims and divergent natures. Specifically, the distinction between concept and object is applied to the content of an expression, not to the expression itself. What is essential to the notion of concept is its unsaturatedness, which is never attributed to *Begriffsschrift*'s notion of function. While an object is always saturated, a concept is defined as unsaturated, that is, as an entity in need of completion.

Finally, based on the previous analysis, I claim that there are two features that show the virtues of the concept-object scheme in Frege's philosophical framework. First, Frege's notion of concept sets a foundation for the principle of the priority of judgements over concepts. A concept needs the completion of an object, and this act of completion takes place in the formation of a statement. Hence, a categorical statement does not come from the connection of two concepts, as had been maintained in the Aristotelian tradition. In fact, concepts are obtained through the decomposition of a statement. Thus, the original object of logical analysis are statements. The decomposition of the content of a statement in terms of concept and object reveals the semantic structure of statements. As it is often noted, this decomposition is not unique and can be made in different ways. In contrast to the semantic neutrality of *Begriffsschrift*'s distinction between function and argument, the multiplicity of possible analyses in terms of concept and object of a single statement is completely subordinated to the semantic structure of that statement. This means that, while the roles of function and argument can be switched, the subject of a statement can never be taken

---

<sup>3</sup>The reception of *Begriffsschrift* and Frege's response are taken in (Heck; May, 2013) as the main reason for Frege's adoption of the concept-object scheme. However, the reviews did not address the shortcomings of the function-argument scheme. In this sense, I revise the picture given in (Heck; May, 2013) and locate the *internal* reasons that drove Frege to abandon the function-argument scheme.

as a symbol for a concept: only predicates can be associated with concepts. Therefore, the flexibility of the application of the distinction between function and argument must be distinguished from the multitude of decompositions of the content of a statement that can be made according to the concept-object scheme.

Second, the distinction between concept and object allows an adequate semantical analysis of categorical statements. In the first years of the 1880s Frege made efforts to emphasise the advantages of *Begriffsschrift*'s formal system over Boolean logic. He stressed that the concept-script could provide a satisfactory logical analysis of categorical judgements. This was noteworthy, since at this time algebraic logicians could not rigorously render particular statements using Boolean logic. Frege was in disposition to say that a statement such as 'All men are mortal' should be analysed as a universally quantified conditional. However, from the analysis in terms of function and argument proposed in *Begriffsschrift* it cannot be determined whether 'All men are mortal' states the subordination of a concept into another or whether 'Socrates is a man' expresses that an object falls into a concept. Only when the distinction between concept and object is incorporated in the language of the formal system, such a semantical analysis can be made. Therefore, on the basis of the combination of the formal features of the concept-script and the concept-object scheme made in the papers written between 1880 and 1882, Frege claimed that he had provided an adequate analysis of categorical judgements. This move marked a clear departure from *Begriffsschrift* and anticipated one of the core elements of *Grundlagen*.

## References

- BAKER, G. P. (2001). 'Function' in Frege's *Begriffsschrift*: Dissolving the Problem. *British Journal for the History of Philosophy*, 9, pp. 525–544.
- FREGE, G. (1879). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Louis Nebert.
- FREGE, G. (1884). *Die Grundlagen der Arithmetik: eine logisch-matematische Untersuchung über den Begriff der Zahl*. Breslau: Wilhelm Koebner.
- FREGE, G.. (1976). *Wissenschaftlicher Briefwechsel*. GABRIEL, G.; HERMES, H.; KAMBARTEL, F.; VERAART, A. (Eds.). Hamburg: Felix Meiner.
- HECK, R.; MAY, R. (2013). The Function is Unsaturated. In BEANEY, M. (Ed.). *Oxford Handbook of the History of Analytic Philosophy*. Oxford: Oxford University Press, pp. 825–850.
- KANTERIAN, E. (2012). *Gottlob Frege. A Guide for the Perplexed*. London: Continuum.
- TEXTOR, M. (2011). *Frege on Sense and Reference*. London: Routledge.
- WEINER, J. (2004). *Frege Explained*. Chicago: Open Court.

# Contingent composition as identity

**Massimiliano Carrara**

FISPPA - Università di Padova

**Giorgio Lando**

DSU - Università dell'Aquila

The debate about Composition as Identity (henceforth CAI, for a general introduction to the topic see Lando 2017) is focused on various constitutive features of identity, and follows a recurring pattern: the question is asked whether composition enjoys a certain constitutive feature of identity and whether, as a consequence, composition qualifies—from the viewpoint of that feature—as a genuine form of identity. If composition enjoys this constitutive feature of identity, then the friend of CAI marks a point. If, by contrast, composition turns out to lack this constitutive feature of identity, then the foe of CAI marks a point.

In this paper, we are going to show that, when CAI is confronted with the modal features of identity, and in particular with the necessity of identity (NI), we should be careful in following this recurring pattern. In particular, it is methodologically desirable to construe the modal notions involved as standardly as possible. This already happens when, for example, the debate on CAI is focused on the principle of the Indiscernibility of Identicals, as it happens in van Inwagen (1994), Cotnoir (2013), Bohn (2014) and Carrara & Lando (2017): a standard, unrestricted version of this principle is considered, and the question is raised whether composition respects it.

The reason why it would be misleading to assume instead non-standard contentions in the assessment of CAI is not that these non-standard theories are provably false, but simply that they are non-standard. Inasmuch as they are not standard, they are unsuitable as a common, neutral and conservative background for the assessment of CAI.

We will argue that it is equally methodologically desirable to adopt standard, neutral contentions when modality is at stake, and in particular when the question is raised whether composition is as necessary as identity. By contrast, in the extant literature about CAI and NI (Merricks (1999), Borghini (2005) and Wallace (2014)) a non-standard approach to modality is adopted (either counterpart theory or the theory of modal parts), and on this basis CAI is shown to be incompatible (in the case of Merricks) or compatible (in the cases of Borghini and Wallace) with NI. Our analysis assumes as a standard philosophy of modality (SM) a broadly Kripkean approach, in which—in particular—inter-world identity is admitted, a single entity is allowed to exist at more than one possible world, and the extensions of predicates vary from world to world.

Given SM and quite independently of CAI, also the extensions of parthood and composition (as those of any other standard property or relation) should be relativized to possible worlds, in order to avoid incurring violations of the

Indiscernibility of Identicals, when a single entity happens to have different parts at different worlds.

We will show that, once SM and the ensuing relativization of parthood and composition to possible worlds are considered, the picture of the relation between CAI and NI turns out to be more complex than the picture given by the current literature. In particular, in order to give a minimal, initial plausibility to CAI's claim that world-relative composition is a kind of identity, also identity should be relativized to worlds. Otherwise, the various relativized forms of composition would collapse on a single, non-relativized form of absolute identity, thereby losing track of the relativization to worlds which was needed in order to preserve the Indiscernibility of Identicals.

Therefore, in order to make sense of CAI in the context of SM, we need to introduce a world-relative variety of identity and we obtain a form of CAI not yet discussed in the literature, namely the thesis that composition is a world-relative, contingent form of identity. We label this new variety of CAI *Contingent Composition as Identity* (CCAI). If CCAI is adopted, we end up needing two distinct notions of identity: absolute identity is needed in order to express the constitutive thesis of SM, according to which there is inter-world identity; non-absolute, world-relative identity is needed in the formulation of CCAI.

Three main features differentiate absolute identity from non-absolute, world-relative identity which the backer of CAI is forced to introduce in the context of SM. On the one hand, standard identity:

- a1) is not relative to a possible world;
- a2) holds necessarily (NI);
- a3) is governed by an unrestricted principle of indiscernibility.

On the other hand, the kind of world-relative identity which characterizes CCAI:

- b1) is relative to a possible world;
- b2) does not hold necessarily;
- b3) is governed—at best—by a restricted principle of indiscernibility, such as the following (World-Relative Indiscernibility of World-Relative Identicals) ( $=_w$  is world-relative identity,  $P_w$  is a second-order variable for world-relative properties):

$$\forall w \forall x x \forall y y (x x =_w y y \rightarrow \forall P_w (P_w x x \leftrightarrow P_w y y))$$

We will show that these differences between standard, absolute identity and the world-relative kind of identity which is employed in CCAI determine—so to say—both the virtue and the limit of CCAI.

The *virtue* is that CCAI preserves the contingency of composition and of all the other mereological relations, without being revisionary about the necessity of standard identity (NI). The *limit* is that world-relative identity is introduced in a suspiciously *ad hoc* way. Indeed, there is no reason to think that there is a world- relative kind of identity outside of CCAI: the extension of world-relative identity at each world will exclusively include what CCAI is about.

As a consequence, to say that any instance of composition is an instance of world-relative identity is not to assimilate composition to an already known and independently characterized relation. CCAI is not the claim that composition is something else, already known. A foe of CAI might even insinuate that “world-relative identity” is simply a new name for world-relative composition.

We will argue that this foe of CAI is too negative. Admittedly, CCAI ends up assimilating composition to a *specific* form of identity, of which there are no other cases except those postulated by CCAI itself. Indeed, CCAI does not consist in seeing composition as a subcase of a preexisting or more general notion of identity. As a consequence, CCAI is incompatible with so-called strong CAI, namely the radical version of CAI endorsed in various forms by Bøhn (2014), Cotnoir (2013), Wallace (2011), according to which any instance of composition is strictly speaking an instance of standard, absolute identity.

Nonetheless, CCAI might be seen as an interesting development of the so-called weak variety of CAI, originally endorsed by Lewis (1991), according to which composition is not strictly speaking identity, but merely analogous to it. CCAI indeed pinpoints a rather precise and interesting analogy between composition and standard identity. Namely, while standard identity obeys an unrestricted principle of indiscernibility (see a3) above), composition—according to CCAI—obeys a restricted, but still quite comprehensive and highly ambitious principle of indiscernibility, namely (World-Relative Indiscernibility of World-Relative Identicals, b3) above), which quantifies over every world-relativized property.

The result that composition obeys a quite strong indiscernibility principle such as (World-Relative Indiscernibility of World-Relative Identicals) might be an additional reason to think that composition is indeed analogous (as weak CAI contends) to standard identity. Thus, the foe of CAI would be overhasty to conclude that CCAI is devoid of any content or that CCAI is a mere way of rebranding composition as identity, without any philosophical content. By contrast, we will show that CCAI is an interesting development of weak CAI.

It is important to remark that in our analysis, we do not presuppose that the non-standard treatments of modal notions (counterpart theory and the theory of modal parts) adopted in Borghini (2005) and Wallace (2014) are false. Maybe one of them is true, and even jointly true with strong CAI. However, if the purpose of the debate is to assess CAI from a modal viewpoint, it is inadvisable with the usual pattern of the debate on CAI to adopt a non-standard view of modality as a background. Moreover, if we adopt SM, we end up formulating an interesting and so far never discussed variety of weak CAI: CCAI.

## References

- Bøhn, E. (2014). Unrestricted Composition as Identity. In A. Cotnoir & D. Baxter (Eds.), *Composition as Identity* (pp. 143–165). Oxford: OUP.
- Borghini, A. (2005). Counterpart Theory Vindicated: A Reply to Merricks. *Dialectica*, 59(1):67–73.

- Carrara, M. & Lando G. (2017). Composition and Relative Counting. *Dialectica*, 71(4):489–529.
- Cotnoir, A. (2013). Composition as General Identity. *Oxford Studies in Metaphysics*, 8:295–322.
- Lando G. (2017). *Mereology. A Philosophical Introduction*. London: Bloomsbury.
- Lewis, D. (1991). *Parts of Classes*. Oxford: Blackwell.
- Merricks, T. (1999). Composition as Identity, Mereological Essentialism, and Counterpart Theory.. *Australasian Journal of Philosophy*, 77(2):192–195.
- van Inwagen, P. (1994). Composition as identity.. *Philosophical Perspectives*, 8:207–220.
- Wallace, M. (2011). Composition as Identity: Part 2.. *Philosophy Compass*, 6:817–827.
- Wallace, M. (2014). Composition as Identity, Modal Parts, and Mereological Essentialism. In A. Cotnoir & D. Baxter (Eds.), *Composition as Identity* (pp. 111–129). Oxford: OUP.



# La argumentación en torno a la desobediencia civil

Álvaro Domínguez Armas

En 1968 se iniciaba en París una revuelta política con especial relevancia debido a sus protagonistas. Los agentes del movimiento social eran jóvenes estudiantes, hijos de clase media-alta que se levantaban contra cualquier forma de autoridad. Sin embargo, el aspecto más relevante de este acto sociopolítico fue lo que el artista Joan Rabascall<sup>1</sup> decía: “en la calle todo el mundo se hablaba, había una necesidad de comunicarse, por la mañana en las esquinas se colgaban carteles dando informaciones y convocatorias, era como si la autoridad hubiera desaparecido, una sensación extraña”.

Sus palabras se quedan, a mi parecer, un poco cortas. La *necesidad* de comunicarse no se resume en la mera interacción entre individuos, sino que debe entenderse de una manera más profunda. Así, es coherente situar que los estudiantes exaltados de la capital francesa necesitaban constantemente *justificar* y *argumentar* su movimiento. Incluso, me atrevería a decir que esta inferencia es aplicable a todos los contextos políticos y que, en definitiva, toda la esfera política está guiada por la argumentación.

Asimismo, considerar este hecho como un acto de *desobediencia civil* es algo pertinente, ya que se trata de un fenómeno político ilegal, público, no violento y consciente, realizado con la intención de frustrar alguna ley o programa de gobierno, pero que *acepta* y *reconoce* el Estado de Derecho<sup>2</sup> no como una construcción acabada, sino como una empresa accidentada, encaminada a establecer o conservar, a renovar o ampliar un ordenamiento jurídico legítimo en circunstancias cambiantes.

Ahora bien, para analizar este fenómeno como una estrategia argumentativa se ha de partir de unas premisas determinadas que, siguiendo al profesor Luis Vega<sup>3</sup>, son las siguientes:

1. La desobediencia civil se dirime dentro del *discurso público*, esto último se plantea como una esfera *accesible* entendida como un medio en el que se desarrolla una interacción comunicativa basada en el *uso experto de la razón* o, en palabras de Vega, de dar cuenta y razón de algo a alguien o ante alguien con el fin de justificar las propuestas y lograr su asentimiento.

---

<sup>1</sup>(cf.) Playà Maset, J. “50 años de las revueltas de París, Praga y México” en La Vanguardia, 29 de abril de 2018

<sup>2</sup>(cf.) Habermas, J. (1985), *Ensayos políticos*, Ramón García Cotarelo (trad.), Barcelona, Ediciones Península, 1988, p. 57 y ss.

<sup>3</sup>Vega, L. (2013) *La fauna de las falacias*, Madrid, Trotta, p. 120 y (cf.) Vega, L. (2008) “Deliberación y discurso civil: nuevas perspectivas en el campo de la argumentación” en *Revista Laguna* nº 22, pp. 35–51, p. 38, cit. en Marraud, H. “La lógica del discurso civil” en Marraud, H. y Olmos, P., (2015) *De la demostración a la argumentación: ensayos en honor de Luis Vega*, p. 163.

2. Su análisis se desarrollará a partir de la *lógica del discurso civil* que tiene en cuenta las creencias, actitudes, valores y propósitos que dan sentido a los intercambios argumentativos y se interesa por cuestiones prácticas, siempre buscando la calidad de la comunicación intersubjetiva.
3. Por lo tanto, se categoriza como la conclusión respecto a la que se propone una argumentación práctica que toma la forma de deliberación colectiva. Como tal conclusión, no adopta, sin embargo, la forma de una *proposición* (susceptible de atribución veritativa) sino la de una *propuesta* (Vega, Kock) susceptible de ser evaluada como más o menos conveniente. Se entiende, además, que los participantes en el tipo de deliberación pública al que nos referimos deberán comparar y ponderar razones y argumentos ya que es posible que existan *buenas razones* a favor y en contra de la propuesta.

Asimismo, siguiendo a Dima Mohammed<sup>4</sup>, en vista de la consideración argumentativa de la desobediencia civil como un acto *estratégico* y con unos objetivos determinados, los interlocutores construyen sus elecciones argumentativas acorde a los fines imperantes englobados en dos grandes grupos:

- a. Fines *intrínsecos* del acto de argumentar; aquellos que son immanentes a la argumentación: justificar una postura determinada, persuadir racionalmente a un oponente y ser un examen crítico de un estándar establecido.
- b. Fines *extrínsecos* de la interacción argumentativa; se derivan del contexto en el que se desarrolla la argumentación: *usos* del acto de argumentar, es decir, los motivos y cuestiones que llevan al interlocutor a hacer uso de la argumentación y; los *propósitos de la argumentación colectiva*, entendidos como inherentes al tipo de diálogo en el que se encuentran los interlocutores.

En último lugar, para el estudio de la desobediencia civil es necesario ver cómo se concretan y determinan los fines correspondientes a cada una de las categorías en el caso del fenómeno político que vengo describiendo hasta ahora, resultando de la siguiente forma:

1. La *justificación* de la desobediencia civil como un acto sociopolítico que fomenta el desarrollo del Estado de derecho, institución que se concibe como una empresa accidentada e inacabada que necesita de la participación cívica para su desarrollo.
2. La *persuasión racional* de los integrantes del ámbito jurídico y el Estado por parte de los disidentes, para lo que es de especial interés la adhesión de terceros a la causa desobediente. Por lo tanto, la importancia de *convencer racionalmente* se ejemplificará en la suma de agentes a la propuesta disidente.
3. Ser un *examen crítico* a la legalidad jurídica, una *alternativa* al poder político y su ordenamiento legal.
4. Los *propósitos de la argumentación colectiva*, en este caso son los de la *deliberación pública* que busca llegar a un consenso sobre una cuestión

---

<sup>4</sup>Mohammed, D. (2016), "Goals in Argumentation: A proposal for the analysis and evaluation on public political arguments", en *Argumentation* n.º 30(3), pp. 221-245, p. 225

práctica y de interés para toda la comunidad, atendiendo a las propuestas planteadas.

Estas serían, en resumen, las perspectivas pertinentes para el estudio de la justificación argumentativa de la desobediencia civil como propuesta política. Entendiéndola como una acción que, si se quiere justificar, se expresará como una propuesta visualizada como la conclusión de un argumento o argumentación compleja.

# Expressiveness comparisons between logics: an investigation of some intuitions and formal criteria

Diego Fernandes<sup>1</sup>

## Introduction

In the literature, it is common to make comparisons of logics in terms of expressive power. Unfortunately, these comparisons are often made on imprecise grounds and with varying criteria. Now when people are careful enough to define precisely which concept of expressiveness is being used,<sup>2</sup> there is generally no further comment on it, e.g. intuitive motivations or why this particular one was chosen. This gives us reason to think it is not quite known that there can be many *prima facie* plausible but conflicting ways to compare expressiveness of logics.

Indeed, the literature on the subject is scarce. There is a traditional criterion for expressiveness (referred to as  $\preceq_{EC}$ ), issuing from the Lindström theorems [Lin69]. Ebbinghaus presents briefly another one ( $\preceq_{\equiv}$ ) in [Ebb85]. Shapiro questions the adequacy of  $\preceq_{EC}$  in [Sha91] due to its strictness and gives two broader criteria. Recently there was two even wider proposals in [GMV07] and [Kui14b].

The aim of this work is to improve the considerations on the concept of relative expressiveness by selecting some intuitions, comparing putative formal criteria for them, analyzing their material adequacy and possible alternatives. Moreover, it will be shown that the later broader criteria are not adequate.

## What would be a good measure for relative expressive power?

### *Absolute vs relative expressiveness*

Here we consider a logic  $\mathcal{L}$  as a tuple  $(\mathcal{F}, \mathcal{S}, \models_{\mathcal{L}})$  where  $\mathcal{F}$  is a class of sentences,  $\mathcal{S}$  is a class of structures and  $\models_{\mathcal{L}}$  is a consequence relation on  $\mathcal{S} \times \mathcal{F}$ . The main motivation of expressive power comparisons in finite model theory, modal logic and the like is to check what these logics can say about their models. Many such results are then about an *absolute* measure of expressiveness, which consists in

---

<sup>1</sup>Universidade Federal de Goiás. This research is partially supported by CAPES/Brazil

<sup>2</sup>E.g. in [KW99], [Koo07], [vBTCV09], [AFFM11] and [Kui14a].

a certain list of relevant properties expressible in the logic. For example, if they are able to say that a given model is finite, has an even number of elements, if there is a path between any of its elements, etc. Thus we have a first intuition for a measure of expressiveness (an absolute one):

- (E<sub>1</sub>) The expressiveness of a logic  $\mathcal{L}$  is measured through the capacity of  $\mathcal{L}$  to express certain properties.

This capacity of expressing properties can be captured through the following criterion

**Definition 1** (Expressing properties). *A property  $P$  is expressible in  $\mathcal{L}$ , if for all  $\mathcal{L}$ -structures  $\mathfrak{U}$  and  $\mathfrak{B}$ , whenever  $\mathfrak{U}$  has  $P$  and  $\mathfrak{B}$  has not, then there is an  $\mathcal{L}$ -sentence  $\phi$  that can distinguish  $\mathfrak{U}$  and  $\mathfrak{B}$ , e.g.:  $\mathfrak{U} \models_{\mathcal{L}} \phi$  and  $\mathfrak{B} \not\models_{\mathcal{L}} \phi$ .*

Thus, in order to compare the expressiveness of two logics  $\mathcal{L}_1$  and  $\mathcal{L}_2$ , one could select a set  $\mathcal{P}$  of relevant properties, and then  $\mathcal{L}_2$  would be at least as expressive as  $\mathcal{L}_1$ , if every property in  $\mathcal{P}$  expressible in  $\mathcal{L}_1$ , is also expressible in  $\mathcal{L}_2$ . The evident issue with this method of comparison is that it depends on the arbitrary choice of  $\mathcal{P}$ . There are, however, more reasonable ways to obtain truly relative notions of expressiveness, they will be exposed in the sequence.

### *Capacity of distinguishing structures*

In order to obtain a more objective and general method of comparing expressiveness between logics, let us extract and adapt the main intuition used in the above method. Since the capacity of expressing some properties is based on the capacity of a logic to distinguish structures, then *the greater the number of structures distinguished by some formula, the greater properties are expressible*. This could be made into a first intuition for relative expressiveness:

- (E<sub>2</sub>)  $\mathcal{L}_2$  is at least as expressive as  $\mathcal{L}_1$  whenever every structure that can be distinguished by  $\mathcal{L}_1$  can also be distinguished by  $\mathcal{L}_2$ .

Now we can sharpen  $E_2$  straightforwardly into the following criterion (which appears in [Ebb85]), where  $Str[\tau]$  is the class of structures in the signature  $\tau$ , and  $\equiv_{\mathcal{L}}$  means  $\mathcal{L}$ -equivalence of structures:

**Definition 2** ( $\preceq_{\equiv}$ ).  $\mathcal{L}_1 \preceq_{\equiv} \mathcal{L}_2$  iff for all  $\tau$  and all  $\mathfrak{U}, \mathfrak{B} \in Str[\tau]$ , if  $\mathfrak{U} \not\equiv_{\mathcal{L}_1} \mathfrak{B}$  then  $\mathfrak{U} \not\equiv_{\mathcal{L}_2} \mathfrak{B}$ .

It is straightforward to prove that if  $\mathcal{L}_1 \preceq_{\equiv} \mathcal{L}_2$ , then every property expressible in  $\mathcal{L}_1$  is also expressible in  $\mathcal{L}_2$ .

Now it turns out that  $\preceq_{\equiv}$  is not adequate in all circumstances, as it does not capture a nuance that could be interesting when comparing logics. There are cases where logics  $\mathcal{L}_1$  and  $\mathcal{L}_2$  can distinguish the same structures. Nevertheless, some class of structures  $\mathcal{K}$  may be *characterizable* in  $\mathcal{L}_2$ , but not in  $\mathcal{L}_1$ . Let us see what this amounts to in the next sections.

## Capacity of characterizing classes of structures

The former considerations lead us to analyse another intuition for comparing expressiveness of logics, which no longer rests on the capacity of distinguishing structures, which is the basis of  $E_2$ , but on the stricter capacity of *characterizing* classes of structures:

- ( $E_3$ )  $\mathcal{L}_2$  is at least as expressive as  $\mathcal{L}_1$  whenever every class of structures that is characterizable in  $\mathcal{L}_1$  is also characterizable in  $\mathcal{L}_2$ .

Now we have to clarify when a class of structures  $\mathcal{K}$  is *characterizable* in a logic  $\mathcal{L}$ . It will be seen that this concept can be formally captured in many ways.

### Elementary classes

The most straightforward criterion for when a class  $\mathcal{K}$  is characterizable in  $\mathcal{L}$ , is when it is an *elementary class* of  $\mathcal{L}$ . Let  $\text{Mod}_{\mathcal{L}}^{\tau}(\phi)$  be the class of  $\tau$ -structures that satisfy the sentence  $\phi$  in  $\mathcal{L}$ . Then we have that:

**Definition 3** (*EC-class*). A class  $\mathcal{K}$  of  $\tau$ -structures is an elementary class in  $\mathcal{L}$  (for short  $\mathcal{K} \in \text{EC}_{\mathcal{L}}$ ) iff there is an  $\mathcal{L}[\tau]$ -sentence  $\phi$ , such that  $\mathcal{K} = \text{Mod}_{\mathcal{L}}^{\tau}(\phi)$ .<sup>3</sup>

Then one obtains the corresponding sharpening of  $E_3$ , which we call  $\preceq_{EC}$ :

**Definition 4** ( $\preceq_{EC}$ ).  $\mathcal{L}_1 \preceq_{EC} \mathcal{L}_2$  iff  $\text{EC}_{\mathcal{L}_1} \subseteq \text{EC}_{\mathcal{L}_2}$ .

One can easily see that  $\mathcal{L}_1 \preceq_{EC} \mathcal{L}_2$  implies that  $\mathcal{L}_1 \preceq_{\equiv} \mathcal{L}_2$ . As it was mentioned above, it turns out that the reverse implication does not hold. Then a logic  $\mathcal{L}_1$  can be as expressive as  $\mathcal{L}_2$  according to  $\preceq_{\equiv}$  yet  $\mathcal{L}_1$  can be more expressive than  $\mathcal{L}_2$  according to  $\preceq_{EC}$ .

The criterion  $\preceq_{EC}$  was very fruitful, as it has been the basis of many important expressiveness results.<sup>4</sup> Despite this, for some comparisons,  $\preceq_{EC}$  turns out to be very strict and people have argued for a relaxation.

One way to do so, is to come back to  $\preceq_{\equiv}$ , but this move was not followed, at least by the people who complained. A possible reason is that there was a recurrent need to allow for changes in signature in expressiveness comparisons, and such changes were not permitted neither by  $\preceq_{EC}$  nor by  $\preceq_{\equiv}$ . Thus, another concept came into the front, the concept of a projective class. We will present it in a moment, after a motivating example.

### Projective classes

Due to the compactness theorem of first-order logic, (for short,  $\mathcal{FOL}$ ), the class of infinite structures is not characterizable in it. Nevertheless, if we add some

---

<sup>3</sup>Another possibility is to allow  $\Delta$ -elementary classes, i.e. classes  $\mathcal{K}$ , such that  $\mathcal{K} = \text{Mod}_{\mathcal{L}}^{\tau}(\Delta)$ , where  $\Delta$  is a denumerable set of  $\mathcal{L}$ -sentences. This possibility holds also for the two remaining notions of characterizability we examine (mention) in this section, e.g. *PC* and *RPC* (only mentioned). However, for the sake of simplicity, their  $\Delta$ -versions will not be examined here.

<sup>4</sup>One of Lindström's results is that first-order logic has maximal expressiveness among the countably compact logics having a downward Löwenheim-Skolem theorem. For more, see [BF85].

non-logical symbols, it becomes characterizable. Then it is not unreasonable to say that infiniteness is expressible in  $\mathcal{FOL}$  once some minimum resources are available.

The same happens in many other relevant concepts and in other logics as well. Some examples are given in [Sha91], involving the logics  $\mathcal{L}(Q_0)$  and  $\mathcal{L}(A)$ .<sup>5</sup> The class of infinite structures is not elementary in  $\mathcal{L}(A)$ , and the transitive closure of relations is not elementary in  $\mathcal{L}(Q_0)$ . Thus, neither  $\mathcal{L}(Q_0) \preceq_{EC} \mathcal{L}(A)$  nor  $\mathcal{L}(A) \preceq_{EC} \mathcal{L}(Q_0)$ . Concerning this Shapiro says (*ibid*, p. 232): “I would suggest that the ‘non-inclusions’ here are artefacts of an unnatural restriction on the non-logical terminology.” His point is that the referred classes are projective in the respective logics. Also in the case of  $\mathcal{FOL}$  mentioned above, the class of infinite structures is not a *EC*- but a projective class (*PC*-class) of  $\mathcal{FOL}$ :

**Definition 5** (*PC*-class). *A class  $\mathcal{K}$  of  $\tau$ -structures is a projective class in  $\mathcal{L}$  (for short,  $\mathcal{K} \in PC_{\mathcal{L}}$ ) iff there is  $\tau' \supseteq \tau$  and a  $\tau'$ -class  $\mathcal{K}'$  such that  $\mathcal{K} = \{\mathfrak{U} \upharpoonright \tau \mid \mathfrak{U} \in \mathcal{K}'\}$  and  $\mathcal{K}' \in EC_{\mathcal{L}}$ .*

In this way a new sharpening of  $E_3$ , based on *PC*-classes can be given:<sup>6</sup>

**Definition 6** ( $\preceq_{PC}$ ).  $\mathcal{L}_1 \preceq_{PC} \mathcal{L}_2$  iff  $EC_{\mathcal{L}_1} \subseteq PC_{\mathcal{L}_2}$ .

Notice that  $EC_{\mathcal{L}} \subseteq PC_{\mathcal{L}}$ , and thus  $\preceq_{EC} \subseteq \preceq_{PC}$ , the opposite inclusions do not hold for many logics, an exception being second-order logics, where  $EC = PC$ .<sup>7</sup>

As regards the formerly mentioned logics, now we have that both  $\mathcal{L}(Q_0) \preceq_{PC} \mathcal{L}(A)$  and  $\mathcal{L}(A) \preceq_{PC} \mathcal{L}(Q_0)$  [Sha91]. Besides Shapiro, a few others have defended  $\preceq_{PC}$  as a better measure of expressiveness than  $\preceq_{EC}$ . For example, in [Mak80, p. 420], Makowski defends the use of *PC*-classes for expressiveness comparisons and says it could be more natural than the use of *EC*-classes. Recently Bresolin et al. [BMVS16] went on similar lines.

We have seen until now two possible sharpenings of the intuition for expressiveness  $E_3$ :  $\preceq_{EC}$  and  $\preceq_{PC}$ . The last one was proposed as issuing a broader and yet reasonable criterion of relative expressiveness. Notice, however, that it is still possible to obtain an even broader criterion using the notion of a *relative projective class* (*RPC*). This latter possibility will not be treated here as it is still not wide enough for our purposes.

All mentioned intuitions for expressiveness and respective formal criteria require the logics to be defined in the same class of structures. However, this is very restrictive, for example,  $\mathcal{FOL}$  and modal logics could not be compared in terms of expressiveness. Due to this restriction, we say these intuitions and criteria are in the “uni-class framework” of expressiveness.

---

<sup>5</sup> $\mathcal{L}(Q_0)$  is  $\mathcal{FOL}$  incremented with the quantifier “for infinitely many” and  $\mathcal{L}(A)$  is  $\mathcal{FOL}$  incremented with the ancestor operator  $A$ .

<sup>6</sup>The relation  $\mathcal{L}_1 \preceq_{PC} \mathcal{L}_2$  is defined in [Ebb85, p. 43] as  $PC_{\mathcal{L}_1} \subseteq PC_{\mathcal{L}_2}$ , rather than  $EC_{\mathcal{L}_1} \subseteq PC_{\mathcal{L}_2}$ . However they can be shown to be equivalent, which is not the case for  $\preceq_{RPC}$ , see footnote below.

<sup>7</sup>The reason for this is that whenever the characterization of a class of structures needs additional relation symbols  $R_1, \dots, R_n$ , one could replace them by relation variables of the same arity and existentially quantify over them.

## Allowing translations of structures and formulas: still broader criteria

To overcome this limitation, some criteria were proposed in what we will call a “multi-class framework” of expressiveness, so as to allow the comparison of logics defined on different classes of structures. For this, we could use an intuition related to  $E_3$ , but much more flexible:

( $E_4$ )  $\mathcal{L}_2$  is at least as expressive as  $\mathcal{L}_1$  whenever for every  $\mathcal{L}_1$ -sentence  $\phi$ , there is an  $\mathcal{L}_2$ -sentence  $\psi$  having an equivalent meaning.

Now the task of obtaining a formal sharpening of  $E_4$  in the multi-class framework is much more complex and new tools are called for. One must employ a translation  $f : \mathcal{S} \rightarrow \mathcal{S}'$  of  $\mathcal{L}$ -structures into  $\mathcal{L}'$ -structures or vice-versa; and a translation  $\mathcal{T} : \mathcal{F} \rightarrow \mathcal{F}'$  of  $\mathcal{L}$ -formulas into  $\mathcal{L}'$ -formulas, or vice-versa.

Thus, now the weight goes on the notion of translation  $(f, \mathcal{T}) : \mathcal{L} \rightarrow \mathcal{L}'$ . For an  $\mathcal{L}$ -structure  $\mathfrak{U}$  and  $\mathcal{L}$ -sentence  $\phi$ , it has to be guaranteed that e.g.  $\phi$  stands to  $\mathfrak{U}$  in the source logic  $\mathcal{L}$  as  $\mathcal{T}(\phi)$  stands to  $f(\mathfrak{U})$  in the target logic  $\mathcal{L}'$ .

The problem is that it is very difficult to establish the nature of the congruence between the pairs  $(\mathfrak{U}, \phi)$  and  $(f(\mathfrak{U}), \mathcal{T}(\phi))$ . Basing it only on satisfaction ( $\mathfrak{U} \models_1 \phi$  iff  $f(\mathfrak{U}) \models_2 \mathcal{T}(\phi)$ ) is far away from being sufficient, since we can always devise translation functions such that if  $\mathfrak{U}$  satisfies  $\phi$ , then  $f(\mathfrak{U})$  satisfies  $\mathcal{T}(\phi)$ , no matter how far the source and target structures are from each other. On the other hand, imposing requirements on  $(f, \mathcal{T})$  is a complex enterprise, because either it under-generates or, by a little breach, it over-generates.

Some proposals for a multi-class expressiveness/sub-logic relation were made by J. Väänänen & M. GarcMatts [GMV07] ( $\preceq_{vg}$ ) and L. Kuijer [Kui14b] ( $\preceq_k$ ). Unfortunately, though imposing different conditions, it can be shown that both over-generate.

- Using  $\preceq_{vg}$ , weak classical propositional logic<sup>8</sup> ( $\mathcal{WPL}$ ) and classical propositional logic ( $\mathcal{CPL}$ ) turn out to be one sub-logic of the other. This is not plausible since they are intuitively different logics.
- Using  $\preceq_k$ , classical propositional logic would be as expressive as a propositional relevance logic<sup>9</sup> and the classical propositional logic would be as expressive as the modal logic KT.

## Conclusion

The criteria  $\preceq_{\equiv}$  and  $\preceq_{EC}$  are the main criteria for expressiveness used in the current literature on extended logics, fragments of first-order logic, modal logics

<sup>8</sup>For  $\mathcal{WPL}$  see [Hum05].

<sup>9</sup>The logic referred to here is the system  $\mathcal{R}$  described in [Eps13].



and the like. As we noticed above, they are strict criteria, not allowing e.g. changes of signature. Because of this, we can expect people to end up making small tacit relaxations to allow a bigger range for comparing logics, thus generating misleading results and confusion in the literature.

We can note an apparent early sign of confusion between the two criteria for expressiveness in Barwise’s seminal paper on abstract model theory [Bar74, p. 260]. At some point the notion of “strength” (which we take to be a synonym for “expressiveness”) is defined in terms of  $\preceq_{EC}$ . Afterwards, commenting on the  $\Delta$ -closure operation<sup>10</sup> he says that the logic  $\Delta(\mathcal{L})$  is as strong as  $\mathcal{L}$ , apparently due to Makowski et al.’s result that  $\mathcal{L} \preceq_{PC} \Delta(\mathcal{L})$  and vice-versa [MSS76, p. 166].

A recent example appears in [AFFM11], where  $\preceq_{EC}$  is explicitly declared to be the criterion used for relative expressiveness, but towards the end of the paper (p. 307), the criterion is relaxed, to accommodate a small change of signature. The problem is that the resulting laxer criterion turns out to be essentially  $\preceq_{PC}$ , and this can be misleading since we know that  $\preceq_{EC}$  and  $\preceq_{PC}$  usually give very different results on the relative expressiveness. Yet other recent examples can be found in [DKV16] and [Yan13], that treat  $\preceq_{EC}$ - and  $\preceq_{PC}$ -results from [Vää11], in an indiscriminate manner, as expressiveness results.

The intuition behind  $\preceq_{PC}$  is that allowing the introduction of new non-logical symbols may help to bring about the latent expressive power of a logic, so to speak. The expressiveness of a logic is a certain mixture of logical and non-logical machinery. Sometimes the former aspect is so powerful that one barely needs non-logical symbols for certain tasks.<sup>11</sup> Then it seems that  $\preceq_{PC}$  is a good measure for expressiveness. It remains however to investigate whether it is not an arbitrary stopping point in the quest for wider criteria, after all, we could use relative projective classes as well, and so on.

Now if we allow functions  $(f, \mathcal{T})$  translating formulas and structures, we would obtain an even wider criterion. However, it is very difficult to find the right requirements on  $(f, \mathcal{T})$ , because either they under-generate or, by a little breach, they over-generate. As it was mentioned above, the later happens with  $\preceq_{vg}$  and  $\preceq_k$ .

## References

- [AFFM11] Carlos Areces, Diego Figueira, Santiago Figueira, and Sergio Mera. The expressive power of memory logics. *The Review of Symbolic Logic*, 4(2):290–318, 2011.
- [BF85] Jon Barwise and Solomon Feferman, editors. *Model-theoretic logics*. Perspectives in mathematical logic. Springer-Verlag, 1985.
- [BMVS16] D. Bresolin, E. MuVelasco, and G. Sciavicco. On the expressive power of sub-propositional fragments of modal logic. In D. Cantone and G. Del-

---

<sup>10</sup>The  $\Delta$ -closure is an operation taking a logic  $\mathcal{L}$  and generating another. In  $\Delta(\mathcal{L})$ , whenever  $\mathcal{K}$  and  $\bar{\mathcal{K}}$  are PC-classes, then one of them is elementary. For more see [MSS76, p. 166].

<sup>11</sup>For example, as in  $\mathcal{L}(Q_0)$  the concept of infinitely many is embedded in a logical constant whereas in both  $\mathcal{L}(A)$  and  $\mathcal{FOL}$ , it must be “constructed” with the help of non-logical symbols.

- zanno, editors, *Proceedings of the Seventh Symposium on Games, Automata, Logics and Formal Verification*, pages 91–104, 2016.
- [Ebb85] H.D. Ebbinghaus. Extended logics: The general framework. In Jon Barwise and Solomon Feferman, editors, *Model-theoretic logics*, Perspectives in mathematical logic. Springer-Verlag, 1985.
- [Eps13] R.L. Epstein. *The Semantic Foundations of Logic Volume 1: Propositional Logics*. Nijhoff International Philosophy Series. Springer Netherlands, 2013. Contributors: Carnielli, W. and d’Ottaviano, I.M. and Krajewski, S. and Maddux, R.D.
- [GMV07] Marta García-Matos and Jouko Väänänen. Abstract model theory as a framework for universal logic. In Jean-Yves Beziau, editor, *Logica Universalis*, pages 19–33. Birkhäuser Basel, 2007.
- [Hum05] Lloyd Humberstone. Bau’s translation paradox. *Theoria*, 71(2):138–181, 2005.
- [Koo07] Barteld Kooi. Expressivity and completeness for public update logics via reduction axioms. *Journal of Applied Non-Classical Logics*, 17(2):231–253, 2007.
- [Kui14a] Louwe B. Kuijer. The expressivity of factual change in dynamic epistemic logic. *The Review of Symbolic Logic*, 7(2):208–221, 2014.
- [Kui14b] Louwe Bart Kuijer. *Expressivity of Logics of Knowledge and Action*. PhD thesis, University of Groningen, Groningen, Netherlands, 2014. Available at [http://www.rug.nl/research/portal/files/15492763/Complete\\_dissertation.pdf](http://www.rug.nl/research/portal/files/15492763/Complete_dissertation.pdf)
- [KW99] Marcus Kracht and Frank Wolter. Normal monomodal logics can simulate all others. *Journal of Symbolic Logic*, 64(1):99–138, 1999.
- [Lin69] P. Lindström. On extensions of elementary logic. *Theoria*, 35(1):1–11, 1969.
- [Mak80] J. A. Makowsky. Measuring the expressive power of dynamic logics: An application of abstract model theory. In Jaco de Bakker and Jan van Leeuwen, editors, *Automata, Languages and Programming: Seventh Colloquium Noordwijkerhout, the Netherlands July 14–18, 1980*, pages 409–421. Springer Berlin Heidelberg, Berlin, Heidelberg, 1980.
- [Sha91] S. Shapiro. *Foundations without Foundationalism: A Case for Second-Order Logic*. Oxford Logic Guides. Clarendon Press, 1991.
- [vBTCV09] J. van Benthem, B. Ten Cate, and J. Väänänen. Lindström theorems for fragments of first-order logic. *Logical Methods in Computer Science*, 5(3), 2009.

# Cat, cats, catses and beyond: In defence of higher-level plural logic

Berta Grimau  
USC

Plural Logic (hereafter, PL) is a powerful tool with several applications in various areas of philosophy and the foundations of mathematics. It is an extension of first-order logic with plural terms, plural quantifiers and a relation of plural membership. Its real advantage arises when its language is given a pluralist interpretation, that is, when plural terms are interpreted at face value, as denoting more than one object at once. The key consequence of this reading is that it deems PL ontologically innocent: its plural quantifiers do not require a domain of their own, but range over the first-order domain of quantification, albeit in a distinctive way, i.e. plurally. Given that PL is equi-interpretable with Monadic Second-Order Logic, it gives us its expressive power at the low ontological cost of First-Order Logic.

Some authors believe that PL can be extended into an even more expressive language, Higher-Level Plural Logic (hereafter, HLPL), by adding higher-level plural terms and quantifiers to it.<sup>1</sup> The idea behind higher-level plural reference is that second-level plurals do not stand to plurals like predicates, but rather like plurals stand to singulars (analogously for higher levels). In other words, second-level plurals are pluralized plurals. If English allowed for the iteration of the plural suffix, ‘catses’ would be the plural of ‘cats’ and the second-level plural of ‘cat’. Intuitively, just as lists of singular terms (‘Venus and Serena’) are plural terms, lists of plural terms (‘the philosophers and the historians’) are second-level plurals. Thus just as plural terms denote many objects at once, second-level plural terms denote many *manys* at once. Put in a somewhat misleading manner, but in proper English, second-level plural terms denote many pluralities of objects at once.

Allegedly, HLPL enjoys the expressive power of type theory while, again, committing us to nothing more than the austere ontology of a first-order language. Were this really the case, HLPL would be an extremely useful tool, with various applications in the philosophy of mathematics and in metaphysics. For example, it would let us go beyond the monadic fragment of Second-Order Logic – i.e. by modelling ordered pairs as higher-level pluralities. This, in turn, could open the door to a strengthened development of the neo-Fregean programme (i.e. substituting Second-Order Logic with Third-Level Plural Logic). Another application would be in a nominalist account of mathematics – roughly speaking, by taking talk of sets as elliptical for plural talk, talk of sets of sets as elliptical for second-level plural talk, and so on.<sup>2</sup>

---

<sup>1</sup>See Rayo (2006) and Oliver and Smiley (2016, ch. 15).

<sup>2</sup>See, for example, the higher-level plural development of Cantorian set theory in Oliver and Smiley (2016).

However, while the notions of plural reference and quantification enjoy widespread acceptance today, their higher-level counterparts have been received with scepticism. The fact that the notion of HLP reference has been looked at with apprehension should not come as a surprise; were it legitimate, it could help us settle some big philosophical debates.

The main objection raised against higher-level plural reference is that it is inherently unintelligible, that it is not possible for us to make sense of such alleged linguistic device. This has been argued on two bases: (1) that there are no higher-level plurals in natural language and (2) that the apparent higher-level plurals found in natural languages can always be eliminated. In turn the latter has been argued in two different ways: (2.1) that they can be paraphrased away and (2.2) that they can be semantically analysed away.

In this talk, I will argue for the legitimacy of the notion of higher-level plural reference (and thus of a formal language which makes use of it) on three different grounds. Firstly, I will show that some natural languages (Finnish, Breton and Khamtanga, among others) clearly contain these expressions and thus that the claim to the contrary is likely to arise from the Anglophone framework in which the debate around higher-level plurals takes place. Secondly, I will show that higher-level plurals are not in general paraphrasable away by giving examples of sentences involving higher-level plurals and showing that they cannot be paraphrased away in favour of merely plural or singular expressions. Finally, I will argue that higher-level plurals are not easily analysable away at the level of the semantics. Usually, one would attempt to do so by arguing that higher-level plural reference is really plural reference to complex objects, such as sets. But I will show that this faces some difficulties. The main argument to this effect will turn on the fact that statements which are apparently about pluralities which are too big to form sets cannot be analysed in the usual set-theoretic fashion. This is a common move in the defence of plural reference and I will show that it can be transferred to the debate around higher-level plural reference.

By means of these three lines of argumentation, I will try to convince the audience that higher-level plural reference is an acceptable linguistic apparatus and that it can legitimately be incorporated into a formal language. A formal language which, given its promising applications, should be adopted as part of the philosophical toolkit. My arguments will be especially convincing for those who advocate plural reference and PL in the first place, since the very same reasons typically advanced to defend plurals lead to the acceptance of higher-level plurals. Thus, an underlying topic of my talk will be that the advocate of plurals who dislikes higher-level plurals is in an awkward position.

## Bibliografia

- Oliver, A., & Smiley, T. (2016). *Plural logic: Second edition, revised and enlarged*. Oxford University Press.
- Rayo, A. (2006). Beyond plurals. In A. Rayo & G. Uzquiano (Eds.), *Absolute generality* (pp. 220–54). Oxford University Press.

# Las lógicas híbridas como herramienta formal para el desarrollo de sistemas lógicos temporal-epistémicos

**José Rafael Herrera González**  
rahego@ull.es  
Grupo de Investigación LEMA  
Universidad de La Laguna<sup>1</sup>

## Presentación

Hace ya bastante tiempo que se ha puesto de manifiesto el interés que presentan los sistemas que combinan distintos tipos de lógicas<sup>2</sup>, en tanto que permiten ampliar el universo del discurso al que se refieren nuestros desarrollos formales y, al mismo tiempo, nos facilitan la tarea de desarrollar sistemas lógicos cuyo comportamiento formal coincide, con bastante aproximación, con lo que podemos esperar de acuerdo con nuestras intuiciones mejor fundadas. En particular, se han subrayado las ventajas que resultan, en el sentido anteriormente apuntado, de combinar distintos sistemas de lógica modal.

En nuestra presentación nos centraremos en la combinación de sistemas de lógica modal temporal y epistémica (o doxástica), por lo que los sistemas lógicos resultantes nos van a permitir hacer referencia a cómo varían a lo largo del tiempo los estados epistémicos de los sujetos cognoscentes y, al mismo tiempo, a los conocimientos y creencias que dichos sujetos pueden albergar acerca del presente, el pasado y el futuro.

Aunque se han propuesto sistemas formales de lógica modal que combinan operadores temporales y epistémicos (o doxásticos), algunos de ellos con importantes aportaciones, dichos desarrollos suelen presentar serias limitaciones en cuanto a su poder expresivo y a la satisfacción de propiedades metalógicas tan fundamentales como la consistencia y la completud. Desde nuestro punto de vista, las principales dificultades que tenemos que afrontar a la hora de combinar sistemas de lógica modal epistémica y temporal derivan del hecho de tener que integrar en un mismo esquema formal un punto de vista temporal absoluto (pues generalmente los instantes de tiempo se determinan en función de un observador situado “fuera del mundo”) y un punto de vista epistémico (o doxástico) relativo

---

<sup>1</sup>Este trabajo ha sido realizado en el marco del proyecto *Puntos de vista, disposiciones y tiempo. Perspectivas en un mundo de disposiciones* (FFI2014-57409-R), financiado por el Ministerio de Economía y Competitividad del Gobierno de España).

<sup>2</sup>Cf. VVAA (1996): *Notre Dame Journal of Formal Logic*, vol. 37, no 2; Blackburn, P. & Rijke, M. de (1997); Sernadas, A. & Sernadas, C. (2004); Kurucz, A. (2007); Carnielli, W. & Coniglio, M. E. (2016).

a cada sujeto epistémico (o conjunto de ellos), en tanto que los conocimientos o creencias son atribuidos a sujetos situados en un momento determinado del flujo temporal lineal (pues recurriremos aquí únicamente a sistemas de lógica temporal para tiempo lineal).

Partiendo de este análisis, mostraremos cómo recurriendo a las denominadas *lógicas híbridas* es posible simplificar, de forma notable, la combinación de sistemas modales temporales y epistémicos, con lo cual podemos evitar el tener que construir modelos semánticos excesivamente complicados y antiintuitivos.

Las lógicas híbridas son lógicas modales que nos permiten hacer referencia, de forma precisa y concreta, a estados (o mundos posibles) del modelo<sup>3</sup>; y en el caso de la lógica temporal nos permiten referirnos a un instante particular<sup>4</sup>, lo que sin duda va a acarrear importantes ventajas a la hora de articular sistemas lógicos temporales-epistémicos. Con ello mejora notablemente el poder expresivo de los lenguajes formales empleados y se facilita la configuración de sistemas que formalicen diversos tipos de conocimiento, creencia y tiempo.

En nuestra presentación, tras hacer referencia a las principales aportaciones que se han realizado en materia de combinación de sistemas modales temporales-epistémicos, analizaremos en profundidad el origen y tipos de dificultades que se nos presentan cuando pretendemos desarrollar este tipo de sistemas lógicos. Finalmente, mostraremos las ventajas que acarrea el recurrir a las lógicas híbridas para configurar dichos sistemas. Asimismo, haremos referencia a los aspectos ontológicos y epistemológicos de los desarrollos formales presentados, con el fin de mostrar cómo las nociones de conocimiento, creencia y tiempo formalizados con nuestras propuestas, resultan muy próximos a las concepciones de dichas nociones que solemos mantener en multitud de contextos relevantes para la reflexión filosófica.

## Algunos precedentes destacables

Al abordar el estudio formal del conocimiento y la creencia resulta interesante poder dar cuenta de los cambios que pueden producirse en un sistema de conocimientos y creencias a lo largo del tiempo, pues una de las características principales del conocimiento y las creencias es que son susceptibles de modificarse, ya sea porque se dispone de nueva información, porque se reelabora la ya disponible, porque unas creencias son refutadas y abandonadas o sustituidas por otras, porque la capacidad de memoria de los sujetos es limitada y olvidan algunos de sus conocimientos, etc. Para conseguir esto es necesario enriquecer la capacidad expresiva de la lógica epistémica mediante la adición de una dimensión temporal. La combinación de distintas lógicas de una dimensión para dar lugar a sistemas bidimensionales constituye una de las temáticas más interesantes y abordadas en la reciente investigación lógica. En el caso que nos ocupa, el de la combinación de sistemas de lógica temporal con sistemas de lógica epistémica, se han dado algunos pasos importantes, pero continúan abiertos

---

<sup>3</sup>Blackburn, P. & Seligman, J. (1998); Blackburn, P. (2000); Areces, C., Blackburn, P. & Marx, M. (2001).

<sup>4</sup>Blackburn, P. (1994).

numerosos problemas de cuya resolución depende la posibilidad de configurar sistemas temporales-epistémicos consistentes y completos que recojan las propiedades formales del conocimiento y la creencia en relación al tiempo de una manera satisfactoria.

En este apartado se analizan algunas de las aportaciones más interesantes que a este respecto se han realizado hasta ahora:

1. El *Sistema temporal-epistémico mínimo de Engelfriet*<sup>5</sup>, que en realidad constituye una temporalización de la lógica epistémica, y no una auténtica combinación, pues en él sólo se permite la ocurrencia de operadores epistémicos bajo el alcance de operadores temporales, pero no lo contrario, con lo cual en esta lógica podemos expresar cómo varía el conocimiento en función del tiempo, pero no los conocimientos de que disponen los sujetos en relación al pasado o al futuro.
2. *Kraus y Lehmann*<sup>6</sup>, por su parte, llevan a cabo una extensión temporal de su sistema para el conocimiento y la creencia, distinguiendo entre distintos tipos de creencias según sean más o menos centrales para los agentes y, en función de ello, puedan ser abandonadas con menor o mayor facilidad. Pero Kraus y Lehmann no nos proporcionan una semántica adecuada para el sistema temporal-epistémico que proponen.
3. *Fagin y Halpern*<sup>7</sup>, en cambio, sí consiguen desarrollar un modelo semántico apropiado para la extensión temporal de su lógica del conocimiento consciente, lo cual permite formalizar la adquisición y la pérdida o modificación de conocimientos y creencias por parte de los agentes.
4. En *Herrera y Vázquez* (1999) se presenta un modelo que permite combinar el Sistema S5 de la lógica epistémica o el Sistema KD45 de la lógica doxástica con un sistema de lógica temporal para tiempo lineal<sup>8</sup>. Lo destacable de esta propuesta es que permite la ocurrencia sin restricciones de operadores epistémicos bajo el alcance de operadores temporales y viceversa, lo cual posibilita expresar los conocimientos y creencias que los sujetos epistémicos albergan acerca del presente, el pasado o el futuro, así como la forma en que dichos conocimientos y creencias van variando a lo largo del tiempo.

## Sistemas híbridos temporal-epistémicos

Como se muestra en Herrera y Vázquez (2011), los principales obstáculos a la hora de configurar sistemas formales de lógica temporal-epistémica de base modal, surgen a la hora de “intentar evaluar las fórmulas en las que operadores temporales aparecen bajo el alcance de operadores epistémicos, es decir, aquellas fórmulas que expresan el conocimiento (o las creencias) que los agentes albergan acerca del pasado o el futuro”<sup>9</sup>. Este tipo de inconvenientes pueden superarse

---

<sup>5</sup>Engelfriet, J. (1996).

<sup>6</sup>Kraus, S. & Lehmann, D. (1988).

<sup>7</sup>Fagin, R. & Halpern, J.Y. (1988).

<sup>8</sup>Herrera, R. y Vázquez, M. (1999).

<sup>9</sup>Herrera, R. y Vázquez, M. (2011), p. 41.

recurriendo a las denominadas *lógicas híbridas*, que nos van a permitir hacer referencia a estados o mundos posibles concretos y, asimismo, a instantes precisos del tiempo.

Para ello distinguimos, a nivel semántico, los diferentes “estados actuales” vinculados a cada instante del tiempo (siendo  $\wp s$  el “estado actual” en el instante  $s$ ,  $\wp t$  el “estado actual” en el instante  $t$ , y así sucesivamente).

Asimismo, debemos tener en cuenta el conjunto de estados o mundos posibles concebidos por un sujeto  $x$  en relación a cualquier instante del presente, el pasado o el futuro:

- a)  $\mu x(t, t)$  representa el conjunto de estados que el sujeto  $x$  considera posibles en el instante  $t$  en relación al presente (es decir, en relación a dicho instante  $t$ ).
- b)  $\mu x(t, s)$  representa el conjunto de estados que  $x$  considera posibles en el instante  $t$  en relación a otro instante  $s$  anterior (pasado) o posterior (futuro):  $s < t$  o  $t < s$ .

Si bien se han propuesto distintas combinaciones de lógica temporal y lógicas híbridas, es mucho menos frecuente incorporar operadores híbridos a una lógica temporal-epistémica, aunque el empleo de lenguajes híbridos simplifica significativamente este tipo de combinaciones, y nos permite articular sistemas temporales-epistémicos de gran capacidad expresiva.

Para lograr este objetivo, definimos  $KF\varphi$  como  $K^{\rightarrow i}\varphi$ ,  $KP\varphi$  como  $K^{\leftarrow i}\varphi$ ,  $MF\varphi$  como  $M^{\rightarrow i}\varphi$  y  $MP\varphi$  como  $M^{\leftarrow i}\varphi$ <sup>10</sup>.

De esta forma, el lenguaje del nuevo sistema híbrido temporal-epistémico incluirá los siguientes operadores:

- a) Los operadores proposicionales clásicos  $\rightarrow \neg \wedge \vee \leftrightarrow$ .
- b) Los operadores temporales  $P$  (“al menos una vez en el pasado”),  $F$  (“al menos una vez en el futuro”),  $H$  (“siempre en el pasado”),  $G$  (“siempre en el futuro”).
- c) Los operadores epistémicos  $K$  (“saber que”),  $M$  (“considerar posible que”).
- d) Los nuevos operadores temporales-epistémicos  $K^{\rightarrow i}$  ( $\equiv KF$ ),  $K^{\leftarrow i}$  ( $\equiv KP$ ),  $M^{\rightarrow i}$  ( $\equiv MF$ ),  $M^{\leftarrow i}$  ( $\equiv MP$ ).

Por otra parte, distinguiremos entre dos tipos de *fórmulas bien formadas* (fbf): *fórmulas bien formadas de tipo 1* (fbf1) y *fórmulas bien formadas de tipo 2* (fbf2), teniendo en cuenta que fbfs son aquellas (y sólo aquellas) que se construyen de acuerdo con las siguientes reglas de formación:

- a) Toda variable proposicional ( $p, q, r \dots$ ) es una fbf1.
- b) Si  $\varphi$  es una fbf1, también lo son  $K\varphi$ ,  $K^{\rightarrow i}\varphi$ ,  $K^{\leftarrow i}\varphi$ ,  $\neg\varphi$ ,  $M^{\rightarrow i}\varphi$ ,  $M^{\leftarrow i}\varphi$ .
- c) Si  $\varphi$  y  $\psi$  son fbf1, también lo son  $(\varphi \rightarrow \psi)$ ,  $(\varphi \wedge \psi)$ ,  $(\varphi \vee \psi)$ ,  $(\varphi \leftrightarrow \psi)$

---

<sup>10</sup>Como es habitual,  $KF\varphi$  establece que “el sujeto sabe que (al menos una vez) en el futuro  $\varphi$  será el caso”;  $KP\varphi$  “el sujeto sabe que (al menos una vez) en el pasado  $\varphi$  fue el caso”;  $MF\varphi$  “el sujeto considera posible que (al menos una vez) en el futuro  $\varphi$  será el caso”; y  $MP\varphi$  “el sujeto considera posible que (al menos una vez) en el pasado  $\varphi$  fue el caso”.



- d) Si  $\varphi$  es una fb1,  $\varphi$  es una fb2.
- e) Si  $\varphi$  es una fb2, también lo son  $\neg\varphi$ ,  $P\varphi$ ,  $H\varphi$ ,  $F\varphi$ ,  $G\varphi$ .
- f) Si  $\varphi$  y  $\psi$  son fb2, también lo son  $(\varphi \rightarrow \psi)$ ,  $(\varphi \wedge \psi)$ ,  $(\varphi \vee \psi)$ ,  $(\varphi \leftrightarrow \psi)$

Los paréntesis externos de las fórmulas pueden ser eliminados.

Definimos un modelo de la lógica temporal-epistémica como una estructura  $M = \langle M, R, <, v \rangle$ , donde:

1.  $M \neq \emptyset$  es un conjunto de “estados”. Un “estado” es un mundo posible “situado” en un instante determinado.
2.  $R \subseteq M^2$  es una relación de accesibilidad entre estados, la cual es reflexiva, simétrica y transitiva (relación de equivalencia).
3.  $< \subseteq M^2$  es una relación de ulterioridad (o relación antes/después) entre estados, la cual es irreflexiva.
4.  $v$  es una función de evaluación que asigna el valor verdadero o falso a cada fb en un determinado estado. Escribiremos  $v(\varphi, m) = 1$  si  $v$  asigna el valor de verdad 1 a la fórmula  $\varphi$  en el estado  $m$ , y  $v(\varphi, m) = 0$  si le asigna el valor 0. Para cualquier nominal  $i$  hay un subconjunto unitario de  $M$ ,  $\{m\} \subseteq M$  tal que  $v(i, m) = 1$ ; para cualquier otro  $m' \in M$ ,  $v(i, m') = 0$ . La función de evaluación  $v$  cumple las siguientes condiciones, para cualquier estado  $m$ , nominal  $i$ , variable proposicional  $p$  y fbs  $\varphi$  y  $\psi$ :

- i)  $v(p, m) = 1$  o  $v(p, m) = 0$
- ii)  $v(\neg\varphi, m) = 1$  syss  $v(\varphi, m) = 0$
- iii)  $v(\varphi \rightarrow \psi) = 1$  syss  $v(\varphi, m) = 0$  o  $v(\psi, m) = 1$
- iv)  $v(K\varphi, m) = 1$  syss  $\forall n \in M/mRn, v(\varphi, n) = 1$
- v)  $v(F\varphi, m) = 1$  syss  $\exists n \in M/m < n$  y  $v(\varphi, n) = 1$
- vi)  $v(P\varphi, m) = 1$  syss  $\exists n \in M/n < m$  y  $v(\varphi, n) = 1$
- vii)  $v(M^{\rightarrow i}\varphi, m) = 1$  syss  $\exists n \in M/mRn$  y  $\exists o/n < o$  y  $v(i \wedge \varphi, o) = 1$ <sup>11</sup>
- viii)  $v(M^{\leftarrow i}\varphi, m) = 1$  syss  $\exists n \in M/mRn$  y  $\exists o/o < n$  y  $v(i \wedge \varphi, o) = 1$
- ix)  $v(K^{\rightarrow i}\varphi, m) = 1$  syss  $\forall n \in M/mRn, \exists o/n < o$  y  $v(i \wedge \varphi, o) = 1$
- x)  $v(K^{\leftarrow i}\varphi, m) = 1$  syss  $\forall n \in M/mRn, \exists o/o < n$  y  $v(i \wedge \varphi, o) = 1$

Una fórmula  $\varphi$  es válida ( $\models \varphi$ ) si, para todo modelo  $M$  y todo estado  $m \in M$ ,  $v(\varphi, m) = 1$ .

Las cláusulas vii-x anteriores nos permiten evaluar las fórmulas que reflejan los conocimientos de los sujetos acerca del futuro o el pasado. Así, por ejemplo, si un sujeto sabe (en el “estado actual”) que en el futuro  $i$  será el caso que  $\varphi$ , la cláusula ix nos permite comprobar que cualquier estado considerado posible por el sujeto (desde el estado actual) tiene un estado ulterior, que es la denotación de  $i$ , y en el cual  $\varphi$  es verdadero.

<sup>11</sup>  $V(i \wedge \varphi, o) = 1$  expresa que  $o$  es la denotación de  $i$ , es decir, que  $v(i, o) = 1$ .

## Sistemas lógicos temporal-epistémicos: balance final

Podría parecer que la combinación de las modalidades epistémicas-doxásticas y el tiempo en un sistema lógico único constituye una tarea fácilmente realizable, pues ya disponemos de sistemas lógicos, como los sistemas de tiempo indeterminista, en los que se combina tiempo y modalidad (en particular, las modalidades aléticas). Pero la cuestión no es tan sencilla, pues al tratar de combinar la lógica epistémica con la temporal surgen dificultades que no aparecen al combinar tiempo y modalidad (necesidad y posibilidad lógicas). Esto es así porque, al pretender configurar sistemas temporales-epistémicos, nos vemos obligados a combinar lo que podríamos denominar dos puntos de vista distintos: por un lado, un punto de vista temporal absoluto, pues la sucesión de instantes se determina en relación a un observador que se sitúa fuera del mundo. Y, por otro lado, un punto de vista epistémico relativo a cada sujeto (o grupo de sujetos) cognoscente. Esto dificulta en gran medida la labor de desarrollar modelos semánticos para la lógica temporal-epistémica, lo cual no ocurre al combinar tiempo, necesidad y posibilidad, pues en este último caso todas las nociones involucradas se determinan desde un único punto de vista, que podemos calificar de absoluto.

En este trabajo hemos pretendido mostrar que las lógicas híbridas pueden dar un impulso decisivo a las combinaciones de lógica temporal y epistémica. Y es que los nominales, al permitirnos hacer referencia a puntos concretos, nos facilitan poder referirnos a instantes específicos, o mejor aún, a determinados estados “específicamente situados” en el presente, pasado o futuro. Es por ello que vemos aquí abierta una vía de investigación que puede proporcionar resultados provechosos en el tratamiento formal de las nociones epistémicas y doxásticas, así como en el análisis lógico de sus relaciones con la noción de tiempo.

## Bibliografía

- Areces, C., Blackburn, P. & Marx, M. (2001). Hybrid Logics: Characterization, Interpolation and Complexity. *The Journal of Symbolic Logic*, 66(3):977–1009.
- Blackburn, P. (1994). Tense, Temporal Reference and Tense Logic. *Journal of Semantics*, 11:83–101.
- Blackburn, P. & Rijke, M. de (1997). Why combine logics?. *Studia Logica*, 59:5–27.
- Blackburn, P. & Seligman, J. (1998). What are Hybrid Languages?. In Kracht, M., de Rijke, M., Wansing, H. & Zakharyashev, M. (Eds.), *Advances in Modal Logic*, vol. 1. CSLI Publications.
- Blackburn, P. (2000). Representation, Reasoning, and Relational Structures: A Hybrid Logic Manifesto. *Logic Journal of the IGPL*, 8(3):339–365.

- Carnielli, W. & Coniglio, M. E. (2016) Combining Logics. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2016/entries/logic-combining/>>[consulta 01-04-2018].
- Engelfriet, J. (1996). Minimal Temporal Epistemic Logic. *Notre Dame Journal of Formal Logic*, Spring 1996, 37(2):233–259.
- Fagin, R. & Halpern, J.Y. (1988). Belief, Awareness and Limited Reasoning. *Artificial Intelligence*, 34:39–76.
- Herrera, R. y Vázquez, M. (1999). Combinaciones de lógica temporal y epistémica. Actas del Congreso *La filosofía analítica en el cambio de milenio*, Servicio de Publicaciones de la Universidad de Santiago de Compostela, 1999, pp. 55–64.
- Herrera, R. y Vázquez, M. (2011). Hacia una lógica temporal-epistémica basada en lenguajes híbridos. *Análisis Filosófico*, Volumen XXXI, mayo 2011, 1:33–46.
- Kraus, S. & Lehmann, D. (1988). Knowledge, Belief and Time. *Theoretical Computer Science*, 58:155–174.
- Sernadas, A. & Sernadas, C. (2004). Combining Logic Systems: Why, how, what for?. *CIM Bulletin*, 15.
- Kurucz, A. (2007). Combining Modal Logics. En P. Blackburn et al. (Editors), *Handbook of Modal Logic*, Elsevier, pp. 869–924.
- VVAA (1996) *Notre Dame Journal of Formal Logic*, 37(2).

# Identifying logical evidence

**Ben Martin**  
University of Bergen

**Abstract.** With the recognised weakness of historically prevalent account of logical knowledge, in terms of intuition and analyticity, a renewed interest in logical epistemology has been sparked. One recent suggestion is that we come to know logical claims similarly to how scientists know truths about the world, through abduction. However, at present, it's unclear exactly which data logical theories should have to explain, and how we should weight the competing strengths and weaknesses of logical theories. This talk proposes looking at historically important logical disagreements, and extracting from these debates an account of which data successful logical theories must accommodate.

We take ourselves to know certain logical claims, for example that Socrates is wise and just only if he's wise. However, we currently fail to have a viable account of how we possess logical knowledge. Historical attempts to explain this knowledge, such as appeals to intuition and linguistic proficiency, have been found to be ultimately unsatisfactory, either because they are metaphysically obscure or fail to explain logical disagreements (Williamson, 2007). Yet, it's imperative that we have a complete understanding of logical knowledge. While we use logic to form beliefs in all areas of life, such as when testing scientific theories and engaging in rational debate, we now have many competing logics at our disposal to do so, all of which lead us to reasoning differently in certain situations. Yet, in order to make informed decisions about which logics we should use, we require suitable criteria to adjudicate between them, which can only be developed with a full understanding of what constitutes logical evidence. Without such an account of logical evidence, we lack the resources to make principled and holistic decisions about the correct logic to use. Consequently, a new, more complete, explanation of logical knowledge is needed.

In order to supply such an explanation, prominent figures such as Timothy Williamson (2013), Graham Priest (2014) and Ole Hjortland (2017), have recently argued for a new account of logical knowledge, *logical anti-exceptionalism*, which emphasises that such knowledge isn't special in any sense, and that logic's method is akin to that of the natural sciences. Just as science proceeds by advancing theories attempting to best explain the relevant data, by a process known as *abduction*, so logic proposes theories to explain its own domain of data as lucidly and coherently as possible. Thus, we come to be justified in our logical beliefs by recognising which available logical theory best explains the relevant data.

Unfortunately, however, there is little agreement between proponents of *logical anti-exceptionalism* over what constitute these *relevant data* that logical theories must explain, and no clear indication yet of how we should settle the matter

of which data are relevant. But, without a detailed account of what these data are, *logical anti-exceptionalism* cannot hope to provide the means to adjudicate between competing logics, a major motivation for any modern theory of logical epistemology. Thus, we need to know what type of data, exactly, logical theories must explain.

This talk argues that we can look to logical practice for help in both providing support for *logical anti-exceptionalism* and pinpointing the types of data logical theories must explain. While using the practice of researchers has proven a useful method to study how knowledge is acquired in the natural sciences (Burian, 2001) and mathematics (Mancosu, 2008), the same method has yet to be extensively used in the study of logic. Yet, just as philosophers of science have used historical scientific experiments and disputes as their data to infer how we come to know empirical claims, so we can use a *practice-based method* in studying logical knowledge. By taking logical arguments as our data, we can infer from these arguments the methodological principles that logicians rely upon, and the data their theories attempt to explain. The rationale for using practice to inform an epistemology of logic is the presumption that generally, as with scientists, logicians provide suitable reasons for their claims even if, ultimately, they are not wholly satisfactory. Thus, we should expect logicians' arguments to provide insight into how we can come to know logical truths, and the data logical theories must accommodate.

To show the fruitfulness of this *practice-based* approach, the talk considers as a case study arguments from one of the most significant debates in modern logic, the dispute between classical logic and dialetheism over the truth of inconsistent theories. Concentrating particularly on Priest's (2006) initial arguments for dialetheism from the liar and Russell-set paradoxes, and classical replies to the arguments, it's proposed that both Priest and his classical opponents rely upon at least three methodological principles: Firstly, that linguistic and mathematical puzzles, such as the liar sentence and Russell set, can form part of a logical theory's *explanandum*; secondly, that linguistic norms form part of logical evidence, for example in admitting the need to take the meaningfulness of the liar sentences seriously; and thirdly, that mathematical concepts and findings form part of logical evidence, for example by suggesting that only classical logic can underpin mathematical results.

The talk concludes that these initial results from the *practice-based* approach provide both support for *logical anti-exceptionalism*, and details on the types of evidence a logical theory should accommodate. To offer support for their logical views, rather than attempting to settle disputes on purely definitional or intuitional grounds, logicians appeal to their logic's ability to explain certain relevant phenomena, including linguistic norms and findings from mathematics. We suggest that with yet further consideration of important logical disputes, we can hope to build an even fuller picture of logical epistemology and evidence.

## References

- Burian, R. M. (2001). The Dilemma of Case Studies Resolved: The Virtues of Using Case Studies in the History and Philosophy of Science. *Perspectives on Science*, 9:383–404.
- Hjortland, O. T. (2017). Anti-Exceptionalism about Logic. *Philosophical Studies*, 174:631–658.
- Mancosu, P. (2008). *The Philosophy of Mathematical Practice*. Oxford: Oxford University Press.
- Priest, G. (2006). *In Contradiction: A Study of the Transconsistent* (2nd ed.). Oxford: Clarendon Press.
- Priest, G. (2014). Revising logic. In P. Rush (Ed.), *The Metaphysics of Logic* (pp. 211–23). Cambridge: Cambridge University Press.
- Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.
- Williamson, T. (2013). Logic, Metalogic and Neutrality. *Erkenntnis*, 79:211–231.

# Instagram y la pereza argumentativa

Silvia Martínez Fabregat

Con el desarrollo de las TIC la forma en la que concebimos la comunicación ha experimentado cambios que sugieren interesantes cuestiones. En primer lugar, la amplitud del alcance y la diversidad de canales, así como el uso cotidiano de formas comunicativas híbridas entre lo verbal y lo escrito, han suscitado la reflexión acerca de cómo estos nuevos odres contienen nuestras viejas aguas. Desde la Teoría de la Argumentación en concreto, se ha planteado por ejemplo, el análisis de estas formas expresivas facilitadas por las nuevas tecnologías explorando el uso de la imagen en territorios que hasta la fecha parecían ser terreno exclusivo de la palabra sondeando a su vez su capacidad argumentativa tanto como elemento que acompaña y enriquece la argumentación verbal como el hecho de ser actos argumentativos en sí mismas (Groarke et al., 2016). Este debate continúa hoy abierto pues las diferentes posturas enfrentadas se alimentan del recurso visual inagotable que suponen la televisión, el cine, la publicidad y la Red desde donde cada día se generan cientos de nuevos ejemplos abiertos al estudio. Los espacios virtuales de comunicación social como *Twitter*, conforman también un objeto de estudio interesante para la Argumentación entre otras cosas porque su propia naturaleza parece muy adecuada para el análisis en tanto que nos permite rastrear intervenciones, interacciones y el impacto de estas generando imágenes nítidas de su alcance social. Con las herramientas adecuadas estas participaciones sociales pueden quedar retratadas incluyendo ciertos rasgos *ethoicos* del orador (Bamman y Smith, 2015), retóricos (Rosso et al., 2013) y de poder persuasivo que nos arrojan datos que reflejan en lo virtual la realidad de nuestras interacciones. Más allá de suponer un espacio de análisis, también ha generado nuevas formas de interacción. En base a la limitación en el uso de caracteres y la sensación continua de inmediatez, agilidad y actualidad, nos ofrece productos que han renovado la forma de relación entre el orador y su público. En el campo literario encontramos por ejemplo la interesante renovación de los relatos por entregas, sea el caso de la propuesta de Manuel Bartual. En el ámbito argumentativo, la composición de argumentos nutridos por un incontable número de razones aducidas por múltiples oradores; tomemos por ejemplo el caso de la crítica al uso de plásticos del científico ambiental Andreu Escrivá o la denuncia respecto al acoso sexual sufrido por las mujeres compilada a través del *hashtag* “cuéntalo”. Así mismo se ofrecen nuevos espacios dialécticos en los que algunos advierten el futuro de la política en una forma de *e-democracy*, entre otros. La cuestión que nos planteamos es cuál es el papel que puede adoptar la Teoría de la Argumentación respecto al debate en redes, cuáles son las características de esa argumentación virtual y si su uso plantea un renovado interés por las prácticas dialécticas entre el público general.

Un 37% de los 500 millones de *tweets* que registra *Twitter* diariamente, son conversaciones. Valorar cuándo en ellas se producen actos argumentativos ya es un primer escollo para cualquier análisis (Alonso, 2016). Si bien estamos lejos

de poder generar algoritmos que procesen esta ingente cantidad de datos y nos arrojen evaluaciones semánticas de los actos argumentativos que se producen (Alonso et al, 2013) el estudio de casos particulares de hilos argumentativos es muy posible y hacia este tipo de intervenciones se han dirigido gran parte de los estudios argumentativos (Schneebeli, 2015; Padilla Herrada, 2015; Alonso, 2013; por citar algunos de ellos).

La traslación del antiguo foro convertido ahora en un espacio virtual con peristilo de fibra óptica, completa su estampa al advertir que algunos de los usuarios más prolíficos de la red de microblogging son los grupos políticos. En sus interacciones podemos encontrar movimientos dialécticos que actualizan formas clásicas de estrategias retóricas como la cita, en forma ahora de mención, o el argumento de autoridad, a través del retuit de discursos de otras personas. También en numerosas ocasiones observamos el recurso a la burla a través de ironías, juegos de palabras o directamente al insulto (Padilla Herrada, 2015). Estas últimas prácticas, notablemente más abundantes, dan la espalda a aquellas buenas prácticas argumentativas deseables que Paul (1993) condiciona a determinados rasgos del propio orador entre las que contaríamos la humildad, la empatía, la imparcialidad o la integridad y que Gascón (2015) entre otros reivindican como una tercera perspectiva a tener en cuenta en la valoración de nuestras argumentaciones. En este sentido podríamos inferir que en el espacio virtual, la implícita distancia entre los participantes puede facilitar la escalada hacia las malas prácticas y la pérdida del objetivo que se persigue. En el cenagal virtual podemos encontrar argumentaciones buenas que brillan como *raras avis*, falsas informaciones, peleas dialécticas que desvirtúan todo el proceso y la permanente sensación de que todos los hilos conversacionales son estériles, pues solo mueren por desinterés, sin resolver jamás el conflicto que, lejos de culminarse en una resolución razonable y meritoria del conflicto, como sugiere la pragma-dialéctica, encona más si cabe las posturas. Es más, parte además de la idea más que cuestionable según la cual nuestras participaciones son ampliamente recibidas de forma exitosa, entendiendo por tal que logran éxito persuasivo y comunicativo. No obstante en la mayoría de las ocasiones esto no ocurre bien por desinterés, ignorancia o falta de acceso a la Red.

Es destacable que los *tweets* que más impacto generan en base a la cantidad de retuits reciben, no agotan los 140 caracteres de rigor y suelen estar acompañados por una imagen que acompaña retóricamente. Ello sugiere que el público en general digiere y difunde mejor piezas de información ligeras y efectistas, haciéndolas circular con prisa sin apenas juicio previo, lanzándolas a una vida tan fugaz como la del interés con el que las consumimos. Valorar la nocividad de este gesto de difusión con piezas de información que pueden fácilmente ser falaces o falsas, excede en mucho el propósito de este resumen pero no deja de ser una cuestión a tener en consideración y más si atendemos a proyectos tales como la *e-democracy*. La discusión pública virtual, así como la posibilidad abstracta de que potencialmente todos tenemos voz en este espacio, genera la falsa sensación de participar con aportaciones muy elevadas, pero sin verdadero sustento teórico, lo cual se traduce en un amplio tráfico de datos que pese a su volumen no dejan de ser incontables intervenciones muy superficiales. Además genera la insidiosa tendencia a pensar que todo es opinable por todos. Ello, ya advertía Platón, es uno de los peligros de la democracia que puede convertirse fácilmente en una oclocracia, pero además genera episodios que pueden ser tan



nefastos como la opinabilidad de la ciencia (Luque y Martínez, 2017).

Así pues, si las prácticas argumentativas que se producen en las redes sociales han de ser modelo de algo, principalmente lo son de aquello que las buenas prácticas pretenden evitar. El nuevo espacio virtual, aunque se ofrece como una ventana muy interesante para recabar datos y rastrear analíticamente hilos argumentativos que además renuevan las viejas formas de producirse, no parece que suscite un incremento o mejora de estas. Tal vez el ansia participativa unida la pereza argumentativa que vacía nuestras aportaciones de contenido reduciéndolas a cascarones retóricos manidos y sin sentido, sea el motivo por el cual *Instagram* gana cuerpos de ventaja a *Twitter*, especialmente entre los jóvenes. La imagen en estado puro, un acto de habla casi pre-lingüístico que expresa, sin los ambages teóricos que las más de las veces nos exige un concurso mental para el que no tenemos tiempo.

## Referencias bibliográficas

- Alonso, E. (2016). “El estudio de la argumentación en los medios digitales. Herramientas y metodología”. *DILEMATA*, año 8, nº 22, 121–135.
- Alonso, E., Blanco, N. Cárdenas, S., Rodríguez, A. (2013). “El debate público en las Redes Sociales. Twitter España como estudio de caso”. *Revista Iberoamericana de Argumentación*, 7:1–13.
- Bamman D. y Smith, N. (2015). “Contextualized Sarcasm Detection on Twitter”. En *Proceedings of the Ninth International AAAI Conference on Web and Social Media*.
- Gascón, J. Á. (2015). “¿Es posible (y deseable) una teoría de la virtud argumentativa?” En *Actas I Congreso internacional de la Red española de Filosofía* (Vol. XI, pp. 41–51).
- Groarke, L. Palczewski C. H. y Godden, D. (2016). “Navigating the Visual Turn in Argument”. *Argumentation and Advocacy*, 52(4):217–235.
- Luque, V. J. and Martínez, S. (2017). “El desequilibrio estable de la democracia”. *Ludus Vitalis*, 47, 251–254.
- Paul, R. (1993). “Critical thinking, moral integrity and citizenship: Teaching for the intellectual virtues”. En *Critical thinking. How to prepare students for a rapidly changing world* (pp. 255–267). Santa Rosa, CA: Foundation for Critical Thinking.
- Padilla Herrada, M. S. (2015). “La argumentación política en Twitter”. *Discurso & Sociedad*, Vol. 9(4), 419–444.
- Reyes, A., Rosso, P. y Veale, T. (2013). “A multidimensional approach for detecting irony in Twitter”. *Lang. Resour. Eval.*, 47(1):239–268.
- Schneebeli, C. (2015). “Twitter rhetoric? Argumentation in a Twitter debate: a case study”. *6th International Language in the Media Conference*, Sep, Hamburg, Germany.

# Emotions without language. The case of infants and animals

Marie Michon

IHPST – Université Paris 1 Panthéon-Sorbonne

By considering emotions as propositional attitudes, cognitivists consider that emotional content is akin to propositional content (D’Arms & Jacobson, 2004; Lycan, 2012; Solomon, 1973). From this conception entails an interdependency between the two objects. If we consider that more or less all humans are capable of emotions, the cognitivist view does not appear as an issue, but we are missing two important points. First, *humans* in this acception seems to refer only to adults and that is a misconception because it should encompass all the developmental stages from infancy through adulthood, stages that would lead to the question of propositional thoughts in infants. Second, if it is true that more or less all humans are capable of emotions, they are not the only creature able to do so, animals are and they do not have propositional thoughts either (Davidson, 2001). The cognitivist hypothesis stresses too much the necessity of language for emotions. There are interesting approaches to explain emotions in non-linguistic entities, notably if we focus on developmental stages as Deigh’s does. What matters is what happens between the stimulus and emotional response, and I aim to show that there can be emotions without language because the link between the two is not trivial. I will discuss some of the issues raised to show that this idea of Davidson is not required: “*we can continue to explain the behaviour of speechless creatures by attributing propositional attitudes to them while at the same time recognizing that such creatures do not actually have propositional attitudes.*”<sup>1</sup> I do not believe that Davidson’s claim is correct or even functional, and in what follows I weave together Deigh’s arguments with some of my own to suggest a way of arguing for emotions without language.

Even if the link between propositions mastering and emotional abilities, following the cognitivist view, fails to account for emotions in infants and animals, there is a decisive distinction between the two entities. Namely, there is a distinction in their respective inability to master language because infants are at a pre-language stage of their life whereas animals are not. What is very interesting in Deigh’s theory is that it can give a satisfactory response to the problem mentioned despite this distinction. I will first focus on infant’s emotions, then I will shift to animal’s emotions, and I will finish by wondering about the dependency of concept mastering for language and what we can learn from it.

---

<sup>1</sup>Davidson 1985, 477–8

## Emotions in infants

Being at a pre-language stage, infants do not experience emotions as adults, and there may be different developmental kinds of emotions that would progress along the mastering of more and more complex abilities. Deigh's conception of pre-language emotions stems from this idea (Deigh, 2011). There are mature emotions—which are experienced by adults—and primitive emotions—which are experienced by infants and animals. He does not go into details to explain the transition between the two kinds but there is no specific threshold for language mastering in children and the process goes back and forth. The difference between primitive and mature emotions is in the properties, or the proficiency, of emotional occurrences, not so much as what emotion can be felt and one cannot<sup>2</sup>. Propositional attitudes give a special access to the world by mean of language, hence one could wonder how infants can make the world intelligible for themselves without this mean of language. Primitive emotions can be connected to a perceptive way of representing the world. In this way there would be no need of concepts such as *surprising* to experience surprise. In fact, there are examples of surprise recorded for infants (Wynn 1992). This shows that they have an idea of what a situation should have been, hence some expectations on the unfolding of the situation. A great part of the research on emotions focus on them being a set of neurophysiological cues and reactions (Ekman 1980), this is not in contradiction with the way babies apprehend the world around them and react to it. Primitive surprise would be akin to a primitive representation of a situation and its extent. Infants being at a biologically immature developmental emotional stage that consists of only feelings, stage that would then be implemented by a cognitive part when they grow up and acquire linguistic abilities. Absence of complex emotions would be linked to absence of complex thoughts, the development of both being linked. The absence of propositional attitudes entails the absence of evaluative judgments (Solomon 1973) but there can be other kinds of judgments, especially if we consider judgments in the sense of non-linguistic inferences. In this sense, the implications of Deigh's idea can echo Tappolet's Perceptual Theory (Tappolet 2016).

## Emotions in animals

Because animals are not at a pre-language stage of their life, one can wonder how a developmental explanation can be satisfactory to account for their emotions. The obvious way around lays with the fact that 'primitive' does not necessarily entail that there will be an evolution. Primitive animal emotions can be of the same nature than those of infants but with a finality in the stage. In this sense, the absence of language is also countered with some sort of representations. A similar idea is found in Dummett (2006) with his animal *protothoughts* consisting in spatial representations. Davidson (2001) also agree to some mental representation for animals. Keeping the example of surprise, we can refer to Malcolm's (1972) famous example of animal surprise: a dog chases a cat until they reach a

---

<sup>2</sup>That being said, it is clear that some complex emotions need a certain comprehension of social relations and functions, so they would not be found in infants even in a primitive way.

tree, the cat jumps at the last minute to another tree nearby but the dog keeps barking under the wrong tree. The dog has an idea of the situation—even a wrong one—and he acts upon it, then we can say that he would be surprised to see the cat in the next tree because he did not expect it. What is important here is that the dog cannot think but he can *believe* that the cat went up the tree (Malcolm 1972). Animals have perceptive thoughts about the objects of their behaviour even if they do not *think*. This explains why we have no difficulties in attributing basic emotions to animals such as joy or fear, and complex ones like shame to a certain extent. The fact that they manage to understand a situation and act accordingly denotes cognitive capabilities. The lack of language is not a problem because there are mental states that mediate between perception and action, namely the representations, and that is enough to account for emotional response. Furthermore, the fact that the response is adequate in regard of the situation shows that there is some kind of control in the emotional response. There are also other supports for this with Deona & Terroni (2012) who propose that emotions are bodily attitudes and animal need only to perceive and react (accordingly) to have them. This does not mean animals have a complete understanding of what happens in a situation and do it does not allow us to say they have some kind of concept mastering or reflexion, but that is not necessary to have emotions. So, Deigh's dichotomy also applies here but with the possible add-on of a developmental notion of conceptual mastery: primitive emotions allow for some watered-down concepts that may evolve into full concepts if the entity were to develop language. Note that this point is delicate because animals' emotions will not evolve from primitive to mature during their life, but in a way, we could infer that, if they did develop linguistic abilities then they would reach mature emotions. What is delicate is the fact that we think of linguistic abilities in a human sense; if we were to consider that some animals in fact possess a whole language—in any sense of the concept—then our distinction would need reframing, or our definition of emotions would.

## Link to concept mastering

Concept mastering is of interest because it is linked both to linguistic abilities and to emotions according to cognitivists, and it also offers an interesting view of Deigh's distinctions. Both primitive and mature emotions support some sort of concept mastering in order to have emotion, the distinction does not lay here. The difference is in the manner concepts are tied with representations. Indeed, mature emotions involve concepts and beliefs, but primitive emotions only involve some sort of representations. Then, one can ponder what happens when the transition from primitive to mature emotions happens: do we suddenly need concepts like *surprising* to be surprised because representations become insufficient? Is the addition of language the key component? Deigh's distinction needs to account for the transition for humans in a way consistent with what happens for animal emotions. According to the Brentanian tradition, to *believe* is to believe *something*, this raises issues for propositional attitudes because they are the link to the something that is believed. Animals do not have propositional attitudes because they lack language (Davidson, 2001), but they possess mental representations (Call & Kaminski, 2004), so at least a *certain kind of*

*belief*. Infants also possess a kind of beliefs (Wynn, 1992) which is illustrated by the fact that they can be surprised: they believe a situation is going to unfold a certain way and are surprised when it does not. Beliefs being propositional attitudes entails that there is a relation between a subject and an entity<sup>3</sup>. This kind of reification links the capacity of having beliefs to concept possession (Glock 2010) because having beliefs signifies that the creature can be correct or mistaken about a situation. Indeed, it would be incongruous to consider that for humans grasping a situation would mean grasping the concepts involved and not for animals; Glock argues for this based on a non-ambiguous definition of the verb ‘believe’. There is a lot support against this for animals (Deigh 1994, Tappolet 2016), and the fact that animals could in a way believe x, and hence possess some kind of concepts, does not entail that they have the thought that x (Malcolm 1972). An illustration of this lays in the examination of what can be the span of cognitive abilities for non-linguistic creatures, thus necessarily via a non-linguistic way of assessing cognitive abilities. The emotion of surprise is a good candidate. To grasp reality enough to be surprised does not require language (Wynn, 1992), (Hauser, MacNeilage, & Ware, 1996), nor the concept of surprising (Tappolet 2016). To be surprised, any subject must meet some requirements: understand a given situation, have some expectations about what is going to happen next, and the new situation must differ enough from the expectations to provoke the surprise. The presence of expectation leads to two cognitive assumptions: the strong one is belief; the weak one is mental representation. The strong assumption is that the subject has a belief about a situation and about what is going to happen, and it is a belief that is contradicted and generates surprise. The weak assumption is that the subject has a representation of a situation and the potential outcomes. This is in accordance to what has been said about two different kind of emotions: primitive emotions involve representations and mature ones involve beliefs. The transitions from primitive to mature is accounted for in the manner one considers the implications of *expectations* are. This is consistent for primitive emotions of animals and the passage to mature emotions from infancy to childhood and adulthood.

Language mastering is not a necessary condition for emotions because not all emotions are cognitively demanding, or rather they don’t always need to be. Describing emotions for non-linguistic creatures like animals proves a challenge for two main reasons: we need a non-linguistic way of assessing emotions, and, there is an inescapable anthropocentrism that bias our understanding. Nevertheless, this paper has shown that there are theories that prove useful to account for emotions for creatures without language, whether or not they will develop it later in their development. We can use a developmental scale to account for emotional complexity. Other ideas like Shepard (1987) offer an interesting point of view by establishing generalization laws for all animals (humans included) based on the assumptions there are regularities across species that may be evolutionary accommodations to universal properties of the world.

Another way of shifting the problem is by rethinking the category of propositional attitudes with a new stance on intentionality following Glock argument: “*one can believe something which is not the case, intend to do something which*

---

<sup>3</sup>The *something* that is believed.

*never happens, and love someone who does not exist.*"<sup>4</sup> In the first case there is a propositional attitude and that is not a surprise for a belief, but the other two cases are interesting because the second is an action-oriented attitude, and the third is an object-oriented attitude and it expresses an emotion. Then we could consider emotions not to revolve mainly around language but in to be object oriented by the entities feeling them. This follows the brentanian tradition and it is true that we then would have to question the idea of intentionality for animal which is delicate, but the work of some of the followers of Brentano can shed light on this idea.

## References

- Call, J., & Kaminski, J. (2004). Word learning in a domestic dog: evidence for "fast mapping.". *Science*, (304 5677).
- D'Arms, J., & Jacobson, D. (2004). Emotion and Action. In R. C. Solomon (Ed.), *Thinking about Feeling: Contemporary Philosophers on Emotions* (1st ed.). Oxford; New York: Oxford University Press.
- Davidson, D. (2001). *Subjective, Intersubjective, Objective*. Oxford University Press.
- Deigh, J. (1994). Cognitivism in the Theory of Emotions. *Ethics*, 104(4):824–854.
- Deigh, J. (2011). *Emotions, Values, and the Law*. Oxford University Press.
- Deonna, J., & Teroni, F. (2012). *The Emotions. A philosophical introduction*. Routledge.
- Dummett, M. (2006). *Thought and Reality*. Oxford University Press.
- Ekman, P., Friesen, W. V., & Ancoli, S. (1980). Facial Signs of Emotional Experience. *Journal of Personality and Social Psychology*, 39(6):1125–1134.
- Glock, H.-J. (2010). Can Animals Judge?. *Dialectica*, 64(1):11–33.
- Hauser, M. D., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Sciences*, 93(4):1514–1517.
- Lycan, W. G. (2012). Desire considered as a propositional attitude. *Philosophical Perspectives*, 26(1):201–2015.
- Malcolm, N. (1972). *Thoughtless brutes*, 46(September).
- Shepard, R., (1987). Toward a Universal Law of Generalization for Psychological. *Science*, 1–7.
- Solomon, R. C. (1973). Emotions and Choice. *The Review of Metaphysics*, 27(1):20–41.
- Tappolet, C. (2016). *Emotions, Value, and Agency*. Oxford University Press.
- Wynn, K. (1992). Addition and subtraction by human infants. *Open MIND*.

---

<sup>4</sup>Glock 2010 p. 13

# Existence as a property of intensions

Manuel Crescencio Moreno Gómez  
manuelcrescencio@gmail.com

## The Informal Account

Much has been said and questioned about the predicate of existence in the past. Our claim is that the predicate of existence should be applied to intensions and not to individuals in order to be non trivial. When you consider an individual of the domain  $D_i$  and you want to predicate existence of it, the only thing you are doing is situating the individual at a given world, but this is not a remarkable information. We defend that the predicate of existence must provide us with more singular information. Any individual of the domain of the model is always located at some world of the model, so it exists at one world or another. It is possible that the individual is not situated at your present world of evaluation but it is situated at another world, where it exists. Therefore, to say that an individual exists does not seem very noteworthy, given that to state the existence of an individual is only a way of saying that it is located at a given world.

We think that the predicate of existence has to convey more notable information and this is better done when it is attributed to intensions. To say that an intension exists is to state something really significant. In principle, you do not know if an (extensionalized) intension denotes an existent or not, you have only a function that can be defined at some worlds but not defined at others (remember that intensions are considered to be partial functions). Consequently, to state the existence of an intension at a given world is to give crucial information. An information that is not trivial because it gives information about the denotation of the intension and about the location of its denotation. The predicate of existence declares not only a location statement but also, and mainly, a denotation claim. In this case, we manage with a predicate of existence that could be defined as:

$$\text{Existence} = \text{Denotation} + \text{Location}$$

We also suppose that definite descriptions can be formalized as intensions in our logical language. It seems that we are more interested in the intension determining a concrete person than in the proper name of this person. It is not trivial to know that "the man who thinks to kill you exists", independently of his name. Proper names are informative to a lesser degree than intensional expressions. We feel real interest in knowing if the president of the United States exists or if the person who discovered the penicillin exists or if the lover of your wife exists, and they are all intensions whose existence matters. One way of testing if the existence of individuals matters is just to read the death notices

in a newspaper to see if they are significant to you, it is possible that you find some death notice interesting just in case a remarkable description is indicated as, for example, that the dictator of North Korea has died.

Intensions also, unlike proper names, allow us to determine the necessary conditions which have to be fulfilled for an individual in order to be designated by the intension. For example, the intension: the Pope, gives us the requisites needed by an individual in order to be Pope: he has to be a man, to be a Bishop, to be Catholic, to be the winner of a Cardinals' Conclave, etc. Between these properties, we can think of one property that can be not only a necessary condition but also a sufficient one, we call this peculiar property the *essence*. We can think of the essential property as determining all the other properties that an individual has to possess in order to be designated by a particular intension.

## The Formal Account

We also offer a formal account of existence in the language of our Intensional Hybrid Type Theory with a constant domain semantics.

**Definition 1** (Predicate  $\Delta$ ). We define the predicate  $\Delta_{\langle(t_1 t) o_1\rangle}$  to be an abbreviation of the following predicate abstract

$$\langle \lambda x_{t_1}, x_t. \downarrow a (@_a x_{t_1} = x_t) \rangle$$

The formula at the core of the predicate abstract is

$$\downarrow a (@_a x_{t_1} = x_t)$$

What this formula is claiming is that an intensional term of intensional type  $t_1$  designates an object of type  $t$  at world  $w$ . The formula can be read as saying that the intension of type  $t_1$  designates the object of type  $t$  at world  $w$ . Predicate  $\Delta$  is an intensional predicate which takes as arguments an intension and the extension of the intension. As a result,  $\Delta(x_{t_1}, x_t)$  can be read as: the intension  $x_{t_1}$  designates the object  $x_t$ . We can also impose to a given model the condition that every object is the designation of some intension at some world.

As we move in a constant domain semantics we can not differentiate by means of quantifiers between the objects that are situated at a certain world and those that are not situated at that world. If we want to do so we need to introduce a new predicate into the language:

**Definition 2** (Predicate **E**). The location predicate  $\mathbf{E}_{\langle \iota o_1 \rangle}$  is true at any world  $w$  of the objects of type  $\iota$  that are actually located at  $w$ .

Given predicates  $\Delta$  and **E** we can establish the following definition:

**Definition 3** (Existence). The existence predicate  $\mathcal{E}_{\langle \iota_1 o_1 \rangle}$  is an abbreviation for this predicate abstract

$$\langle \lambda x_{\iota_1}. \exists x_{\iota} (\mathbf{E}_{\langle \iota o_1 \rangle}(x_{\iota}) \wedge \Delta_{\langle (\iota_1 \iota) o_1 \rangle}(x_{\iota_1}, x_{\iota})) \rangle$$

Note that the existence predicate  $\mathcal{E}$  is only applied to first-order intensions of type  $\iota_1$ . What it essentially says is that an intension exists at a given world  $w$



provided the intension designates an object at  $w$  and the object designated is located at  $w$ .

# Expressivism: the Frege-Geach problems and new goals

Alba Moreno Zurita  
University of Granada

## Introduction

We can find in literature a plentiful amount of information on what is known as the touchstone of expressivism: The Frege-Geach problem (Frápolli & Villanueva 2012; Hom & Schwartz 2013; Horwich 2005; Schroeder 2008). The aim of this paper is to defend that there are at least two ways (Interpretation A and Interpretation B) to understand the so-called Frege-Geach problem and show that, within the contemporary expressivist options, the hybrid forms of expressivism are not affected by any of these two versions of the problem.

## Interpretation A (Embedding and truth-conditions)

According to Geach, what expressivist theories have in common is that they individuate the meaning of ethical terms in terms of what speakers do when they use them. This conception of the meaning of ethical terms poses certain difficulties when the terms are embedded within logical connectives (see Geach 1965). What happens when you embed non a truth-conditional string of symbols as an argument of a truth-functional expression?

These type of assertions do not express any content that is susceptible to being declared true or false. If that is the case, they cannot function as statements in truth-functional expressions or act as premise or conclusion in an argument like *modus ponens*. In other words, we cannot say that the truth of the conclusion follows from the truth of the premises; in order for the conditional statement to make sense, it must have truth value and to follow that, its antecedent must have truth value as well.

## Interpretation B (Embedding and sameness of meaning)

In Geach's critique of expressivist theories (see Geach 1960 and 1965), you can find reasons to argue that the problem Geach attributes to such theories is not strictly related to the truth-conditional character of the expressions which

contain ethical terms. Geach's main thesis, from which his critique stems, is that expressivist proposals are incompatible with a distinction between the *content* and the *force* of a proposition (see Geach 1965: 223).

The conditional is the paradigmatic example of the problems generated by embedding. Geach, like Frege, maintains that when an expression is embedded within a conditional, the expression loses the assertoric force that comes with it (see Geach 1965: 461–462). According to Geach, the theories which defend the contrary, that is, which defend that the *force* of a proposition determines its *content*, cannot give a satisfactory explanation of how certain expressions behave when they are inserted in an inferential process such as the *modus ponens* (Geach 1965: 462–463). The *modus ponens* is a clear example of an argument where the use or *force* of an expression varies, but not its *content*. According to Geach's critique, if, for expressivist theories, *content* depends on use, we are forced to say that the *content* also varies, and, so, we would end up with arguments like the following:

If it is wrong to lie, then it is wrong to lie to our parents ( $p \rightarrow q$ )

It is wrong to lie ( $r$ )

Thus, it is wrong to lie to our parents ( $s$ )

This kind of argument, construed in this manner, is not a *modus ponens*. In order for it to be a *modus ponens*, the propositional content must be the same for each case, despite the difference in use of the expressions.

If expressivism claims that the difference of meaning between both sentences depends on the different uses of the expression, then expressivism is to blame for a fallacy of equivocation. In order to preserve the validity of the argument, expressivism must provide the same meaning in both cases, that is, both when the expression “It is wrong to lie” appears isolated and when the expression is a part of a complex sentence (see Geach 1965: 462–464).

## Hybrid expressivism is not affected by any of these two versions of the problem

Hybrid expressivism (Chrisman (2007) avoids these two versions to approach the Frege-Geach problem, given the two following reasons. First, according to this sort of expressivism, a statement such as “It is wrong to lie” express two mental states. On the one hand, the belief that the act of lying is wrong in relation to certain standards (contextualist proposition). And, on the other hand, the acceptance of those standards (conative attitude).

This way, ethical statements would express content susceptible of being declared true or false, as they express a belief that has propositional content. Secondly, from the point of view of this proposal, the expression “It is wrong to lie” would have the same meaning when isolated as when it constitutes the antecedent of the conditional statement “If it is wrong to lie, then it is wrong to lie to your parents”. What happens, in this case, is that the same content is expressed through different uses. As in the previous case, when we utter “It is wrong to

lie” we express the belief that the act of lying is wrong in relation to certain standards, as well as expressing the acceptance of those standards. In this case, an evaluative use of the term “to lie” is expressed. On the other hand, when we utter “If it is wrong to lie, then it is wrong to lie to your parents”, there is a non-evaluative use of the terms, but the set of accepted standards is the same given that the standards are a part of the expressed proposition. In consequence, this is a case in which, with different forces (namely an evaluative use of the term and a non-evaluative use of the term), the same content is expressed.

## References

- Chrisman, M. (2007). From epistemic contextualism to epistemic expressivism. *Springer*, 135:225–254.
- Frápolli, M.J. & Villanueva, N. (2012). Minimal Expressivism. *Dialectica*, 66(4):471–487.
- Geach, P.T. (1960). Ascriptivism. *The Philosophical Review*, 69(2):221–225.
- Geach, P.T. (1965). Assertion. *The Philosophical Review*, 74(4):449–465.
- Hom, C. & Schwartz, J. (2013). Unity and the Frege-Geach problem. *Philosophical Studies*, 163:15–24.
- Horwich, P. (2005). The Frege-Geach Point. *Philosophical Issues*, 78–93.
- Schroeder, M. (2008). What is the Frege-Geach Problem?. *Philosophy Compass*, 3/4: 703–720.

# La lógica del siglo XIX desde la perspectiva de Edmund Husserl y su fenomenología

David Orellana Viñambres

El texto que se presenta a continuación, forma parte de un proyecto de investigación ya concluido y ha sido publicado por el autor en formato libro bajo el título *El estudio de la lógica formal en la fenomenología de Edmund Husserl* (2018). La tesis rectora de dicha publicación se presenta, en esta ocasión, a modo de resumen explicativo del proceso de revisión crítica al que el propio Husserl sometió a la historia de la lógica hasta finales del siglo XIX y que, en dicha obra, el lector podrá encontrar un estudio en profundidad al respecto, del cual ahora se presentan las conclusiones.

El proyecto fenomenológico se reveló como la crítica definitiva a la lógica formal, en tanto la convierte en tema mismo de predicación re-flexiva acerca de su sentido y de la forma como se constituye originariamente. En este proceso, se ha evidenciado el reduccionismo de la concepción lógica tradicional fuertemente influenciada por las corrientes psicologistas y metafísicas. Husserl reconoce que la lógica no debe ser un conocimiento aislado de la realidad en base a un formalismo que, en su distanciamiento de la realidad, ha descartado toda posibilidad de alcanzar el saber auténtico. Gracias a la crítica a la lógica formal, nuestro autor ha mostrado toda una nueva concepción de la lógica desde la fenomenología, no solo escapando del idealismo ingenuo en el que otros autores, como el propio Husserl señala, han caído sino, precisamente, afirmando la experiencia del mundo como origen mismo de la predicación. La lógica no puede ser concebida como algo simplemente atemporal y extra-mundano, al modo de la ciencia garante del en-sí, debe abrirse al mundo desde el que la propia fenomenología encuentra su razón de ser.

La fenomenología se concibe, desde la revisión de la lógica formal, como una teoría crítica del conocimiento en general, que tiene que ver con el intento por descubrir en qué medida es posible el conocimiento sobre lo real. Ya se sabe, tras la crítica a la razón lógica, que la lógica formal no tiene la última palabra al respecto del ser de las cosas mismas, pues en ella simplemente éstas adquieren sentido en tanto existían previamente al acto mismo por el cual son conocidas, esto es, a través del cual son referidas por la predicación.

El profesor San Martín (1986), dice al respecto en *La estructura del método fenomenológico*:

La motivación estructural de la fenomenología no puede ser otra que el resultado negativo de una reflexión sobre las ciencias al constatar que éstas no pueden realizar el ideal teórico de un conocimiento capaz de fundamentar la totalidad de la praxis humana. (pp. 48–49).

Este conocimiento de la acción humana no se puede reducir a la explicación de ciencia concreta alguna, por ello la lógica, en tanto se concibe como sistema de principios en que se fundamenta la forma por la que se accede al conocimiento por medio de la predicación, debe abrirse a la experiencia del mundo no para suplir con sus resultados a ciencia alguna sino, precisamente, para hacer que toda ciencia concreta pueda contribuir, a partir de la forma que le es propia, al auténtico conocimiento acerca del ser de los objetos del mundo a los que se refiere. Si la lógica formal no se amplía desde la fenomenología husserliana, queda reducida a una ciencia particular más y, por tanto, no contribuye al proyecto epistemológico de la fenomenología de fundamentar el saber auténtico a partir de unos principios formales que permitan superar toda concreción y puedan identificarse como *a priori* y de validez universal.

La lógica tradicional, desde la perspectiva de la crítica husserliana a la lógica formal, debe ser ampliada para poder integrarse en el cuerpo fenomenológico. En esta perspectiva de la lógica formal se reconocen tres niveles de categorías lógicas (forma de los juicios, consecuencia interna del juicio, y verdad de los juicios) pero, además, al convertirse los objetos mismos (objetividades ideales) en tema del juicio, ello permite dotar de sentido a los propios objetos y, por tanto, lograr auténtico conocimiento. Exactamente, lo que ocurre es que Husserl lleva hasta las últimas consecuencias su búsqueda radical del saber auténtico y, por ello, supera la “ingenuidad” inicial de la lógica tradicional basada en una concepción meramente objetivo-formal que constituyera una lógica pura como teoría del conocimiento. Avanzada la crítica husserliana a la lógica formal, nuestro autor descubre que en este proceso interviene también el sujeto (ideal, no empírico-real), en tanto los objetos adquieren su sentido a partir de su aparecerse a la conciencia (intencional), una de las tesis principales de la fenomenología husserliana.

Así, mientras la lógica formal tradicional quedaba reducida al análisis de la forma del juicio y su consecuencia interna (no contradicción), la significación, en base al carácter intencional de la conciencia, va a asentar la búsqueda de la forma como se presenta el conocimiento en los principios formales de esta lógica formal pura o lógica ampliada, que se conocerá como lógica trascendental. No es tema de la presente comunicación pasar a exponer en detalle la constitución de esta lógica trascendental de Husserl, basta con indicar que, como se sabe, el conocimiento se presenta en la forma de la predicación a través de la cual los objetos ideales (conceptos formales de “objeto”, “relación”...) adquieren sentido y se conciben como correlato intencional de la propia conciencia.

La lógica, desde este punto de vista, no es simplemente un sistema de reglas formales por el cual se presenta el conocimiento sobre el mundo como un conjunto de principios formales (leyes) que permiten fundar la forma como se da el conocimiento de aquellos objetos, los cuales se dan sí mismos y por sí mismos (con independencia del juzgar) con evidencia y desde la experiencia del mundo. A medida que se desarrolla la crítica husserliana se aprecia que la lógica formal tradicional no funda el conocimiento a partir de parámetros formales objetivos, sino que, en realidad, este conocimiento se fundamenta en toda una experiencia pre-predicativa del mundo por la cual los objetos se tornan evidentes, esto es, siendo tal y como son. Pero entonces, ¿qué papel le reconoce el propio Husserl a la lógica formal tradicional hasta el siglo XIX? ¿Qué pretendía Husserl al de-

batir con los principales lógicos de su época a lo largo de toda su bibliografía? El de dotar de sentido a los objetos mismos a partir de su significación ideal, a partir de la forma de los juicios sobre objetividades que se fundamenta en unos principios formales universales y *a priori*. Así, la analítica ampliada contribuye, en cierta medida, al esclarecimiento de la fundación originaria de los juicios, a partir de la forma como las objetividades que tiene por tema adquieren significación y validez, adquiriendo el carácter evidente del auténtico conocimiento.

La lógica husserliana, tal y como aparece presentada en *Prolegómenos*, pretendía ser una teoría de la ciencia en general, por lo que al propio Husserl se le hacía indispensable determinar previamente cómo los objetos adquieren sentido en su tornarse evidentes. No obstante, reconociendo que la existencia de dichos objetos (ideales) precede a la predicación misma y, por tanto, al análisis de la misma.

La lógica formal, a diferencia de lo pensado por la tradición, no funda el saber, sino que éste se fundamenta en la intencionalidad de la conciencia por la cual los objetos adquieren sentido en su darse sí mismos a la conciencia y ésta descubre el ser verdadero de los objetos (del mundo, del yo, de la comunidad intersubjetiva...), y los conoce a partir de la forma como se dan, que la propia lógica formal ha proporcionado. Esta es precisamente la principal contribución de la lógica formal al proyecto fenomenológico.

En definitiva, desde la concepción husserliana, se habrá de superar la idea según la cual existen determinados parámetros formales por los que se accede al ser-en-sí de los objetos, del mundo, de las cosas mismas. Hay que reconocer, primero, que éstas ya se dan con independencia de la conciencia, pero que en ella adquieren su propio sentido (se convierten verdaderamente en lo que son a partir de su darse previo siendo ello mismo). El conocimiento de ello se presenta en la forma de la predicación, por lo que la lógica formal no puede, según Husserl, seguir defendiendo la idea de una verdad absoluta como fundamento de todas las ciencias sin reconocer que en su fundamentación misma interviene la subjetividad y todo un horizonte previo de los objetos supuestos. Lo objetivo del conocimiento deja de concebirse como absoluto- formal y se hace depender del carácter intencional de la conciencia, de la actividad de la propia subjetividad. Así, si realmente se pretendía en la época constituir una teoría de la ciencia en general, tuvo que venir la fenomenología para sacar a la lógica formal del reduccionismo de la analítica apofántica y descubrir que ésta misma se funda en la subjetividad, en el darse los objetos pre-existentes a la propia conciencia a partir de la cual adquieren su verdadero sentido. Así, la lógica trascendental no es punto y aparte de la lógica formal, sino fundamentación de ella a partir de la subjetividad trascendental, de tal forma que no sólo se accede a lo predicable de los objetos sino a su ser verdaderamente tal y como son, de lo cual lograr conocimiento a partir de su tornarse evidente. No se puede, por tanto, sostener la idea de que la fenomenología se origina en el camino desde la lógica formal hacia la lógica trascendental. Más bien al contrario, la lógica trascendental viene a fundamentar a la propia lógica formal desde el reconocimiento de la forma como los objetos se dan sí-mismos a la propia conciencia desde un horizonte de pre-existencia por el cual adquieren auténtico sentido.

¿En qué momento Husserl se da cuenta de la necesidad de reflexionar sobre la forma como se dan los objetos? Precisamente, a partir de la crítica a la lógica

formal en su concepción tradicional del siglo XIX. Por ello, se defiende en esta comunicación la importancia de reconocer toda la trascendencia de la crítica husserliana a la lógica formal en su forma tradicional hasta el siglo XIX, precisamente en el momento en que comenzaba a edificarse el proyecto fenomenológico. La crítica a la lógica formal es un tema permanente en la obra de Husserl y el estudio de la lógica trascendental de las obras más tardías del profesor de Friburgo no viene sino a recordárnoslo; por ejemplo, teniendo en cuenta que la predicación se refiere a los objetos (ideal-categoriales) en su darse formalmente, con evidencia, a la conciencia del *ego* trascendental.

La forma como los objetos se dan nos pone en aviso de su ser sí mismos, conocimiento acerca de lo cual solo se puede lograr a partir de relacionar el sujeto que conoce y el objeto que lo trasciende. Así, vemos que la lógica formal no se supera para quedar relegada de la fenomenología de Husserl, sino que se reconoce en ella la importante labor de determinar la forma esencial a partir de la cual poder identificar el conocimiento como válido, esto es, a partir de la evidencia de los objetos siendo sí mismos. Tal y como el propio Husserl señala en sus *Meditaciones Cartesianas* (1931), “la evidencia de la experiencia del mundo requiere en todo caso de una previa crítica de su validez y alcance; o sea, que no debemos admitirla sin examen como inmediatamente apodíctica” (pp. 24–25<sup>1</sup>). La crítica de la experiencia del mundo será, precisamente, la labor de la lógica fenomenológica a partir de la subjetividad trascendental. Por ello, le corresponde a la lógica formal la tarea de determinar, a partir de sus principios formales, cómo se accede al conocimiento válido por medio de la predicación en la cual se encuentran (una vez ampliada la analítica apofántica) el juicio y los objetos, la mención y lo mentado siendo sí mismo.

Gracias a esta crítica de la razón lógica es posible reconocer la experiencia previa del mundo de la que se ocupará la lógica trascendental a partir de la actividad constituyente de la subjetividad trascendental. Así, la crítica a la razón lógica transcurre a lo largo de la amplia bibliografía del autor, y es mucho más que un tema residual del universo husserliano; es parte indispensable para la comprensión del origen y desarrollo de la fenomenología trascendental legada por Edmund Husserl.

---

<sup>1</sup>Extracto de Husserl, E. (2009). *Meditaciones cartesianas*. Mario A. Presas (Trad.). Madrid, España: Editorial Tecnos.



# Déficit de conocimientos médicos y no fallos en las estrategias de razonamiento como principales responsables de los errores diagnósticos

Carmen Pascual

Los estudios de la psicología cognitiva nos han revelado que los humanos no somos razonadores probabilísticos naturales. Los estudios descriptivos del razonamiento humano, al menos desde el hito que supuso la publicación de *Judgment under Uncertainty: Heuristics and biases* por Tversky y Kaneman en 1974 (1) han tratado de comprobar y replicar nuestra utilización de heurísticos y su tendencia a producir errores en la estimación de probabilidades. Los heurísticos no son fuentes habituales de error para todos los autores. Algunos, como G. Gigerenzer, los consideran herramientas adaptativas (2). Este último autor acuñó el término de *fast and frugal heuristics*, que caracteriza al heurístico como una regla que selecciona información relevante y es capaz de llegar a soluciones correctas, a veces con más eficiencia que otros métodos más laboriosos y que utilizan toda la información disponible –como, por ejemplo, modelos de regresión múltiple– si se aplica al tipo de problemas y en el contexto adecuados. Hay que evaluar el contexto en que un heurístico es aplicado para entender su racionalidad antes de dictaminar su tendencia a la producción de sesgos. Por ejemplo, los médicos piden más pruebas y procedimientos de los que parecerían necesarios en entornos en los que el riesgo de ser denunciados es alto. En cuanto a los análisis de decisión clínica basados en razonamiento bayesiano, requieren datos estadísticos fiables, de los que no siempre se dispone.

Los modelos duales, en sus diversas versiones, postulan dos fases en el proceso cognitivo. La primera, rápida y automática, engloba tanto el reconocimiento de patrones y evocación de categorías como el uso de reglas heurísticas. La segunda, rigurosa y reflexiva, actuaría como revisora y correctora de las conclusiones alcanzadas durante la primera (3). Desde hace varias décadas predomina la atribución de los errores diagnósticos médicos a exceso de confianza en la primera y a negligencia en la segunda. Sin embargo, no ha podido demostrarse que el uso de heurísticos sea responsable de los errores médicos. No se han realizado estudios en los que se definieran de forma algorítmica y se comprobara que su uso y consiguiente sesgo suceda de la manera que los defensores de la teoría de los Heurísticos y Sesgos mantienen (4). Además, un metanálisis realizado en 2015 (5) sobre los estudios en sesgos cognitivos y uso de heurísticos en la decisión médica mostró que el 77 % de ellos habían sido realizados con viñetas hipotéticas y sólo en el 34 % los participantes habían sido médicos.

Los programas destinados a potenciar la fase dos o analítica del proceso de razonamiento en la toma de decisiones médicas, basados en la premisa de que

si se incentiva el proceso reflexivo y se advierte a los médicos del peligro de ciertos heurísticos, los errores médicos se reducirán significativamente, no han dado los resultados esperados. Ningún estudio ha demostrado ningún efecto de esta estrategia sobre el índice de errores. Un único estudio llevado a cabo por Mamede et al. en 2008 observó que la práctica reflexiva era útil cuando el médico se enfrentaba a casos especialmente complejos o inusuales, pero que no incidía en el índice de diagnósticos correctos en su práctica clínica habitual (6).

A pesar de que no se ha demostrado la responsabilidad de los heurísticos en la ocurrencia de los errores médicos ni su disminución tras proporcionar directrices de los modelos de razonamiento dual, sigue actualmente presente en la literatura. Durante mucho tiempo el conocimiento médico se ha identificado con el conocimiento biomédico y el juicio clínico se ha considerado implícito; algo que el médico realiza, pero cuyos contenidos y procesos no es capaz de explicitar. Entre los diferentes acercamientos a la semántica del diagnóstico médico con la intención de describir el tipo y estructura de los conocimientos puestos en juego en la decisión clínica, destacaremos la propuesta de los *Illness Scripts* o Esquemas de Enfermedad. Los *Illness Scripts* describen la secuencia de procesos cognitivos, en el sentido de actividades mentales que se ponen en marcha de forma ordenada cuando un médico se enfrenta al diagnóstico de un nuevo caso. El conocimiento de estos esquemas se adquiere a través de la exposición repetida a casos similares. Incluyen el reconocimiento y discriminación de patrones, la formación y recuerdo de prototipos o categorías, las acciones que se han de realizar para confirmar o descartar las hipótesis diagnósticas, correspondientes a los patrones o categorías evocados, además de todos las exploraciones y procedimientos necesarios para llegar a establecer un diagnóstico con la mayor certeza posible (7). El médico dispone de múltiples esquemas, adecuados a los diferentes tipos de problemas diagnósticos a los que se enfrenta. Las redes de asociación de los síntomas y signos entre sí y a su vez de constelaciones de ellos con determinadas enfermedades, así como los pasos necesarios para obtener, seleccionar y organizar la información relevante para identificar un modelo de problema o un tipo de enfermedad no son solo el inicio del proceso diagnóstico, sino el núcleo del mismo. Si no se considera la enfermedad del paciente entre las posibilidades diagnósticas en la primera fase del proceso, las probabilidades de alcanzar finalmente el diagnóstico correcto decrecen (8).

La adquisición de estos *scripts* y su activación selectiva forman parte del conocimiento médico más implícito. Es decir, aquél que aún no se ha podido expresar algorítmicamente de forma eficaz, a pesar de los avances en este sentido por parte de los sistemas basados en lógica difusa.

La formación médica que obtiene mejores resultados es la basada en casos, reales o recreados. Un estudio de revisión de 2012 encontró que la mayoría de los errores médicos están relacionados con *actos cognitivos subóptimos*, que corresponden en su mayoría a errores, acciones incorrectas realizadas a causa de desconocimiento de su incorrección o de falta de conocimiento de la acción correcta (9).

Por lo tanto, el error diagnóstico parece más una consecuencia de un defecto de conocimiento médico que el resultado de un fallo de razonamiento.

## Bibliografia

1. Tversky A, Kaneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974; 185: 1124-31.
2. Marewski JN, Gigerenzer G. Heuristic decision making in medicine. *Dialogues in Clinical Neuroscience* 2012; 14(1): 77-89.
3. Monteiro SM, Norman G. Diagnostic Reasoning: Where We've Being, Where We've Going. *Teaching and Learning in Medicine* 2013;25(S1): S26-S32.
4. Fiedler K, von Sydow M. Heuristics and biases: Beyond Twersky and Kaneman's (1974) judgment under uncertainty. *Cognitive Physiology: Revisiting the classic studies*. SAGE, 2015.
5. Blumental-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Med Decis Making* 2015; 35(4): 539-57.
6. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ* 2008; 42(5): 468-75.
7. Charlin B, Boshuizen HP, Custers EJ, Feltovich J. Scripts and Clinical Reasoning. *Med Educ* 2007; 41: 1178-84.
8. Neufeld VR, Norman GR, Feightner JW, Barrows HS. Clinical problem solving by medical student: a cross-sectional and longitudinal analysis. *Med Educ* 1981; 15(5): 315-22.
9. Zwaan L, Thijs A, Wagner C. Van del Wal G., Timmermans DR. Relating faults in diagnostic reasoning with diagnostic errors and patient harm. *Acad Med* 2012; 87(2): 149-56.

# Lógica y razonamiento no lingüísticos: el ejemplo de los gráficos existenciales de Ch. S. Peirce

Carmen Sánchez Ovcharov  
Karim Gherab Martín

## Resumen:

El propósito de esta comunicación es atraer la atención sobre la principal propuesta de elaboración de una lógica no lingüística: los gráficos existenciales de Ch. S. Peirce. Dentro de su proyecto semiótico estos gráficos se clasifican como un tipo de signo icónico: el “diagrama”, cuya característica esencial es que guarda una relación de semejanza estructural –isomorfía– con el objeto de su representación. Comúnmente los diagramas han sido considerados como una herramienta auxiliar, secundaria y prescindible, puesto que se han empleado para presentar de forma simplificada expresiones formales complejas. No obstante, vamos a defender que la diagramatización de las fórmulas y argumentos lógicos no constituye una simplificación, sino un aglutinamiento que puede revelar, a través de un razonamiento no lingüístico, estructuras comunes a las expresiones representadas.

## Palabras clave:

Lógica simbólica, lógica diagramática, lógica icónica, Peirce, gráfico existencial, lógica no lingüística.

En el panorama clásico o tradicional de la Lógica (resultado del programa logicista de depuración del lenguaje natural u ordinario), los diagramas, gráficos, figuras, etc., han desempeñado un papel meramente auxiliar, siendo empleados en ocasiones con una finalidad pedagógica para ofrecer una representación más simplificada de determinadas relaciones sistémicas o estructurales, comúnmente relativas al lenguaje natural o formal. Toda teoría de la deducción se ha considerado una teoría *sobre* estructuras lingüísticas, con sus correspondientes componentes sintácticos y semánticos. Hay que recalcar que, de modo análogo, en física, la diagramatización se ha empleado desde el mismo enfoque: como una técnica de simplificación de manejo del lenguaje formal matemático y, por lo tanto, prescindible y auxiliar (p.e., los diagramas de Feynman).

Charles Sanders Peirce, en su proyecto semiótico, toma como punto de partida la ruptura con dos tradiciones: la *eidética* (las ideas son de la mente y están en la mente: implica el problema de establecer –forzar– la relación entre la idea y el objeto o referente) y la *atomista* (búsqueda de las unidades últimas carentes de estructura; Frege, Russell,...). En su lugar, toma la vía de los *signos* (monismo sígnico): *todo* es signo, incluidos sistemas y entidades no lingüísticos. Esta nueva perspectiva, de carácter holístico, vuelve a abrir la ventana a la

dimensión epistemológica no centrada en el lenguaje (ventana entreabierta por los diagramas de Aristóteles, Leibniz, Euler, Venn, entre otros): la apropiación significativa de la realidad por vías **no lingüísticas**, lo que afecta a diferentes campos, incluido el de la Lógica.

Peirce no fue el único, pero sí el que ofreció el sistema más completo para llevar a cabo el gran paso de la *Lógica Simbólica* (lingüística) a una *Lógica no lingüística*: la *Lógica Diagramática* o *Icónica*. Este sistema es el denominado por él mismo sistema de Gráficos Existenciales; Alfa – lógica de orden cero, Beta – lógica de primer orden, y Gamma – lógica modal, que fue un proyecto no acabado. En esencia se trata de un sistema *iconográfico* de símbolos que guardan *semejanza* con las relaciones de objetos mentados por ellos; se trata de una *sintaxis diagramática*.

Obsérvese abajo [Imagen 1] la representación diagramática (izq.) de las seis fórmulas de lógica simbólica (dcha.). Hay una clara diferencia cualitativa, que se debe al cambio de un sistema signico *lingüístico* a uno *no lingüístico*:



Imagen 1: Gráfico alfa (izquierda) correspondiente a las fórmulas de lógica simbólica (derecha).

Como vemos en la Imagen 1, no solo en la deducción se revelan “relaciones no advertidas” sino también en las relaciones de equivalencia lógica entre fórmulas moleculares. Se puede afirmar que el gráfico alfa de la izquierda revela la estructura común –**la estructura de equivalencia**– entre las fórmulas moleculares de la derecha.

Para entender las implicaciones de ese paso, de lo simbólico a lo icónico, debemos precisar las nociones de “símbolo” e “ícono”, dentro de la semiótica peirceana. Peirce plantea una constitución *triádica* del signo, en el que se articulan tres componentes indisolubles: el objeto, el representamen (signo) y el interpretante (signo superior), de modo que  $x$  (representamen) es el signo de  $y$  (objeto) para  $z$  (interpretante). La fusión de la tríada objeto-signo-interpretante es el fundamento constitutivo del fenómeno semiótico. Entre estos tres componentes se pueden realizar tres tricotomizaciones, según tipo de relación entre ellos:

1. Relaciones de *comparación*: relaciones de la naturaleza de las **posibilidades lógicas**. Son las relaciones del signo (representamen) consigo mismo y sufren la siguiente tricotomía: *cualisigno* (cualidad) – *sinsigno* (hecho, objeto) – *legisigno* (ley, regla).
2. Relaciones de *actuación*: relaciones de la naturaleza de los **hechos**. Son relaciones del signo con el objeto y sufren la siguiente tricotomía: *ícono* (semejanza con el objeto) – *índice* (indicio del objeto) – *símbolo* (significado establecido por convención).

3. Relaciones de *pensamiento*: relaciones de la naturaleza de las **leyes**. Son relaciones del signo con el interpretante y sufren a la siguiente tricotomía: *rema* (genera en la mente la idea de una propiedad) – *decisigno* (desarrolla en la mente una proposición) – *argumento* (desarrolla en la mente un razonamiento).

Las lógicas lingüísticas –en tanto que lenguajes artificiales– estarían en una relación *simbólica* con su objeto, es decir, en una relación basada en la convención, característica específica de la *terceridad* (legisigno, símbolo, argumento), por lo que constitutivamente estarían *alejadas* de lo representado (el objeto), por fundamentarse en acuerdos de una comunidad. La idea de Peirce es traer la significación, del nivel más alejado del signo respecto al objeto significado (nivel *simbólico*), al más cercano, el de *semejanza*: el nivel *icónico*. De esta forma, la significación *procedería* –de manera icónica– del objeto significado, pues íconos son un tipo especial de signos: su manipulación proporciona información acerca de su referente (lo significado por ellos). Aplicado al campo de la lógica, *construir* un signo icónico –el grafo existencial instanciado– tendrá como función hacer *visible* la estructura del razonamiento que representa.

Queda por precisar qué tipo de ícono es el grafo existencial para comprender cómo la significación procede del objeto significado. Para ello debemos profundizar en la tricotomización que sufre el signo icónico, en función del tipo de semejanza que guarda, como signo, con su objeto:

- a. **Imagen**: el signo icónico y el objeto guardan una relación de semejanza entre sus *cualidades* simples (color, forma,...). Ejemplo: un retrato.
- b. **Diagrama**: el signo icónico y el objeto guardan una relación de semejanza –*isomorfismo*– entre sus elementos constituyentes. Ejemplo: un mapa, un plano, etc.
- c. **Metáfora**: el signo icónico y el objeto guardan una relación de semejanza entre sus *modos de representar*; poseen la capacidad de generar los mismos interpretantes (interpretantes similares en efectos). Ejemplo: la vida es una montaña rusa.

Un grafo existencial sería un *diagrama*, que guarda una relación de semejanza estructural –isomorfía– con lo significado. Siendo hechos o estados de cosas lo que la formalización lógica procura reflejar estructuralmente (tradicionalmente mediada por el lenguaje) los diagramas reflejarían visual e isomórficamente las relaciones constitutivas de hechos.

Precisemos:

“Un diagrama es un ícono de un conjunto de objetos racionalmente relacionados. [...] El diagrama no solo representa los correlatos vinculados, sino también, y de manera mucho más definida, la relación entre ellos. [...] El razonamiento necesario lleva a una conclusión *evidente*. ¿Qué es esta ‘evidencia’? Ella consiste en el hecho de que la verdad de la conclusión es percibida, en toda su generalidad, y en la generalidad de cómo y por qué la conclusión es *percibida*. [...] Es [...] un rasgo muy extraordinario de los diagramas que ellos *muestran* [...] que se sigue una consecuencia. [...] De todos modos, no es el diagrama-ícono estático que muestra directamente esto, sino

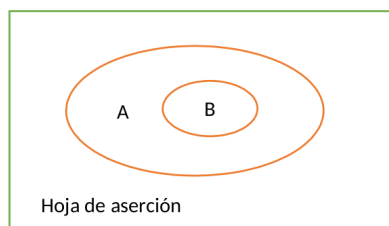
el diagrama-ícono construido con una intención.” (Peirce, NEM IV: 316)

Ahora la cuestión a resolver es dar cuenta de la relación que guarda la forma lógica diagramática con el razonamiento concreto (incluido el expresado en lenguaje ordinario). La respuesta de Peirce está en otro tipo de signo: el denominado **decisigno**, que proporciona un elemento indexical a la proposición; es un índice que contiene un ícono, es una proposición que *dice* un hecho. El conjunto de decisignos (proposiciones) estructurado según reglas específicas, dará lugar a un argumento: un razonamiento deductivo.

Veamos cómo se despliega diagramáticamente un *decisigno*:

- I. GRAFO EXISTENCIAL (*tipo* general, no existe; determina cosas que existen, es la generalización de *tokens*: algo concreto dado en un espacio y un tiempo determinados). Es la posibilidad de cualquier gráfico existencial particular.
  - i. Es primeridad (el blanco en sí – vacío).
  - ii. Es un **legisigno** icónico remático.
- II. GRAFO INSTANCIA (instanciación del tipo; concreto), es el **DECI-SIGNO**.
  - i. Es segundidad (un espacio en blanco).
  - ii. Es un **sinsigno** icónico remático.
  - iii. Consta de:
    - HOJA DE ASERCIÓN (primeridad: posibilidad)
    - LETRAS y CORTES (segundidad: existencia actual)
    - REGLAS (terceridad: reglas semióticas de transformación)

Un ejemplo de gráfico instancia (decisigno) es el que tenemos abajo: inscrito sobre la hoja de aserción y con dos cortes. Las reglas básicas expuestas al lado permiten leer correctamente el gráfico y establecer la naturaleza de la proposición que significa diagramáticamente, que no es otra que un condicional:



Reglas básicas de los gráficos alfa:

**Afirmación:** la escritura de una letra sobre la hoja de aserción significa la afirmación de la fórmula representada por la letra [A,B].

**Negación:** encerrar una letra o gráfico en un recorte [óvalo naranja] significa negarlo; recortarlo de la existencia.

**Conjunción:** dos letras/gráficos juntos en una hoja de aserción significan su conjunción.

Ejemplo de gráfico (fórmula diagramatizada:  $P \rightarrow C$ ) y las reglas básicas de construcción de gráficos

Para la *deducción alfa* existen otras reglas que permiten derivar un gráfico de otro y construir la secuencia gráfica de la deducción (los gráficos *beta* y *gamma* poseen sus correspondientes reglas). Se trata de las reglas de *inscripción* y de *borrado* en la hoja de aserción. Constituyen las reglas del cálculo formal diagramático. Las principales son tres:

1. Regla de *inserción*: en un área, anidada dentro de cortes, con un número impar de cortes puede insertarse cualquier gráfico (letra o letra con recorte(s)).
2. Regla de *iteración*: si un caso o réplica de gráfico A aparece en un área, puede insertarse otra réplica de A en el mismo área o en un área anidada en A.
3. Regla de *doble corte*: si entre dos cortes no aparece ningún gráfico A, entonces los dos cortes pueden ser borrados o insertados en cualquier gráfico.

Aplicando estas reglas sobre un gráfico dado (premisa) se construye el gráfico-conclusión. Así, la deducción cobra un carácter icónico-diagramático:

“Por razonamiento diagramático entiendo el razonamiento que construye un diagrama de acuerdo con un precepto expresado en términos generales, realiza experimentos sobre este diagrama, toma nota de sus resultados, se asegura de que experimentos realizados sobre cualquier razonamiento que sea construido con el mismo precepto tengan los mismos resultados, y expresa esto en términos generales.” (Peirce, NEM IV: 47–48)

Y:

“Todo razonamiento deductivo [...] contiene una relación de observación; es decir, la deducción consiste en construir un *ícono* o *diagrama*, la relación de cuyas partes presenta una completa analogía con la de las partes del objeto de razonamiento, en experimentar sobre esta imagen en la imaginación y en observar el resultado, de modo de descubrir relaciones no advertidas y ocultas entre las partes”. (Peirce, CP. V. 165; 3.363)

Los gráficos permiten no solo manejar diagramáticamente estructuras de razonamiento deductivo, sino que *muestran* y *hacen visibles* estructuras unificadoras de varias fórmulas proposicionales; podríamos decir que los gráficos *aglutinan* conjuntos de fórmulas (proposiciones) lógicamente equivalentes y ese aglutinamiento es a su vez una estructura **—forma—** lógica hallada *sin el recurso a lo simbólico-lingüístico*. Un gráfico existencial es al grupo de equivalencia de proposiciones que formaliza diagramáticamente, lo que la proposición es al grupo de equivalencia de enunciados que formaliza simbólicamente. Así, el gráfico existencial puede ser considerado una forma lógica de orden superior manifestando la posibilidad de elaborar razonamientos a nivel gráfico: diagramáticamente.



## Bibliografía

- Jiménez Moskovitz, L. (2003). *Gráficos existenciales en Gama Color y algunos sistemas de Lógica Modal.* Univ. Nacional de Colombia, facultad de Matemáticas.
- Legrís, J. (2012). “Visualizar y manipular. Sobre el razonamiento diagramático y la naturaleza de la deducción”. *Aparecerá en Visualização nas Ciências Formais.*, comp. por Abel Lassalle Casanave & Frank Thomas Sautter. Londres, College Publications, 89–103.
- Lemon, O. & Pratt, I. (1997). Spatial logic and the complexity of diagrammatic reasoning. *Machine Graphics and Vision*, 6(1):89–108.
- Peirce, C. S., & Eisele, C. (1976). *The new elements of mathematics* (Vol. IV). The Hague: Mouton.
- Peirce, C. S. (1974). *Collected papers of Charles Sanders Peirce* (Vol. V). Harvard University Press.
- Peirce, C. S. (1974). *La ciencia de la semiótica*. Ed. Nueva Visión, Buenos Aires (recopilación).
- Peirce, C. S. (1904). *Qué es el pragmatismo*. Traducción de Norman Ahumada, 2004.
- Peirce, C. S. (1905). *La base del pragmatismo en la “faneroscopia”*. Traducción de Sandra Ollo, 2004.
- Redondo Domínguez, I. (2006). *La comunicación en Charles S. Peirce: análisis de sus textos fundamentales*. Univ. De Navarra.

# Revising logic: Anti-exceptionalism and circularity

Maria Paola Sforza Fogliani

According to anti-exceptionalism (AE) about logic, (i) logical laws do not possess the series of properties that have historically been thought to make the discipline to some degree ‘special’ – *viz.*, *aprioricity* (or *unrevisability*), *analyticity*, *necessity* and *normativity*; rather, logic is in many senses continuous with science. Also, (ii) theory-choice in logic is to be carried out by means of an abductive methodology; that is, logical theories are justified, revised and compared with respect to a set of traditional criteria – *e.g.*, *adequacy to the data*, *simplicity*, *strength*, *elegance* and *fruitfulness*.

We’ll first try to clarify claim (i), by reviewing which properties AEs think logical laws should be deprived of. It will turn out that there is a substantial disagreement on what logic (allegedly) cannot be, the only agreed upon feature being non-apriorism; furthermore, it seems that AEs use ‘*a priori/a posteriori*’ in an unspecific sense – in that they do not make reference to empirical kinds of evidence, but rather equate non-aprioricity with revisability.

We’ll then move on to (ii), and try to unpack the abductive methodology and its criteria. In order to do this, we will first review the main implementations of the AE model – namely, Priest’s [2016] and Williamson’s [2017].

Priest provides a formal model for theory-choice in logic. Consider a set of criteria  $\{c_1, \dots, c_n\}$  and a set of theories  $\{T_1, \dots, T_n\}$ . Let an evaluation scale of a theory  $T$  with respect to a criterion  $c$  be the set  $X = \{x \in \mathbb{R} \mid -10 < x < +10\}$ ; then for every criterion and theory there is a measure function  $\mu_c(T) \in X$ . Since not all criteria might be equally important, we assign a weight to each of them – *i.e.*  $w_c \in X$ . The rationality index of a theory  $T$  is thus defined as:

$$\rho(T) = w_{c_1} \mu_{c_1}(T) + \dots + w_{c_n} \mu_{c_n}(T)$$

Among a set of competing theories, the one displaying the highest rationality index will be the rationally preferable.

Williamson’s account focuses on consequence operators ( $Cn$ ). Let  $\Gamma$  be the set of sentences of a well-confirmed theory (*e.g.*, solid laws of physics), and let a consequence operator  $Cn_1$  for a consequence relation  $\models_1$  and theory  $\Gamma$  be defined as:

$$Cn_1(\Gamma) = \{A \mid \Gamma \models_1 A\}$$

$Cn_1$  takes sentences of  $\Gamma$  as inputs, and outputs  $Cn_1(\Gamma)$  – *i.e.*, the set of sentences that follow from  $\Gamma$  via  $\models_1$ . The model compares consequences relations with respect to how well they do when applied to well-confirmed theories; *e.g.*, a consequence operator yielding, from premises we have grounds to believe, a

sentence we have grounds to reject, will count as evidence against that consequence operator.

Secondly, we'll take a closer look at the abductive criteria, and in particular at the prominent among them – namely, adequacy to data. The so far most detailed account has been presented by Priest, and has it that data for the assessment of a logical theory are provided by our pre-theoretical intuitions about the validity of natural language inferences.

Despite providing a sensible account of theory-change in logic, I will argue that AE has to face a host of objections that stem from a well-know argument in the philosophy of logic – the Centrality Argument (CA; *e.g.*, Putnam [1978]) – which, in a nutshell, runs as follows: logical laws are so central in every rational reasoning that any attempt either to revise or to justify them ends up using those laws themselves and, so, winds up being circular or otherwise illegitimate. We can build versions of CA that are specifically targeted against the AE accounts; moreover, since logical laws are extremely pervasive in reasoning, we expect circularity issues to arise at several different levels of the abductive computation.

Let us start with justification, and consider Priest's implementation of the adequacy to data criterion:

P<sub>1</sub>. If we have favorable intuitions for a logical law  $l$ , then  $l$  is justified;

P<sub>2</sub>. we have favorable intuitions for  $l$ ;

---

C.  $l$  is justified

Obviously, this is a *modus ponens*; so, if  $l$  happened to be *modus ponens* itself, our justification would be circular. As for revision, let us apply CA to Williamson's *Cn*-account:

P<sub>1</sub>. If  $Cn_1(\Gamma)$  includes false sentences,  $\models_1$  has to be revised;

P<sub>2</sub>.  $Cn_1(\Gamma)$  includes false sentences;

---

C.  $\models_1$  has to be revised.

Suppose, moreover, that we could tell the issue with  $Cn_1(\Gamma)$  was specifically caused by *modus ponens*; again, our revision would be circular.

Furthermore, we can device a dull version of the argument that attaches – at a higher level – to the general AE inference:

P<sub>1</sub>. If  $\rho(L_1) > \rho(L_2), \rho(L_3) \dots \rho(L_n)$ , then  $L_1$  is the best theory among the set of logics  $L_1, L_2 \dots L_n$ ;

P<sub>2</sub>.  $\rho(L_1) > \rho(L_2), \rho(L_3) \dots \rho(L_n)$ ;

---

C.  $L_1$  is the best theory among the set of logics  $L_1, L_2 \dots L_n$ .

In case  $L_1$  does not validate *modus ponens*, while its rivals do, we might be in for a clash.

Thus, we appear to be left with two conflicting cases about the revisability of

logic: logical theories have indeed often been revised on the basis of an abductive methodology, which has taken into account their adequacy to our intuitions, their strength and their simplicity; however, the AE account faces some serious threats of circularity.

I will end by proposing a way of reconciling these seemingly opposing intuitions, which appeals to Priest's [2014] distinction between *logica docens* and *logica ens* – that is, between what logicians claim about logic, and what is actually valid. I will argue that AEs seem to submit only that *logica docens* is revisable, while remaining silent on *logica ens*' fate; on the other hand, a minimal version of CA shows only that we can neither revise nor justify the laws of the correct logic – *i.e.*, of *logica ens* – whatever this logic is. I'll thus submit that some compatibility can be worked out, at least between modest versions of the two opposing positions.

## References

- Hjortland O. T., 2017, «Anti-Exceptionalism About Logic», *Philosophical Studies*, 174, 3, pp. 631–658.
- Maddy P., 2002, «A Naturalistic Look at Logic», *Proceedings and Addresses of the American Philosophical Association*, 76, 2, pp. 61–90.
- Priest G., 2014, «Revising Logic», in Rush P. (ed), *The Metaphysics of Logic*, pp. 211–223, Cambridge, CUP.
- Priest G., 2016, «Logical Disputes and the A Priori», *Logique et Analyse*, 59, 236, pp. 347–366.
- Putnam H., 1978, «There Is At Least One A Priori Truth», *Erkenntnis*, 13, 1, pp. 153–170.
- Read S., draft, «Anti-Exceptionalism About Logic», <https://www.st-andrews.ac.uk/slr/Anti-exceptionalism.pdf>.
- Williamson T., 2017, «Semantic Paradoxes and Abductive Methodology», in Armour-Garb B. (ed), *Reflections on the Liar*, pp. 325–346, Oxford, OUP.

# Referencia directa, inocencia semántica y opacidad

José Manuel Viejo

## I

En las últimas décadas, se han propuesto muchas teorías semánticas que se comprometen explícitamente con las dos siguientes tesis:

**Referencia Directa:** El valor veritativo-condicional de nombres propios y deícticos (su contribución a la condición de verdad de las preferencias de los enunciados en que intervienen) es meramente el objeto referido. De otro modo, el valor veritativo-condicional correspondiente a la preferencia de un enunciado simple que contenga nombres propios o deícticos (su condición de verdad) es una *proposición russelliana objeto-dependiente*—un estado de cosas que contiene como constituyentes a los individuos referidos, y no condiciones descriptivas concernientes a ellos. Así (dado que Superman no es otro que Clark Kent), la preferencia de las oraciones siguientes tiene el mismo valor veritativo-condicional, consistente en la proposición russelliana objeto-dependiente (1/2/p):

(1) Superman puede volar,

(2) Clark Kent puede volar,

(1/2/p) <Superman, poder volar>

**Inocencia Semántica:** El valor veritativo-condicional de una expresión al aparecer incrustada en una cláusula-que es su valor veritativo-condicional “usual” (el que habría tenido de no estar incrustada).

Las teorías que se comprometen con Referencia Directa e Inocencia Semántica se enfrentan a un conocido rompecabezas: parecen entrar en conflicto con una tercera tesis, aparentemente verdadera:

**Opacidad:** La sustitución, en una cláusula-que, de un nombre propio o un deíctico por otro con la misma referencia, no siempre preserva el valor de verdad de las preferencias de los enunciados que la contienen.

¿Cómo podrían estas teorías explicar el valor de verdad que intuitivamente corresponde a preferencias de enunciados como (3) y (4)?

(3) Lois cree que Superman puede volar.

(4) Lois cree que Clark Kent puede volar.

El problema debería quedar claro. Queremos ser capaces de reconocer que preferencias de los enunciados (3) y (4) podrían tener diferente valor de verdad, pero ¿cómo podría ser esto posible si dichos enunciados difieren tan solo en que en (3) ‘Superman puede volar’ ha sido reemplazado por una oración, ‘Clark Kent puede volar’, con el mismo valor veritativo-condicional? Ya que Superman = Clark Kent.

El reto consiste en desarrollar una teoría de la semántica de las atribuciones de creencia que nos permita ver cómo Referencia Directa, Inocencia Semántica y Opacidad no son genuinamente inconsistentes entre sí. Uno de los intentos más prometedores de conseguir esto es lo que Schiffer (1992) ha llamado *teoría del deíctico oculto*. Fue formulada por primera vez por Schiffer en 1977, y ha sido vigorosamente defendida por Crimmins y Perry (1989)<sup>1</sup>. La cuestión que pretendo examinar aquí es si realmente esta teoría proporciona la mejor explicación de cómo puede eliminarse la aparente inconsistencia.

El artículo se desarrollará de la siguiente manera. Primero, explicaré brevemente en qué consiste la teoría del deíctico oculto y cuál es su principal motivación. En segundo lugar, haré un repaso breve de sus principales problemas, y finalmente propondré una teoría alternativa de la semántica de las atribuciones de creencia a la que llamaré el *modelo deíctico*, examinaré cómo evita los problemas mencionados, y argumentaré que es la mejor estrategia disponible para disolver la aparente inconsistencia.

## II

La condición de verdad que la teoría del deíctico oculto asigna a las atribuciones de creencia puede representarse de la siguiente manera<sup>2</sup>:

**(TDO)** Una preferencia  $u$  de una oración de la forma ‘ $S$  cree que  $a$  es  $P$ ’ es verdadera syss  $\exists m(\Phi m \ \& \ RTC(S, [a \text{ es } P]_u, m))$ ,

donde  $[a \text{ es } P]_u$  es el valor veritativo-condicional que la oración subordinada ‘ $a$  es  $P$ ’ tiene en  $u$ ; RTC es una relación triádica que se aplica a una triada consistente en  $S$ ,  $[a \text{ es } P]_u$  y un modo de presentación  $m$  syss  $S$  acepta  $[a \text{ es } P]_u$  bajo un modo de presentación  $m$ ; y  $\Phi$  es una clase de modos de presentación, contextualmente determinada y a la que se hace referencia implícitamente, de  $[a \text{ es } P]_u$ .

(TDO) nos dice bastantes cosas. Nos dice que una teoría semántica para la atribución de creencia tiene que interpretar ‘cree’ como un predicado relacional diádico que se da entre creyentes y proposiciones. Nos dice también que la cláusula-que en una preferencia de ‘ $S$  cree que  $a$  es  $P$ ’, ‘que  $a$  es  $P$ ’, es un término referencial cuyo referente es el valor veritativo-condicional que la oración subordinada, ‘ $a$  es  $P$ ’, tiene en dicha preferencia. Y nos dice que una preferencia de un enunciado de atribución de creencia exige la referencia a una clase de modos de presentación contextualmente determinada, y que esta referencia es

<sup>1</sup>Véase también Crimmins 1992.

<sup>2</sup>En este artículo nos centramos en atribuciones de creencia contenedoras de *oraciones subordinadas singulares*, esto es, oraciones de la forma ‘ $a$  es  $P$ ’, donde ‘ $a$ ’ es un término singular –un nombre propio, un deíctico o una descripción definida.

implícita, en tanto que no hay ningún elemento sintáctico en el enunciado que refiera a dicha clase.

He aquí, pues, nuestra situación. Queremos dejar abierta la posibilidad de que preferencias de (3) y (4) puedan diferir en valor de verdad. Al mismo tiempo, queremos hacer esto sin renunciar necesariamente a Referencia Directa y a Inocencia Semántica. Como diría Schiffer, ahora hace su entrada sobre su blanco corcel la teoría del deíctico oculto, que satisface todos los *desiderata*. Incluso si aceptamos Referencia Directa, y asumimos en consecuencia que el valor veritativo-condicional de ‘Superman puede volar’ y de ‘Clark Kent puede volar’ es una y la misma proposición russelliana objeto- dependiente (a saber <Superman, poder volar>), e incluso si, de acuerdo con los dictados de Inocencia Semántica, estos enunciados mantienen su valor veritativo-condicional al aparecer subordinados en (3) y en (4), una preferencia de (3) no tiene por qué tener el mismo valor de verdad que una preferencia de (4). Esto es debido, por supuesto, al rasgo deíctico oculto: en el caso normal, la clase de modo de presentación al que se hace referencia en una preferencia de (3) será diferente de aquel al que se hace referencia en una preferencia de (4). De manera que no hay rompecabezas para la teoría del deíctico oculto.

### III

Lo anterior es lo que puede decirse a favor de la teoría del deíctico oculto. Plantearé ahora tres problemas para la teoría.

*El problema de la forma lógica.* En su famoso artículo de 1992, Schiffer plantea el así llamado problema de la forma lógica en contra de la teoría del deíctico oculto, que puede resumirse en el siguiente argumento<sup>3</sup>:

- (i) Si la teoría del deíctico oculto es correcta entonces ‘modo de presentación *m*’ en una preferencia de
  - (5) Lois cree que bajo un modo de presentación *m*funciona como la especificación del tercer lugar de argumento de una relación triádica.
- (ii) Pero ‘un modo de presentación *m*’ en (5) no parece ser tal cosa (es más bien un modificador adverbial de una relación diádica).
- (iii) Por lo tanto, la teoría del deíctico oculto no es correcta.

Este argumento es claramente válido, y la premisa (i) es claramente verdadera, de modo que falta por ver qué se puede decir a favor de la premisa (ii). La justificación de la premisa (ii) proviene de un criterio sintáctico que distingue entre argumentos y adverbios según que puedan extraerse con o sin ambigüedad de las cláusulas subordinadas dubitativas: cuando no hay ambigüedad nos encontramos ante un argumento, y cuando la hay, ante un adverbio. Así, el carácter de argumento de ‘su hermana’ en una preferencia de

- (6) Juan le dio la casa a su hermana

---

<sup>3</sup>Me baso en el esquema de la formulación del argumento dado por Schiffer (2000:18).

se revela por el hecho de que la oración interrogativa

(7) ¿De quién te preguntabas si Juan le dio el libro?

tiene una única lectura: solo puede estar preguntando acerca de quién es la persona que uno piensa que es el beneficiario de la generosidad de Juan. Por el contrario, el carácter adverbial de ‘un modo de presentación  $m$ ’ en una preferencia de (5) se revela en el hecho de que la oración interrogativa

(8) ¿Bajo qué modo de presentación te preguntabas si Lois cree que Superman puede volar?

claramente puede estar preguntando cuál es el modo de presentación bajo el cual uno se pregunta si Lois cree que Superman puede volar, o puede estar preguntando cuál es el modo de presentación bajo el que Lois cree que Superman puede volar.

*El problema de la restricción lingüística.* En su capítulo “Intensional Contexts” (2014), Michael Nelson presenta lo que llamaré el problema de la restricción lingüística en contra de la teoría del deíctico oculto. Si la teoría del deíctico oculto es correcta, entonces la preferencia de un enunciado de atribución de creencia exigiría hacer referencia a una clase de modos de presentación  $\Phi$ , donde  $\Phi$  sería un constituyente de la condición de verdad de la preferencia tal que no habría ninguna expresión en el enunciado proferido que contribuyera esa clase como su valor veritativo-condicional. El problema aquí es que esto es incompatible con lo que muchos consideran una restricción importante en los efectos que el contexto puede tener sobre la condición de verdad de una preferencia, a saber, es incompatible con lo que Nelson llama el *principio de restricción lingüística*, de acuerdo con el cual todos los efectos del contexto sobre la condición de verdad de la preferencia de un enunciado son siempre rastreables hasta algún elemento sintáctico del enunciado proferido; esto es, los efectos del contexto en la condición de verdad de una preferencia se restringen a asignar valores veritativo-condicionales a los elementos sintácticos del enunciado proferido, donde cada elemento trae consigo reglas que gobiernan qué y qué no puede asignarle el contexto. A pesar de que estas cuestiones son complejas y rápidamente nos llevan a cuestiones delicadas en los debates contemporáneos de filosofía del lenguaje, una razón para aceptar el principio de restricción lingüística es que ayuda a eliminar información pragmáticamente comunicada mediante una preferencia pero que no entra a formar parte de la condición de verdad de la misma, evitando con ello problemas de sobregeneración que amenazan a otras teorías que, como la teoría del deíctico oculto, violan este principio. Alguien, por ejemplo, puede proferir la oración ‘ $2+2=4$ ’ para comunicar algo que es verdadero si Pepito está yendo al estanco. Incluso entonces, parece altamente implausible que la preferencia misma tenga dicha condición de verdad, ya que nada en el enunciado proferido contribuye los elementos relevantes de la información pragmáticamente comunicada.

*El problema de las atribuciones de re.* Si la teoría del deíctico oculto es correcta, entonces, para que una preferencia  $u$  de un enunciado de atribución de creencia sea verdadera es necesario que el sujeto de la atribución acepte una proposición que tenga la misma condición de verdad que la que tiene la oración subordinada en  $u$ . Pero esto es falso, tal como pone de manifiesto el siguiente



ejemplo<sup>4</sup>. Supongamos que Alicia, sobre bases completamente generales, sabe que exactamente un estudiante obtuvo un 10 en el examen (supongamos que el profesor ayudante de Alicia se lo ha dicho mientras le hacía saber el fracaso que ha supuesto el esfuerzo realizado). Reflexionando sobre la dificultad de examen, Alicia profiere:

(9) El estudiante que obtuvo un 10 en el examen es un genio.

Considero que es bastante claro que, al proferir (9), Alicia está utilizando atributivamente la descripción definida ‘el estudiante que obtuvo un 10’, y que, en consecuencia, el valor veritativo-condicional de esta preferencia es una proposición russelliana objeto-independiente. Ahora supongamos que el compañero de Alicia, Nicolás, escucha a Alicia proferir (9). Nicolás sabe que el estudiante que obtuvo un 10 es su hermano, Javier. En base a esto, Nicolás le dice a su hermano:

(10) Alicia cree que tú eres un genio,

donde ‘tú’ es claramente un término singular directamente referencial, y, en consecuencia, el valor veritativo-condicional correspondiente a la oración subordinada es una proposición russelliana objeto-dependiente (i.e., <Javier, ser un genio>). La atribución de creencia parece verdadera en un caso como éste. Pero, ¿cómo puede el teórico del deíctico oculto explicar esto? De acuerdo con la teoría del deíctico oculto, para que una preferencia de (10) sea verdadera es necesario que Alicia acepte la proposición que es el valor veritativo condicional de la oración subordinada, a saber, la proposición russelliana objeto-dependiente <Javier, ser un genio>. No obstante, parece claro que Alicia no acepta la proposición russelliana objeto-dependiente en cuestión. Alicia, después de todo, solo parece tener creencias objeto-independientes sobre Javier, ya que su única concepción de esta persona lo representaría por medio de una descripción definida usada atributivamente. Así, es difícil ver cómo el teórico del deíctico oculto puede explicar la intuición de que una preferencia de (10) es verdadera en la situación descrita.

## IV

Formularé ahora una teoría alternativa, a la que llamaré el *modelo deíctico*, que reúne todas las virtudes de la teoría del deíctico oculto a la vez que evita sus problemas. Primero, dejadme introducir la noción de dominio intencional adoptando la siguiente definición:

**Dominio intencional:** Dada una preferencia  $u$  de un enunciado de la forma ‘ $a$  es  $P$ ’, el dominio intencional de  $u$  consiste en toda proposición  $PROP$  tal que (i)  $PROP$  es una proposición con la misma condición de verdad que la preferencia  $u$  de ‘ $a$  es  $P$ ’, o (ii)  $PROP$  es directa o indirectamente acerca del objeto designado por ‘ $a$ ’ en el contexto de  $u$  y  $PROP$  entraña atribuirle a dicho objeto la propiedad de ser  $P$ .

<sup>4</sup>El de Alicia es una modificación de un ejemplo de King (2001). Para más ejemplos de este tipo, véase Bach 2012: 45; Eaker 2002, 2009; Kaplan 1989: 555, n. 72. See also Kaplan 2013: 33 and Stalnaker 2009: 245.

La idea es que una proposición *PROP* es directamente acerca de algún individuo si es una proposición objeto-dependiente cuya condición de verdad es un estado que contiene a ese individuo como uno de sus constituyentes. Por otro lado, una proposición *PROP* es indirectamente acerca de algún individuo si es una proposición objeto-independiente cuya condición de verdad es un estado de cosas que contiene una condición descriptiva que es satisfecha por ese individuo.

Equipados con la noción de dominio intencional, el modelo déictico se puede formular de la siguiente manera:

(MD) Una preferencia  $u$  de un enunciado de la forma ‘ $S$  cree que  $a$  es  $P$ ’ es verdadera syss  $RDC(S, SCSDI-[a \text{ es } P]_u)$ ,

donde  $SCSDI-[a \text{ es } P]_u$  es la subclase contextualmente saliente de las proposiciones que constituyen el dominio intencional que ‘ $a$  es  $P$ ’ tiene en  $u$ , y donde  $RDC$  es una relación diádica que se aplica a una diada consistente en  $S$  y  $SCSDI-[a \text{ es } P]_u$  syss  $S$  acepta alguna proposición específica de  $SCSDI-[a \text{ es } P]_u$ . De modo que:

$RDC(S, SCSDI-[a \text{ es } P]_u)$  syss  
 $\exists p(p \in SCSDI-[a \text{ es } P]_u \ \& \ RDC^*(S, p))$ ,

donde  $RDC^*$  es una relación diádica que se aplica a una diada consistente en  $S$  y alguna proposición específica syss  $S$  acepta dicha proposición.

(MD) nos dice bastantes cosas. Nos dice que una teoría semántica para la atribución de creencia tiene que interpretar ‘cree’ como un predicado relacional diádico que se da entre creyentes y (clases de) proposiciones. Nos dice también que una preferencia de un enunciado de la forma ‘ $S$  cree que  $a$  es  $P$ ’ exige la referencia a una clase contextualmente saliente de proposiciones, y que esta referencia contextualmente determinada a una clase de proposiciones es “déictica” en tanto que hay una expresión en el enunciado que trae consigo dicha referencia, a saber, la cláusula-que, ‘que  $a$  es  $P$ ’.

El modelo déictico nos permite ver cómo una preferencia de (3) puede diferir en valor de verdad con una preferencia de (4), incluso si, *ex hypothesi*, ‘Superman puede volar’ en (3) tiene el mismo valor veritativo-condicional que ‘Clark Kent puede volar’ en (4). Que preferencias de (3) y (4) puedan diferir en valor de verdad se debe a la naturaleza déictica de las cláusulas-que: en el caso normal, la clase de proposiciones a la que se hace referencia mediante el uso de ‘que Superman vuela’ en una preferencia de (3) será diferente de la clase a la que se hace referencia mediante el uso de ‘que Clark Kent vuela’ en una preferencia de (4).

Ahora estamos en posición de ver cómo se las arregla el modelo déictico con los tres problemas que ha de enfrentar la teoría del déictico oculto.

*El problema de la forma lógica.* El modelo déictico no tiene ningún problema con la forma lógica, ya que representa ‘cree’ en una preferencia  $u$  de un enunciado de la forma ‘ $S$  cree que  $a$  es  $P$ ’ como lo que parece ser: un predicado relacional diádico con dos lugares de argumento, uno para el referente de ‘ $S$ ’ en  $u$ , y otro para el referente de ‘que  $a$  es  $P$ ’ en  $u$ .

*El problema de la restricción lingüística.* El modelo déictico ofrece un resultado particularmente satisfactorio para el problema de la restricción lingüística. La

teoría del deíctico oculto es incompatible con el principio de restricción lingüística, ya que exige la referencia a una clase de modos de presentación contextualmente determinada, y no hay ningún elemento sintáctico en el enunciado de atribución de creencia que lleve a cabo dicha referencia. Cuando pasamos al modelo deíctico, la cláusula-que ‘que *a* es *P*’ en una proferencia *u* de un enunciado de la forma ‘*S* cree que *a* es *P*’ se interpreta como haciendo referencia a una clase contextualmente determinada de proposiciones. La referencia a una clase de proposiciones es deíctica –está contextualmente determinada– pero no es “oculta”; es el valor veritativo- condicional de una expresión contenida en el enunciado de atribución de creencia, a saber, la cláusula-que, ‘que *a* es *P*’. De manera que, si se acepta el modelo deíctico, no hay ninguna necesidad de rechazar el principio de restricción lingüística, evitando así problemas de sobregeneración no deseados.

*El problema de las atribuciones de re.* La teoría del deíctico oculto no puede explicar la intuición de que la proferencia de (10) llevada a cabo por Nicolás puede ser verdadera incluso si Alicia no acepta ninguna proposición objeto- dependiente sobre Javier, sino una objeto-independiente. Pero el modelo deíctico no tiene ninguna dificultad con ejemplos como éste: la clase de proposiciones referida por ‘que tú eres un genio’ en la proferencia (10) realizada por Nicolás sería simplemente la clase que constituye el dominio intencional que ‘tú eres un genio’ tiene en dicha proferencia. Como esa clase incluye no sólo proposiciones objeto- dependientes, sino proposiciones-objeto- independientes concernientes a Javier, y como aceptar una clase de proposiciones es aceptar al menos una proposición específica de esa clase, una proferencia de (10) puede ser verdadera en virtud de que Alicia acepte solo alguna de las proposiciones objeto-independientes. En otras palabras, lo que Nicolás dice al proferir (10) se interpreta en el modelo deíctico como diciendo simplemente que Alicia acepta alguna proposición, la que fuera, que es directa o indirectamente acerca de Javier y entraña atribuirle la propiedad de ser un genio. De manera que la atribución podría ser verdadera si Alicia acepta, por ejemplo, una proposición que es verdadera syss el estudiante que obtuvo un 10 en el examen es un genio.

## Bibliografía

- Bach, K. 2012. ‘Getting a thing into a thought’, in R. Jeshion (ed.). *New Essays on Singular Thought*, (Oxford: Oxford University Press).
- Crimmins, M. 1992. *Talk about Beliefs*. (Cambridge, Mass: MIT Press).
- Crimmins, M. and J. Perry. 1989. ‘The prince and the phone booth: Reporting puzzling beliefs’. *Journal of Philosophy*, 86:685–711.
- Eaker, E. 2002. *Reporting Attitudes*. Ph.D dissertation, UCLA.
- Eaker, E. 2009. ‘A new starting place for the semantics of belief sentences’, in J. Almog and P. Leonardi (eds), *The Philosophy of David Kaplan*, (Oxford: University Press), 208–232.
- Kaplan, D. 1989. ‘Demonstratives’, in J. Almog, J. Perry, and H. Wettstein (eds.). *Themes from Kaplan*, (Oxford: Oxford University Press), 481–563.

- Kaplan, D. 2013. ‘*De re* belief’, in R. Hull (ed.). *The American Philosophical Association Centennial Series. Volume 9. Presidential addresses of The APA 1981-1990*, (Kluwer Academic Publishers), 25–37.
- King, J. 2001. *Complex Demonstratives: A Quantificational Account*. (Cambridge, MA: MIT Press).
- Nelson, M. 2014. ‘Intensional contexts’, in M. García-Carpintero and M. Kölbel (eds.), *The Bloomsbury Companion to the Philosophy of Language*, (New York)
- Schiffer, S. 1977. ‘Naming and knowing’. *Midwest Studies in Philosophy*, 2:28–41.
- Schiffer, S. 1992. ‘Belief ascription’. *Journal of Philosophy*, 89:499–521.
- Schiffer, S. 2000. ‘Propositional attitudes in direct-reference semantics’, in K. M. Jaszczolt (ed.), *The Pragmatics of Propositional Attitude Reports*, (Oxford: Elsevier Science),
- Stalnaker, R. 2009. ‘What is *de re* belief’, in J. Almog and P. Leonardi (eds), *The Philosophy of David Kaplan*. (Oxford: University Press), 208–232.

# A stochastic process explanation of conditionals

Anna Wójtowicz  
Krzysztof Wójtowicz  
Institute of Philosophy  
Warsaw University

Any formal account of evaluating probabilities of conditional sentences has (at least) two important problems to solve:

1. Accounting for Lewis' triviality arguments, concerning the PC=CP thesis (i.e. *The probability of the conditional is conditional probability*).
2. Understanding and incorporating different interpretations of the conditional into the formal model.

The classic distinction are the subjunctive *versus* indicative mode, as exemplified by the famous Oswald-Kennedy sentences [Adams 1970]. [Khoo 2015] considers epistemic *versus* metaphysical interpretations, and [Kaufmann 2004, 2005, 2009, 2015] distinguishes global *versus* local interpretations. Leaving aside the relationships between these three sets of concepts, we choose the last one as the most natural one for our study.

In our account, the intuitive illustration of the conditional is given in terms of a game. Its rules depend on the interpretation of the conditional (local *versus* global). Formally it is represented as a Markov process, which gives a very simple method of evaluating the probabilities.

## The one-urn case

A ball is drawn from the urn containing 10 White, 8 Green and 2 Red balls. What probability is a rational subject going to ascribe to the sentence *If the ball is not White, it is Green* ( $\neg W \rightarrow G$ )?

We can think of this probability in terms of estimating the price for playing a certain fair game—The Conditional Game. Obviously, before we bet, we have to define the rules. It is clear, that:

- if we draw a Green – we win;
- if we draw a Red – we lose.

But what happens when we draw a White? There are four *a priori* possibilities:

- (1) We win.

- (2) We lose.
- (3) The game is undecided and stops.
- (4) We restart the game, i.e. put the ball into the urn and draw the ball again.

Choices (1) and (2) would amount to identifying the conditional with a Boolean combination of W,G,R, which is not acceptable. Option (3) “annihilates” the problem of probabilities of conditionals (at least for this approach). So the only reasonable (and quite standard (cf. [van Fraassen 1976]), definition of The Conditional Game  $\neg W \rightarrow G$  is (4).

We can represent the flow of the game in a natural and convenient way as a stochastic process. It gives rise to an appropriate probability space, which resembles van Fraassens and Kaufmann’s constructions ([van Fraassen 1976], [Kaufmann 2004, 2005, 2009, 2015]), but is much simpler and easy to handle. Importantly, an event in this space corresponds to the conditional  $\neg W \rightarrow G$ —and the probability of this event is exactly the initial conditional probability  $P(G|\neg W)$ . So the PC=CP principle is given a precise form, and becomes a theorem within our formalization.

## The two-urn case. Global and local interpretations of the conditional

The following example is “isomorphic” to the example analyzed in [Kaufman 2004] and [Khoo 2016]:

*There are two urns:  $U1$ ,  $U2$ . In urn  $U1$  there are 2 White, 9 Green and 1 Red ball. In the urn  $U2$  there are 50 White, 1 Green and 9 Red balls. Assume that urn  $U1$  can be chosen with probability  $3/4$  and urn  $U2$  with probability  $1/4$ .*

What is the chance of winning The Conditional Game  $\neg W \rightarrow G$ ? We need a precise description of the rules of the game. Obviously, an urn must first be chosen and then a ball from this urn has to be drawn. Again:

- if we draw a Green – we win;
- if we draw a Red – we lose.

but what happens when we draw a White? In the formal model, we consider two possible continuations, which correspond to the global and local interpretation of the conditional. They are formally captured in terms of Markov processes, giving rise to appropriate probability spaces, where the probabilities of the (suitably interpreted) conditional  $\neg W \rightarrow G$  can be computed in a simple way. Importantly, they correspond to the two intuitive (but not formally correct) formulas from [Kaufmann 2004]:

$$P(\neg W \rightarrow_g G) = P(\neg W \rightarrow G|I)P(I|\neg W) + P(\neg W \rightarrow G|II)P(II|\neg W) = 0,6$$

$$P(\neg W \rightarrow_l G) = P(\neg W \rightarrow G|I)P(I) + P(\neg W \rightarrow G|II)P(II) = 0,3$$

for the global and local interpretations of the conditional.

## Lewis's triviality result: PC = CP?

The PC=CP thesis was considered to be the expression of certain intuitions, which are often presented by quoting the famous passage from Ramsey:

*If two people are arguing 'If p will q?' and both are in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q... We can say that they are fixing their degrees of belief in q given p.* (Ramsey 1929, 247).

We can say, that it expresses the view, that the probability of the conditional is dependent on the conditional probability/ies. So the PC=CP principle must be reformulated in appropriate ways (depending on the interpretation of the conditional). The appropriate reformulation(s) of the PC=CP principle(s) allows to give their formal justification(s).

Our considerations have also an important metaphilosophical aspect: they contribute to the vivid discussion concerning the explanatory (explicatory) role of mathematics. The discussion on mathematical explanations is vast (see e.g. [Lange 2013]). Our account exemplifies a general phenomenon: notions and results from the theory of Markov processes have been used to explain some philosophically important issues. We show how to give formal, precise explications of the probability of conditionals, taking different interpretations and variants into account. So our account can be viewed as a contribution to the idea of mathematical philosophy ([Leitgeb 2013]).

**Acknowledgment:** The preparation of this paper was supported by a NCN grant 2016/21/B/HS1/01955.

## References

- Adams, E. (1965). "The logic of conditionals", *Inquiry*, 8, 166–197.
- Adams, E. W. (1970). "Subjunctive and Indicative Conditionals", *Foundations of Language*, 6, 89–94.
- Adams, E. (1975). *The Logic of Conditionals*, Reidel.
- Hajek, A. (2011). "Triviality pursuit", *Topoi*, 30(1), 3–15.
- Kaufmann, S. (2004). "Conditioning against the grain: abduction and indicative conditionals", *Journal of Philosophical Logic*, 33(6), 583–606.
- Kaufmann, S. (2005). "Conditional predictions: a probabilistic account", *Linguistics and Philosophy*, 28(2), 181–231.

- Kaufmann, S. (2009). "Conditionals right and left: probabilities for the whole family", *Journal of Philosophical Logic*, 38, 1–53.
- Kaufmann, S. (2015). "Conditionals, conditional probability, and conditionalization". In: Zeevat, H., Schmitz, H.C. (eds.). *Bayesian natural language semantics and pragmatics*. Springer, 71–94
- Khoo, J. (2015). "On Indicative and Subjunctive Conditionals", *Philosopher's Imprint*, 15(32), 1–40.
- Khoo, J. (2016). "Probabilities of conditionals in context", *Linguistics and Philosophy*, 39, 1–43.
- Lange M. (2013). "What Makes a Scientific Explanation Distinctively Mathematical?", *British Journal for the Philosophy of Science*, 64(3), 485–511.
- Leitgeb, H. (2013). "Scientific Philosophy, Mathematical Philosophy, and All That", *Metaphilosophy*, 44(3), 267–275.
- Lewis, D. (1976). "Probabilities of conditionals and conditional probabilities", *Philosophical Review*, 85, 297–315.
- Ramsey, F.P. (1929). "General propositions and causality", Reprinted in *Philosophical Papers*, Mellor, D. H. (ed.) Cambridge: Cambridge University Press, 1990, 145–163.
- Van Fraassen, B.C. (1976). "Probabilities of conditionals", in: Harper, W.L., Stalnaker, R., Pearce, G. (eds.) *Foundations of probability theory, statistical inference, and statistical theories of science*, The University of Western Ontario Series in Philosophy of Science, vol. 1, D.Reidel, 261–301.



# FILOSOFÍA DEL LENGUAJE

# The Russell, the John, and the Katherine: The-defenders of the-Predicativism

Eno Agolli

Recently, the philosophical world has witnessed a lively debate about the meaning of proper names. Most philosophers have thus far considered predicative uses of names as deviations from what they regard as their standard uses. In standard uses, they say, names have a kind of meaning radically different from that of predicates—in standard uses, names are singular, referential terms. Here, joining a minority of philosophers of language and linguists, I argue that this is wrong. Names are predicates, and can be argued to be predicates in all of their uses in natural language despite recent objections.

In §I, I briefly outline the two views about names under discussion. The standard view, Referentialism, claims that the main function of a name in language is to refer to an individual (directly or otherwise). As for predicative uses of names, Referentialism postulates a kind of semantic ambiguity. Another view, the one I endorse, Predicativism, denies this. Predicativism claims that names do not refer and are predicates everywhere they occur, thus avoiding the need to postulate semantic ambiguity.

In §II, I give the details of Predicativism. I outline its main motivation, which is that names and common count nouns have exactly the same syntax in natural languages. I then define their meaning as predicates: a name *N* is satisfied by an individual *x* iff *x* is called *N*. Finally, I outline the linguistic mechanism via which names can be used to refer, as I admit they do, to a thing (rigidly): names refer (rigidly) by being the predicative component of incomplete definite descriptions. The version of Predicativism that embraces this last idea is known as the- Predicativism, and it will be the one I endorse.

In §III, I consider a semantic objection to the-Predicativism: two sentences which, according to Predicativism, have exactly the same semantics, though different syntax, may exhibit a feature which is not predicted by the-Predicativism semantically: they can make salient either the name of an individual or the individual itself. I argue that this is a pragmatic phenomenon, however, captured and explained by the pragmatic notion of topic or, as I call it, pragmatic stress, which I define as what is relevant or important in the context of utterance of a sentence. The difference in pragmatic stress of such sentences, I argue, is tracked by their different syntax.

In §IV, I use the notion of pragmatic stress to address two crucial objections. According to the-Predicativism, the definite article is used with names according to a certain syntactic rule. This syntactic rule (due to Fara (2015)) is what sustains the main motivation behind the-Predicativism. The syntactic objections

are counterexamples to that rule. In response, I treat both of these counterexamples in a uniform manner: I show that both involve pragmatic stress on the name, rather than the individual itself. I then recommend a modification of the original syntactic rule that accounts for pragmatic stress and thus accommodates the relevant counterexamples.

# Actos de habla y estatus social. Cómo el lenguaje puede convertirse en un estigma

José Alhambra Delgado

El tema general de mi trabajo es poner en común los resultados de un tipo de sociología que denominaré *interaccionista* (i.e. que estudia las relaciones entre individuos con un determinado estatus social y en un marco de interacción delimitado) con algunas concepciones de la filosofía del lenguaje, en concreto del ámbito de la pragmática filosófica. Dentro de este planteamiento general, el tema en torno al cual girará mi ponencia es el siguiente: estudiar el fenómeno social del estigma en el ámbito de los intercambios lingüísticos. Para ello me apoyaré principalmente en dos fuentes: por un lado, las investigaciones de E. Goffman<sup>1</sup> acerca de la marginalidad social llevadas a cabo en la década de los 50; por otro, las investigaciones filosóficas de J. L. Austin<sup>2</sup> y de L. Wittgenstein<sup>3</sup>. Siguiendo con este esquema dual, estructuraré la exposición a partir de dos tesis complementarias, una de *ida* y otra de *vuelta*. La primera, que da título a la ponencia y que presentaré al comienzo de la misma, es la siguiente: el lenguaje es un elemento fundamental de estigmatización social. La generalidad de los términos en los que enuncio esta primera proposición es intencionada: el desarrollo de la exposición tendrá como objetivo explicar las nociones que la componen y proporcionar un mapa conceptual que permita una interpretación más precisa.

El itinerario que seguiré para realizar esta aclaración conceptual se articula a partir de tres bloques. El primero de ellos estará dedicado a la noción de estigma social. El estigma es un cierto atributo, rasgo o cualidad que descalifica al individuo o grupo de individuos que lo poseen, poniendo de manifiesto una incongruencia entre dos dimensiones de la identidad, siempre en un contexto de interacción determinado. Estas dimensiones de la identidad son, por un lado, la «identidad social virtual» (i.e. el estereotipo en el que, en un primer momento, clasificamos a los individuos con los que interactuamos) y, por el otro, la «identidad social real» (i.e. las características que realmente posee ese individuo, que pueden o no coincidir con la identidad social virtual). A su vez este marcador social descalificador puede ser de diferente naturaleza: física, conductual, simbólica, etc. Por poner algunos ejemplos: que se descubra un pasado ligado a la prostitución es una deshonra para una mujer que desempeña el rol de madre *responsable* en una familia de clase media; es una mancha en la biografía de un individuo haber pasado por un hospital psiquiátrico o por la prisión, y la desviación de la norma sexual vigente es considerada un estigma en algunas sociedades.

---

<sup>1</sup>Erving Goffman: *Estigma: la identidad deteriorada*, Buenos Aires, Amorrortu, 2006.

<sup>2</sup>John L. Austin: *Cómo hacer cosas con palabras*, Barcelona, Paidós, 1998.

<sup>3</sup>Ludwig Wittgenstein: *Investigaciones filosóficas*, Madrid, Gredos, 2009.

Así, partiendo de esta primera aproximación, en mi exposición presentaré estos conceptos e introduciré otros como «esferas de interacción» e «identidad personal», «desacreditado-desacreditable» o «modo de encubrimiento-modo de enmascaramiento», que me permitirán mostrar el carácter relacional del estigma y su vinculación, por un lado, a la acción del individuo y, por otro, al contexto en el que se lleva a cabo tal (inter-)acción.

El segundo paso consistirá en explicar *grosso modo* la concepción del lenguaje desde la que trabajo. Esta tiene como trasfondo algunos elementos de la filosofía del lenguaje del segundo Wittgenstein (e.g. «modos de vida», «significado como uso» o las cuestiones en torno al problema de seguir una regla), pero sobretudo se apoya en la noción austiniana de «acto de habla» y en el carácter convencional de la «fuerza ilocucionaria» que está a la base de las denominadas «condiciones de felicidad». Precisamente esta dimensión convencional de los actos de habla me servirá en la tercera y última fase como punto de enlace entre estigma y lenguaje (entendido ahora como herramienta de acción y valoración social). De esta forma intentaré mostrar cómo, a partir de la combinación de diferentes tipos de esferas de interacción y de estatus sociales, la realización de ciertos actos de habla queda vetada a algunos individuos en función de su posición marginal en el intercambio lingüístico; dicho de otra manera: mostraré cómo algunos actos ilocucionarios se convierten en infortunios debido a la entrada en escena del estigma. Para ilustrar esto se podría poner el ejemplo paradigmático de la incapacidad del soldado raso para dar órdenes efectivas («malas apelaciones» en palabras de Austin). No obstante, los casos que me interesan son aquellos en los que, como ocurre con el estigma, el acto ilocucionario hace referencia a ciertas formas de clasificación de los roles asumidas por los agentes en el proceso de socialización y no tanto a instituciones con un aparato explícito de normas regulativas; lo que, a su vez, conlleva la modificación de la noción de «convención de la fuerza ilocucionaria» mencionada por Austin. Algunos ejemplos de este tipo podrían ser los siguientes: la ausencia de fuerza ilocucionaria en los juramentos de un expresidiario en ámbitos en que esta característica es una lacra; la nulidad de las reivindicaciones en el espacio público de un individuo que pertenece a un grupo social estigmatizado, o el infortunio, y la conveniente falta de credibilidad, de las predicciones de alguien cuyo estatus social no se corresponde con la esfera de interacción en la que las emite (e.g. el camarero haciendo predicciones macroeconómicas). Así, en este punto podríamos reformular la primera tesis de la siguiente manera: los actos de habla producen y reproducen las características de los fenómenos de estigmatización social; o dicho con otras palabras: el éxito o el fracaso de determinados actos de habla está mediado por, o depende directamente de, la posición social del agente y, por ende, de su condición de estigmatizado.

Finalmente, presentaré la segunda tesis que mencionaba al comienzo, la cual hará las veces de complementaria de la proposición principal y de conclusión de la ponencia. Se podría enunciar de la siguiente manera: el carácter práctico o convencional que caracteriza las concepciones del significado que he utilizado para elaborar este esquema conceptual debería ser interpretado a partir de los resultados del tipo de sociología que he denominado interaccionista. De esta forma quedará esbozado el doble movimiento al que aludía al principio: por un lado, la defensa de los conceptos de la pragmática en el estudio del estigma y, por el otro, la asunción de nociones tales como rol social, esferas de interacción

o elementos de legitimidad y desprestigio en el análisis filosófico.

# The nature of (covert) dogwhistles<sup>1</sup>

**Manuel Almagro**

**José R. Torices**

University of Granada, Department of Philosophy I

malmagro@ugr.es

jrtorices@ugr.es

## Introduction

‘Dogwhistle’ refers to a kind of political manipulation that some people carry out for political gains. According to Saul (2018), dogwhistles can be either *intentional* or *unintentional* depending on whether the speaker carried out the dogwhistle deliberately or not—although one cannot always recognize whether a particular case was intentional. In addition to being intentional or not, dogwhistles can also be *overt* or *covert* depending on whether the audience is aware or not of the dogwhistle.<sup>2</sup> In the case of overt dogwhistles, the speaker addresses a message to an audience with two possible interpretations. One of these is coded and affects only a subset the audience (Witten 2014). Covert dogwhistles, on the other hand, are not really about sending a “coded message.” Instead, they raise attitudes to salience, so people will act on them without realizing they are being moved on them. As Stanley (2015) points out, these kinds of dogwhistles work as a strategy for undermining democratic ideals without immediate rejection. Our key question is whether covert dogwhistles constitute a special form of implicit communication or whether they can be reduced to already existing forms of implicit communication such as presuppositions or implicatures. We will focus on covert dogwhistles because they seem more difficult to accommodate within the traditional categories of presuppositions and implicatures. To carry out this task, we compare the features of each of the mentioned phenomena and analyse how they behave in the face of retraction.

---

<sup>1</sup>This work was partially funded by the research project “Expresivismos contemporáneos y la indispensabilidad del vocabulario normativo: alcance y límites de la hipótesis expresivista” (MINECO: FFI2016-80088-P), the project “FiloLab-UGR” (UCE-PP2017-04), the FPI Grant (BES-2014-067584) and the FPI Grant (BES-2017- 079933).

<sup>2</sup>The audience to which the dogwhistle is addressed is not necessarily *homogeneous*. Within the audience, we can find subgroups of hearers. That is why it is at least conceptually possible to think that the same dogwhistle can be *overt* and *covert* at the same time for the different groups that constitute the audience.

## The Food Stamp President

According to Neubeck and Cazenave (2001, p. 3), American politicians have long “forged and exploited the link between ‘race’ and ‘welfare’ [or ‘food stamp’]<sup>3</sup> to such a degree that the two terms are now politically and culturally inextricable.” Other authors have also echoed the hidden racial dimension behind the use of the term ‘welfare’, as well as its possible effects on the political attitudes of both African American and White American voters (see e.g. Quadagno 1994, Gilens 1996, Mendelberg 2001, Valentino et al. 2002). They consistently conclude that the expression ‘welfare’ and ‘food stamp’ has racial connotations, mostly for White American voters, which affect their political attitudes without them noticing.

In 2012, Newt Gingrich, one of the candidates for the Republican Party presidential nomination, had to deal with the accusations of racism that a lot of people levelled at him for calling President Obama “the most successful *food stamp president* in American history.” Here’s what he literally said in one of his speeches during the campaign:

Over here you have a policy which, with Reagan and me as speaker, created millions of jobs—it’s called paychecks. Over [t]here you have the most successful *food stamp president* in American history, Barack Obama. (Newt Gingrich, quoted in Elliott, 2012)

Weeks after his statements, the American civil rights activist Al Sharpton, who hosts the political talk show *PoliticsNation with Al Sharpton*, interviewed Newt Gingrich. In the interview, Al Sharpton accuses Gingrich of making a racist use of language in his criticism of President Obama’s social policies, calling Barack Obama the “food stamp president”:

Mr. Gingrich, [...] I have to ask you this kind of talk, this kind of language you use, is it just playing to the right wing, just playing to the far-right as David Gregory ask you with racially tinged language? Do you still defend now what you said? (Al Sharpton, 2012, MSNBC)

Let us put Gingrich’s words into context to understand why the accusations made against him, in this particular case by Al Sharpton, make sense. Although the data show the opposite, a significant portion of the White American population thinks that those who benefit most from food stamps are African Americans.<sup>4</sup> This general misperception serves as a ground for Gingrich’s statements to trigger the implicit biases of that part of the audience that is unknowingly racist. But there are not only implicit biases at work here, there are also many explicit racial prejudices that are simply not seen as such by those who harbour them.

---

<sup>3</sup>“Food stamps” is the popular term for what is currently known in the United States as the Supplemental Nutrition Assistance Program (SNAP).

<sup>4</sup>According to survey data from Delaney and Edwards-Levy for HuffPost, in 2018, 59 % of Americans thought that the most significant welfare recipients were either African Americans or Whites and African Americans alike. However, according to data provided by the same article, 36.2 % of the recipients were white, compared to 25.6 % who were African American. In Al Sharpton’s interview with Gingrich, the data that appear are those provided by the U.S. Census Bureau in 2012. According to this study, 46 % of food stamp recipients are White, compared to 26 % who are African American and 20 % who are Hispanic.



As Henry and Sears (2002) note, the racism of those who believed that African Americans were biologically inferior and supported segregation, lynching and others forms of racial discrimination is virtually non-existent in the United States.<sup>5</sup> However, this old-fashioned racism has been replaced by new forms of racism that Henry and Sears (2002) call *symbolic racism*.<sup>6</sup> They characterise it in the following way:

[A] coherent belief system combining the following ideas: that racial discrimination is no longer a serious obstacle to blacks' prospects for a good life; that blacks' continuing disadvantages are due to their own unwillingness to take responsibility for their lives; and that, as a result, blacks' continuing anger about their own treatment, their demands for better treatment, and the various kinds of special attention given to them are not truly justified. (Henry and Sears, 2002, p. 254)

In such a context, a significant part of Gingrich's audience either endorses symbolic racism or harbours implicit racial biases or both. These attitudes, therefore, are the driving force behind the mobilization of the conservative vote for racist reasons even when this happens, as we have said, surreptitiously. So, this is a clear case of a covert dogwhistle.

## Presuppositions and Implicatures

Covert dogwhistles seem to convey a kind of *not-at-issue* information, i.e., a sort of information that is not explicitly uttered by the speaker (Tonhauser 2011, Stanley 2015). Commonly, presuppositions and implicatures are conceptual tools used to account for how different kinds of implicit communication work. Let's then see whether dogwhistles are reducible to any of them.

A presupposition is a proposition implicitly assumed by uttering a sentence. The truth of the implicit assumption is taken for granted in uttering the sentence that triggers it, and that proposition implicitly assumed is not part of the main propositional content of the speech act (Stalnaker 1973, von Stechow 2004, Potts 2005). For instance, the utterance of sentence **(1)** presupposes the proposition expressed by **(2)**, and normally the later cannot be rejected without affecting the meaning of the former. It is important to note that negating **(1)** does not change its presupposition.  $\neg$ **(1)** also presupposes the proposition expressed by **(2)**.

**(1)** Gingrich's speech has nothing to do with race.

**(2)** Gingrich made a speech.

On the other hand, implicatures are propositions suggested, but not strictly implied, by an utterance. For example, the utterance of **(3)** suggests that Gingrich

---

<sup>5</sup>After Trump's victory in the last election, this assertion may no longer be accurate as explicit racism in the United States has grown notoriously.

<sup>6</sup>Symbolic racism or racial resentment, as Mendelberg (2001) calls it, are forms of racism that find support and justification in the ideology of colorblindness.

made those two assertions in that very same order, but (3) would still be true even if Gingrich had made them in the opposite order.

- (3) Gingrich said that the best way to help others is to teach them to support themselves and that there are people that accept public welfare as a way of life.

Grice (1975/1989) distinguishes three kinds of implicatures, namely: *conventional implicatures*, and *particularized* and *generalized conversational implicatures*. The first are propositions suggested by the meaning of the words used in an utterance. In this sense, conventional implicatures are very similar to presuppositions, that is, they cannot be cancelled without oddity. For example, (5) cannot be rejected (it is not cancellable) without affecting the meaning conveyed uttering (4).

- (4) Gingrich is a politician but honest.

- (5) Politicians are not honest.

The meaning of the word ‘but’ in (4) triggers the implicature (5). To reject (5) after uttering (4) produces oddity, given the linguistic meaning of ‘but’. The two further types of implicatures, particularized and generalized conversational implicatures, are propositions suggested in exploiting conversational rules. The only difference between them is the degree of context-dependence of each one. Unlike conventional implicatures (and presuppositions), both particularized and generalized conversational implicatures are *cancellable*, *calculable* and *non-detachable*. For example, if (7) is a possible conversational implicature of (6), then a speaker can cancel (7) after uttering (6) without affecting the truth of (6), (7) can be calculated from the utterance of (6) in virtue of contextual elements of the conversation, and the substitution of the expression ‘mute’ for ‘unable to speak’ in (6) does not affect what is implicated by uttering it, that is, (7).

- (6) Gingrich was mute.

- (7) Gingrich did not participate in the annual debating competition.

## How (not) to Explain Covert Dogwhistles

In the above quotation, Gingrich’s speech apparently suggests something like African Americans are lazy. Even so he can explicitly reject that African Americans are lazy without obvious contradiction and without seemingly affecting the meaning of what he said. However, both presuppositions and conventional implicatures cannot be rejected without producing a certain oddity in the hearers. Besides, presuppositions and conventional implicatures are semantically linked to the meanings of the words such that it is practically impossible for their contents to be captured only by a selected part of the audience.

On the other hand, conversational implicatures are cancellable and can be rejected without affecting the main content asserted. So, it seems plausible to argue that Gingrich exploits conversational rules and contextual factors to influence a

subset of his audience sending the implicated message that African Americans are lazy.

Nevertheless, assuming that in using the term ‘food stamp’ Gingrich is suggesting that African Americans are lazy is troubling. A notable feature of covert dogwhistles, as we have seen, is their ability to affect a part of the audience without their awareness—that is, without the latter being aware that they are being moved for racist reasons. In the case of implicatures in general, speaker and hearer engage in a game of mutual recognition of intentions, without which the implicated content cannot be successfully captured by the hearer. In the case of dogwhistles, however, the success rests precisely because the hearer does not recognize the speaker’s intention yet is still mobilised for specific political purposes. The type of contribution the speaker makes through a covert dogwhistle instead seems to produce a certain effect on the audience rather than to communicate propositionally articulated content. The effect is not exactly about the recognition of *what is said*, but about *how what is said is said*, that is, how the words used influence the audience, and this is neither cancellable nor truth-apt. One way to confirm this is by analysing the behaviour of the implicated content and the “dogwhistled content” concerning retraction. Consider the following dialogues, the first is a case of retraction of a conversational implicature and the second is an alleged case of retraction of a covert dogwhistle:

- (8) A: Sam does not like to work.  
B: Well, she is African American.  
**(Implicature: African Americans are lazy).**  
B: Wait, wait! Sorry, I was wrong, not all African Americans are lazy. **(B’s retraction)**  
A: Exactly.
- (9) A: Over here you have a policy which, with Reagan and me as speaker, created millions of jobs—it’s called paychecks. Over there you have the most successful *food stamp president* in American history, Barack Obama.  
**(Dogwhistle: African Americans are lazy)**  
B: Well, food stamp recipients are a tiny minority.  
A: It is true, I was wrong, not all African American are lazy. **(A’s retraction)**  
B: Wait, I did not know that you were referring to African Americans.

As we can see, in case (8) the retraction of B does not produce surprise to A, in case (9), however, the retraction of A does produce surprise to B, because A’s retraction in (9) reveals to B something that had previously remained hidden. In fact, as Mendelberg (2001) and Saul (2018) show, when the covert dogwhistle becomes explicit the effect produced begins to change. Therefore, covert dogwhistles cannot be propositional because they lose their persuasive power when translated into a propositionally articulated statement.

## Concluding Remarks

As we have seen, presuppositions and implicatures are not quite suitable for explaining covert dogwhistles.<sup>7</sup> Presuppositions and conventional implicatures cannot be rejected without affecting the meaning of the utterances that trigger them (let's call this feature *deniability without oddity*) and are closely linked to the linguistic meaning of the words uttered (*linked to linguistic meaning*). However, part of the information conveyed by a covert dogwhistle can be rejected without affecting the meaning of the speech, and covert dogwhistles exploit associations and social meanings, not linguistic meaning (although they share with conventional implicatures that changing the words of the expression asserted removes the implicit—they are *not interchangeable*—, that is, the conventional implicature or part of the dogwhistle). Therefore, covert dogwhistles are neither presuppositions nor conventional implicatures. Conversational implicatures, on the other hand, can be rejected and exploit contextual factors, like dogwhistles. However, unlike covert dogwhistles, the success of a conversational implicature depends on the recognition of it by the audience (*successful yet unrecognized*), and retraction of the implicated proposition does not produce a kind of surprise in the audience (*retraction with oddity*). The following table shows the features pointed out.

|                                    | Covert<br>Dogwhistles | Presupposi-<br>tions | Conventional<br>implicatures | Conversational<br>implicatures |
|------------------------------------|-----------------------|----------------------|------------------------------|--------------------------------|
| Not interchan-<br>geable           | ✓                     |                      | ✓                            |                                |
| Deniability<br>without<br>oddity   | ✓                     |                      |                              | ✓                              |
| Successful yet<br>unrecognized     | ✓                     |                      |                              |                                |
| Retraction<br>with oddity          | ✓                     |                      |                              |                                |
| Linked to<br>linguistic<br>meaning |                       | ✓                    | ✓                            |                                |

It can be argued that, in cases of covert dogwhistles, retraction is in fact impossible. In case (9), for example, it can be argued that the speaker A is making explicit something that remained hidden and denying it, rather than retracting. This can be explained by appealing to Charlow's distinction between *locational* and *orientational* information (Charlow 2014). According to Charlow (and Lewis 1979), locational information is information that enables "an agent to self-locate in a space of relevant possibilities" (Charlow 2014: 640) by ruling out some of them. On the other hand, orientational information is information that ranks the relevant possibilities, allowing the agent to form intentions (Charlow 2014: 640). Covert dogwhistles do not convey information which is recognized by the audience, and in this sense, they do not enable the audience to rule out possibilities. Rather, it can be argued that they rank the relevant possibilities, putting some of them as more likely than others, and this is what produces the effect that mobilizes the audience for specific political purposes. Since covert

<sup>7</sup>The arguments advanced by Stanley (2015) and Saul (2018) also support this thesis.

dogwhistles do not convey information that rule out possibilities, they are not propositional. If covert dogwhistles are not propositional, one cannot retract a dogwhistle, strictly speaking, because what characterizes a covert dogwhistle is its implicit and unaware persuasive power.

## References

- Charlow, N. (2014). The problem with the Frege-Geach problem. *Philosophical Studies*, 167(3):635–665.
- Elliott, D. (2012). Food stamp president: Race code, or just politics. *National Public Radio.*, <https://www.npr.org/2012/01/17/145312069/newts-food-stamp-president-racial-or-just-politics?t=1537960323951>
- Gilens, M. (1996). “Race coding” and white opposition to welfare. *The American Political Science Review*, 90(3):593–604.
- Grice, H. P. (1975). “Logic and conversation,” in P. Cole & J. Morgan (eds.), *Syntax and Semantics, 3: Speech Acts Studies in the Way of Words*, pp. 41–58, New York: Academic Press. Reprinted in H. P. Grice (1989), *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, pp. 22–40.
- Henry, P. J. and Sears, D. O. (2002). The symbolic racism 2000 scale. *Political Psychology*, 23(2):253–283.
- Mendelberg, T. (2001). *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton University Press.
- Neubeck, K. and Cazenave, N. (2001). *Welfare Racism: Playing the Race Card Against America’s Poor*. Routledge.
- Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford University Press UK.
- Quadagno, J. (1994). *The Color of Welfare: How Racism Undermined the War on Poverty*. New York: Oxford University Press.
- Saul, J. (2018). “Dogwhistles, political manipulation and philosophy of language”. In Daniel Fogal, Daniel Will-Harris, M. M., (eds), *New Works on Speech Acts*, (pp. 360–384). Oxford University Press.
- Stalnaker, Robert (1973). Presuppositions. *Journal of Philosophical Logic*, 2(4):447–457.
- Stanley, J. (2015). *How Propaganda Works*. Princeton University Press.
- Tonhauser, J. (2011). Diagnosing (not-)at-issue content, *Proceedings of Semantics of Under-represented Languages in the Americas (SULA)*, 6. 239–254.
- Valentino, N. A., Hutchings, V. L., and White, I. K. (2002). Cues that matter: How political ads prime racial attitudes during campaigns. *The American Political Science Review*, 96(1):75–90.

- von Fintel, Kai (2004). “Would you Believe It? The King of France is Back! (Presuppositions and Truth-Value Intuitions)”. In Marga Reimer & Anne Bezuidenhout (eds.), *Descriptions and Beyond* (pp. 315–342). Clarendon Press.
- Witten, K. (2014). Dogwhistle politics: the new pitch of an old narrative.. Unpublished Manuscript.

# Crossed disagreements: A quantitative and qualitative study on the minutes of the sessions of the Spanish Parliament

David Bordonaba  
University of Valparaiso  
Neftalí Villanueva  
University of Granada

## Introduction

A great deal of attention has been paid in the last 15 years to the phenomenon of disagreement, both from the quarters of the philosophy of language and from epistemology, the focus has been, for the most part, on trying to set apart different kinds of disagreements: *faultless disagreement* (Kölbel 2004; Lasersohn 2009; Schaffer 2011; MacFarlane 2009, 2014), *metalinguistic disagreement* (Sundell 2011, 2016), *peer disagreement* (De Cruz & De Smedt 2013; Everett 2014; Davis 2015), *deep disagreement* (Lynch 2010; Kappel 2012) or *non straightforwardly factual disagreement* (Field 2009). While the challenge within the philosophy of language has been to provide an account to accommodate such a taxonomy, the epistemological challenge has rather been what to do once we realize that we are in front of disagreements of a particular kind, mostly peer disagreements. A lot less attention has been paid, though, to the fact that disagreements can be *framed differently*:

“Moreover, people may of course be mistaken or insincere in the way they conceive of the political disagreements that they are involved in. They may say, and even sincerely believe, that a given disagreement is fact-dependent, while it really is not. In some cases, there might be strategic advantages to framing a disagreement as fact-dependent, even when it is not. Or conversely, disagreements may be framed as value-dependent, when they are really fact dependent, and so on.” (Kappel 2017, 317).

The purpose of this paper is to analyze and define a particular linguistic phenomenon that we deem “crossed disagreement”. The aim of this paper is threefold. First, we describe what a crossed disagreement amounts to. Second, we provide a set of examples to support the idea that crossed disagreements often appear in political contexts. Third, we compare the examples to extract the most relevant and usual lexical features in an attempt to set apart possible regularities that could be used to identify crossed disagreements.

## Crossed Disagreements: A Brief Description

When the parties involved in a dispute conceive the disagreement in significantly different terms, they are involved in a *crossed disagreement*. Consider the following examples:

1. *Global warming*: Lisa and Thom discuss publicly what policies to adopt with respect to global warming. Lisa thinks that it is clear that human intervention has boosted global warming, and that policies should reflect that. Thom thinks that it is not clear whether human intervention has boosted global warming, and therefore we cannot make it the focus of our environmental policies. Lisa insists that the evidence backing her position is overwhelming, while Thom asserts that experts are divided, no matter in which percentage, and that we have no way to adjudicate.
2. *Huntington's advice*: Imagine a public debate between Lisa and Thom, concerning the use of gender-inclusive language in public institutions. Lisa, who favors the use of gender-inclusive language, asserts that the changing sensibilities of the population should be taken into account, and that ultimately a proposal should be voted. Thom, who is against the use of gender-inclusive language, claims that only linguists (or other experts) can know the real impact of such changes, can distinguish genuine motivations etc., and therefore the legislator should only listen to their expert opinions.

Whenever this happens, whenever fact-dependent issues are framed as value-dependent (global warming), or the other way around (Huntington's advice), the parties are engaged in, and probably presenting to the public as well, a crossed disagreement.

Three provisos before we move forward. First, in a crossed disagreement, speakers are not talking *at crossed purposes*. A crossed disagreement is a genuine disagreement in which both parties conceive of the disagreement in different ways, but they are still discussing the same issue. Lisa and Thom above discuss the extent of human influence in global warming and the use of gender-inclusive language in public institutions, they are not talking past each other. Second, many public instances of disagreements can contain not just one, but several crossed disagreements. When discussing a particular topic in a public forum, it is common to try to re-frame the issue in more than one way. Crossed disagreements will come thus in different degrees. Finally, crossed disagreements, with the effect that we explore here, are only to be found in public discussions. Related phenomena that rather belong to the private sphere of interpersonal communication are not our target here.

In order to make the identification of crossed disagreements both tractable and theoretically parsimonious, we present a simplified—dynamic—taxonomy of classes of disagreement:

1. Type A. Disagreements in which there is a presumption of commonality with respect to the standards of both parts, e.g., the distance between Granada and London.
2. Type B. Disagreements that become about the standards, once it becomes



obvious that both parties have different standards, e.g., moral philosophers discussing trolley-cases.

3. Type C. Disagreements that neither disappear nor become about the standards, once it becomes obvious that both parties have different standards, e.g., musical appreciation.

Crossed disagreements are thus instances of public discourse in which different parties behave as if they conceived a given discussion in ways that sufficiently accord, respectively, with at least two of the above categories of disagreement.

Public political discourse is an important source of examples of this kind—though not the only one. Our examples are based on data from three different corpora of the Minutes of the Sessions of Spanish Parliament, from 2004 to 2016. Exhaustive analysis based on systematic searches of particular expressions—explicit marks of disagreement, ‘discussion’, ‘debate’—have made possible to depict a rich overview of the complexity of crossed disagreements. A stark contrast is created when one of the parties frames the disagreement as a type A-disagreement (“factual”, “fact-dependent”), while the other frames it as a type B-disagreement (“normative”, “value dependent”), as in the cases included above, but also when one party presents the dispute as a type C-disagreement (“evaluative”, “non-straightforwardly factual”), while the opposing part doesn’t.

## Methodology: A Corpus-Informed Approach

Corpus linguistics (McEnery & Wilson 2001; Parodi 2008; McEnery & Hardie 2012) is a methodology used in studies pertaining to many and diverse fields as theory of literature, lexicography, forensic linguistics or, as in our case, experimental philosophy of language. The use of linguistic corpora for the study of questions of philosophy of language has been explicitly defended on more than one occasion recently (see Bluhm 2013, Hansen and Chemla 2015, Hansen 2015). Our research is mainly a corpus-informed discourse analysis since we use corpora only as a source of examples to prove a certain thesis, that is, that crossed disagreements are a widespread and real phenomenon that take place in many political debates. Our corpora are monolingual (only Spanish) and sample corpora (they are representative of a specific kind of language—language used in parliamentary debates—in specific lapses of time—three different parliamentary terms). We formed three different corpora belonging to Legislatures VIII (2004-2008), IX (2008-2011) and X (2011-2016).

It has been argued (Mollin 2007; McEnery & Hardie 2012) that corpora made up from transcripts of parliamentary debates ‘have not been designed as reliable materials for linguistic exploration of spoken language’ (McEnery & Hardie 2012: 4). They criticize certain studies based on corpora built upon the Hansard reports, that is, transcripts of parliamentary debates in Britain and other Commonwealth countries. As Mollin says, the studies based on Hansard reports are deficient because the transcripts ‘omit performance characteristics of spoken language, such as incomplete utterances or hesitations, as well as any type of extrafactual, contextual talk (e.g., about turn-taking)’ (Mollin 2007: 1). In contrast, our study is based on the Minutes of the Sessions of the Spanish

Parliament, that include a lot of performance characteristics such as incomplete utterances, expressions and taggers marking turn-takings, statements and shouts of other people who are not in their speaking time, or other extrafactual information. Consider the following examples as a way of illustration:

El señor PRESIDENTE: Muchas gracias, señora Esteve. Para la fijación de posiciones, por el Grupo Parlamentario Vasco, PNV, tiene la palabra don Emilio Olabarriá.

El señor OLABARRÍA MUÑOZ: Gracias, señor presidente. Señor presidente, mi grupo no está aquí ubicado...<sup>1</sup>

Usted me dirá: han tardado ocho años y no lo han corregido. Ciertamente, señor Rajoy, pero tiene un origen, viene de los costes —¿se acuerda?— a la competencia. Todo este sistema tiene un origen. **(El señor Nadal Belda hace signos negativos.)** Sí, sí, son ustedes los que inventaron un sistema según el cual se paga menos por la electricidad, las eléctricas cobran y el Estado tiene un déficit. **(El señor Aguirre Rodríguez pronuncia palabras que no se perciben.)**<sup>2</sup> [énfasis en el original]

Pero, por encima de todo, también velaremos para que el producto del genio y del talento de nuestros cineastas, de nuestros músicos, de nuestros artistas, de nuestros creadores... **(Rumores.— Aplausos.)**<sup>3</sup> [énfasis en el original]

## Crossed Disagreements: An Example

We are going to present only one example of crossed disagreement because of the special nature of parliamentary debates. These debates can be taken during long periods and often concern more than two people. The example we set out below is an example of a crossed disagreement where one party conceived the dispute as a type B-disagreement, and the other party conceives it as a type A-disagreement. It's characteristic of Type A-disagreements that the parties act as if the dispute made no sense had they realized that their standards are not shared (presumption of communality). Type B-disagreements critical feature is that the parties seem willing to discuss their standards once they've discovered that they were endorsing different standards. The example is made out of excerpts of a debate Of law 39/2003, of november 17, of railway sector (Expedient number 130/000001):<sup>4</sup>

El señor Ayala Sánchez: Señora ministra, las dos primeras razones, que por cierto son las únicas que se encuentran en la exposición de

---

<sup>1</sup>Página 28, Número 131 del Diario de Sesiones del Congreso de los Diputados, Legislatura IX.

<sup>2</sup>Página 29, Número 2 del Diario de Sesiones del Congreso de los Diputados, Legislatura X.

<sup>3</sup>Página 24, Número 2 del Diario de Sesiones del Congreso de los Diputados, Legislatura VIII.

<sup>4</sup>In capital letters appear the political expressions related with territorial issues, and in small capital letters appear the explicit marks of disagreement.

motivos, literalmente NO SON CIERTAS. [...] Tras esos dos argumentos no válidos, el tercer argumento de la deuda de Renfe no puede dejar de ser más esperpéntico. [...] Hemos visto que estos tres argumentos no son en absoluto atendibles y nos falta llegar al cuarto, que nos parece el más bonito, es decir, que por la interposición de unos recursos de inconstitucionalidad el Gobierno suspenda la vigencia de una ley. Eso es de una gravedad monstruosa. [...] Ahora bien, si ustedes a partir de ahora van a utilizar y van a permitir que se suspenda la entrada en vigor de una ley sólo por la interposición del recurso de inconstitucionalidad, díganoslo, porque entonces, si nosotros interpusiéramos un recurso de inconstitucionalidad ante la suspensión del Plan Hidrológico Nacional, ustedes acordarían la suspensión de la suspensión, y lo mismo se daría con la Ley de Calidad de la Educación. Por tanto, si van a mantener este criterio, a partir de mañana muchas COMUNIDADES AUTÓNOMAS empezarán a interponer recurso de inconstitucionalidad ante la nueva fórmula...

El señor De La Encina Ortega: Usted sabe que la ley, señor portavoz del Grupo Popular, no ha contado con las competencias que las comunidades autónomas tienen constitucionalmente establecidas sobre el transporte ferroviario, y eso se recoge en el informe del Consejo Económico y Social y en su dictamen sobre la ley, señor Ayala. Usted no puede hurtar aquí algo tan serio como lo que está recogido en el informe del Consejo Económico y Social. Por tanto, cuando la ley regula un sistema integrado de transporte lo hace sin tener en cuenta a las COMUNIDADES AUTÓNOMAS, sin darles participación en ninguno de los dos entes que se crean para gestionar el transporte y la infraestructura ferroviaria,...

La señora Ministra de Fomento (Álvarez Arza): El que HA FALTADO A LA VERDAD es usted. En primer lugar, porque me ha citado diciendo que he argumentado que sólo había cuatro razones. Señor Ayala, NO ES CIERTO y me remito al «Diario de Sesiones». He dicho: Hay dos razones fundamentales y otras cuestiones. Señor Ayala, escuche usted cuando intervengamos porque me parece muy poco serio que usted diga a un miembro del Gobierno que FALTA A LA VERDAD, cuando el único que HA FALTADO A LA VERDAD es usted en múltiples ocasiones en esta intervención.

El señor Ayala Sánchez: Señora Álvarez, confirmo, mantengo y sostengo todas las afirmaciones que he hecho en la tribuna en la mañana de hoy, todas y cada una de las que he hecho. (Aplausos.) YO NO HE MENTIDO aquí. USTED SÍ HA MENTIDO aquí. [...] Señora Álvarez, esto no empieza con un rollito de primavera –y sabe usted muy bien lo que le estoy diciendo–, esto es algo bastante más serio. Estamos aquí hablando de cuestiones trascendentales para lo que es la VERTEBRACIÓN TERRITORIAL española, estamos aquí hablando de cuestiones fundamentales para nuestra integración en las redes transeuropeas de transporte. De eso estamos hablando aquí. [...]

## Possible Marks of Crossed Disagreements

In order to obtain possible candidates that could serve as explicit marks of crossed disagreement, we tried to isolate specific characteristics of the different types of disagreement presented above. To do that, we made close readings not only of the above example, but also of other examples we encounter in the corpora where the different parties conceive the disagreement in different terms. The following is an open unfinished list of some characteristics (their translations into English) that could be used to identify different types of disagreements:

1. Expressions that could indicate that the dispute is conceived as a type A-disagreement: references to data, expressions appealing to the law such as “the law  $x$  says:...”; expressions such as “I refer to what I said” or “I have just said”.
2. Expressions that could indicate that the dispute is conceived as type B-disagreement: subjunctive conditionals, expressions such as “transcendent issues” or “fundamental issues”, “criterion”, “standard” (note that our type B covers deep but also certain metalinguistic disagreements).
3. Expressions that could indicate that the dispute is conceived as a type C-disagreement: “should”, “must”, “un/fair” (plus the absence of some of the characteristics of type B-disagreements).

During the close reading, we realize that sometimes the Members of the Spanish Parliament reach a point in the discussion where they try to re-frame the debate. That is, they use specific expressions that seem to state that the discussions have been misconceived. The most used expressions were: “That is not what we are here to debate”, “the debate is not...” or “the discussion must not be...”. We think that this type of expressions could be used to directly identify crossed disagreements.

## Conclusions

The purpose of this paper has been to identify a linguistic phenomenon that concerns the notion of disagreement, and that we take to be underrepresented in the current literature on the issue. Crossed disagreements are, we claim, scenarios of public debate where both parties appear to conceive the dispute that they are engaged in significantly different terms. We propose a new and simplified taxonomy of kinds of disagreement to make our phenomenon both empirically tractable and theoretically parsimonious, and provide an example taken from our work with the Minutes of the Sessions of the Spanish Parliament. This paper is part of a bigger project in which we tackle, both theoretically and empirically, the connections between crossed disagreements and polarization.

## References

- Bluhm R. (2013). Don't Ask, Look! Linguistic Corpora as a Tool for Conceptual Analysis. En Migue Hoeltje, Thomas Spitzley & Wolfgang Spohn (eds.), Was dürfen wir glauben? Was sollen wir tun? Sektionsbeiträge des achten internationalen Kongresses der Gesellschaft für Analytische Philosophie e.V. DuEPublico. pp. 7–15.
- Davis, J. K. (2015). "Faultless Disagreement, Cognitive Command, and Epistemic Peers". *Synthese* 192: 1–24.
- De Cruz, H. & De Smedt, J. (2013). "The Value of Epistemic Disagreement in Scientific Practice. The Case of Homo Floresiensis". *Studies in History and Philosophy of Science Part A* 44(2): 169–177.
- Everett, Th. J. (2015). "Peer Disagreement and Two Principles of Rational Belief". *Australasian Journal of Philosophy* 93(2): 273–286.
- Field, H. (2009). "Epistemology without metaphysics". *Philosophical Studies* 143: 249–290.
- Hansen, N. & Chemla, E. (2015). "Linguistic Experiments and Ordinary Language Philosophy". *Ratio* 28(4): 422–445.
- Hansen, N. (2015). "Experimental philosophy of language". In: *Oxford Handbooks Online*. Oxford University Press.
- Kappel, K. (2012). "The Problem of Deep Disagreement". *Discipline Filosofiche* 22(2), 7–25.
- Kappel, K. (2017). "Fact-Dependent Policy Disagreements and Political Legitimacy". *Ethical Theory and Moral Practice* 20(2): 313–331.
- Kölbel, M. (2004). "Faultless Disagreement". *Proceedings of the Aristotelian Society* 104(1): 53–73.
- Lasersohn, P. (2009). "Relative Truth, Speaker Commitment, and Control of Implicit Arguments". *Synthese* 166(2): 359–374.
- Lynch, M. P. (2010). "Epistemic Circularity and Epistemic Disagreement". In A. Haddock, A. Millar & D. Pritchard (eds.), *Social Epistemology*, Oxford University Press, Oxford, 262–277.
- MacFarlane, J. (2014). *Assessment-sensitivity: Relative Truth and its Applications*. Oxford University Press, New York.
- McEnery, T. & Wilson, A. (2001). *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- McEnery, T. & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Mollin, S. (2007). "The Hansard Hazard: Gauging the Accuracy of British Parliamentary Transcripts". *Corpora* 2(2): 187–210.

- Parodi, G. (2008). “Lingüística de Corpus: Una Introducción al Ámbito”. *Revista de Lingüística Teórica y Aplicada* 46(1): 93–119.
- Schaffer, J. (2011). “Perspective in Taste Predicates and Epistemic Modals”. In: A. Egan and B. Weatherson (eds.), *Epistemic modality*, Oxford University Press, Oxford, 179–226.
- Sundell, T. (2011). “Disagreements about Taste”. *Philosophical Studies* 155(2): 267–288.
- Sundell, T. (2016). “The Tasty, the Bold, and the Beautiful”. *Inquiry: An Interdisciplinary Journal of Philosophy* 59(6): 793–818.

# The difficulty of finding a commonsense answer to the special composition question

Marta Campdelacreu

Peter van Inwagen, in (van Inwagen 1990), famously formulates the ‘Special Composition Question’ (SCQ): When do some things compose something? Chad Carmichael, in (Carmichael 2015), formulates an innovative answer to the SCQ which, he argues—unlike the rest of the answers proposed—accords with common sense about composition.

Carmichael’s proposal:

Necessarily, for all  $xs$ , there is a  $y$  such that the  $xs$  compose  $y$  iff either

- (i) the  $xs$  are lump-like and the  $xs$  are bonded, or
- (ii) the activities of the  $xs$  constitute an event that imposes sufficient unity on the  $xs$ .

Where:

$z$  is lump-like iff  $z$  is a material object and there is no event  $e$  such that the fact that  $z$  exists is grounded in the fact that  $e$  occurs.

Where:

If the fact that  $z$  exists is grounded in the fact that  $e$  occurs, and there are some  $ts$  such that the  $ts$  compose  $z$  and all and only the  $ts$  participate in  $e$ , then  $e$  *imposes sufficient unity* on the  $ts$ .

Where:

$e$  imposes *sufficient unity* on the  $ts$  iff

- (i) the activities of the  $ts$  constitute  $e$ ;
- (ii)  $e$  is *self-sustaining*;
- (iii)  $e$  is *homeodynamic*.

Moreover, the  $zs$  are bonded iff the  $zs$  are related by the ancestral of the *joining relation*.

Now, van Inwagen (1990) states several powerful objections to answers of this kind (i.e., series-style answers) to which Carmichael (2015) provides interesting replies. In this paper I will focus on Carmichael’s answers to two of van Inwagen’s objections. My main aim is to argue that Carmichael’s answers are not satisfactory, and thus reflect the difficulty of finding a commonsense answer to SCQ.

## The transitivity of parthood objection

As Carmichael says, given his proposal, there are two ways one might try to construct a counterexample to the transitivity of parthood. However, he argues, neither is possible. In this paper I focus on the one in which some objects, the *xs*, compose something in virtue of the fact that their activities constitute a unity-imposing event and, so, they are parts of the composed object; but some part of one of the *xs* is not among the objects whose activities constitute the event, and so it is not part of the composed object.

Against the possibility of such a case, Carmichael argues that: if the activities of the *xs* constitute *e*, and *y* is part of one of the *xs*, then there are some *zs* such that the activities of the *zs* constitute event *e*, and *y* is among the *zs*.

First, a definition:

- (C) The activities of the *xs* constitute event *e* iff the fact that *e* occurs is grounded in facts about the activities of the *xs*.

Now, let us suppose that some event *e* is constituted by (the activities of) composite objects *x*<sub>1</sub> ...*x*<sub>*n*</sub>. By (C), the occurrence of *e* is grounded in facts about (the activities of) *x*<sub>1</sub> ...*x*<sub>*n*</sub>.

1<sup>st</sup> premise: the facts about a composite object are grounded in facts about its parts. Then: the facts about *x*<sub>1</sub> ...*x*<sub>*n*</sub> that ground the occurrence of *e* are grounded in facts about their parts.

2<sup>nd</sup> premise: grounding is transitive.

Then: the occurrence of *e* is grounded in facts about the parts of *x*<sub>1</sub> ...*x*<sub>*n*</sub>.

By (C), facts about the parts of *x*<sub>1</sub> ...*x*<sub>*n*</sub> constitute event *e*.

Now, after presenting Carmichael's argument in detail, I will argue that its 1<sup>st</sup> premise contradicts some other relevant theses in his account. Let me explain. The mentioned premise claims that the facts about a composite object are grounded in facts about its parts. For any composite object, one of these facts is its coming into existence. Then, the existence of a composite object is grounded in facts about its parts. Now, let us apply this to a rock, one of Carmichael's examples. This means that the existence of a rock is grounded in facts about its parts. And, in fact, this seems to be in accordance with common sense. The fact that some adequate parts are arranged in the required way seems to ground the existence of the resulting rock. However, it is part of Carmichael's framework that the existence of objects such as rocks is not grounded in the activities of their parts. A rock (for example) is a lump-like object. So, there is no event such that the fact that the rock exists is grounded in the fact that this event occurs. If this were so, given the definitions above, this event would impose sufficient unity on the composers of the rock and this would imply that the relevant event is homeodynamic, which is not possible in such a case; for this would imply that it is possible for the rock to change all its parts.



## The *Ship of Theseus* objection

Carmichael accepts van Inwagen's requirement that an answer to the SCQ has at least to shed some light on ancient puzzles of persistence, like the Ship of Theseus puzzle. He argues that his proposal can solve it using the distinction between *event-based* and *lump-like* objects together with the fact that these notions very plausibly have built-in persistence conditions. He argues that the Original Ship of Theseus (the one we start with) cannot be an event-based object. So, he concludes, we should identify it with the ship which is slowly disassembled and, afterwards, reassembled.

I will argue that the fact that the Original Ship of Theseus is a *lump-like* object leaves completely undetermined the matter of which *later* ship it is identical to. Carmichael's understanding of *lump-like* objects is compatible with the two possible answers to the puzzle (the second one being that it is identical to the ship to which all planks are slowly replaced). For all Carmichael says in the paper (developing the definition above), the ancient puzzle of the Ship of Theseus remains unresolved.

## References

- Carmichael, Chad (2015). 'Toward a Commonsense Answer to the Special Composition Question'. *Australasian Journal of Philosophy*, 93:475–490.
- van Inwagen, Peter (1990). *Material Beings* (Ithaca: Cornell).

# The is-ought gap and the grounds of normative facts

**Samuele Chilovi**

University of Barcelona

**Daniel Wodak**

Virginia Tech

Hume's Law—roughly, that one cannot derive an “ought” from an “is”—has long been declared to pose a serious problem for moral naturalism and legal positivism (hereafter, “naturalism” and “positivism”). James Rachels and Charles Pigden noted the widespread acceptance of the view that if Hume's Law is true then naturalism is false (and, perhaps, vice versa).<sup>1</sup> The status of appeals to Hume's Law in philosophy of law is even more interesting. On the one hand, it has long been considered to be one of the central challenges to positivism.<sup>2</sup> Here's how Scott Shapiro recently explained the nature of that challenge:

According to the legal positivist, the content of the law is ultimately determined by social facts alone. To know the law, therefore, one must (at least in principle) be able to derive this information exclusively from knowledge of social facts. But knowledge of the law is normative whereas knowledge of social facts is descriptive. How can normative knowledge be derived exclusively from descriptive knowledge? That would be to derive facts about what one legally ought to do from judgments about what is the case. Legal positivism, therefore, appears to violate the famous principle introduced by David Hume [...], which states that one can never derive an ought from an is (Shapiro 2011, p. 43).

And on the other hand, Hume's Law has also been invoked against *rivals* to positivism. Brian Bix wrote that it “undermines a major strand of natural law theory”, and “by many accounts” is responsible for pushing “natural law theory to the sidelines” (Bix 2002, pp. 74–75).

As the above suggests, the *nature* of the problem that Hume's law putatively poses for views like positivism or naturalism is often opaque. (Is it, as Shapiro suggests, an epistemological problem?) And as a result, so is the *scope* of the

---

<sup>1</sup>Here's Rachels: ‘There is... another argument that many philosophers believe is devastating to ethical naturalism. David Hume is credited with first observing that we cannot derive “ought” from “is.” [...] Factual judgments and evaluative judgments are fundamentally different, and no amount of purely factual information can logically entail any evaluation. It is commonly assumed that, if this is true, the naturalistic project is doomed’ (Rachels 2000). And here's Pigden: ‘it is often assumed that if moral judgements can be derived from non-moral propositions, naturalism is true. If not, naturalism is false’ (Pigden, 1989).

<sup>2</sup>This challenge became prominent with the work of Hans Kelsen (1934). See Marmor's summary of Hans Kelsen's views on this issue: Marmor 2010 p. 17.

putative problem. (Can it be a problem for positivism and its rivals? What about naturalism and its rival, non-naturalism?)

Our goals in this paper are threefold. First, in §I, we argue that it is a mistake to declare that if Hume's Law is true, positivism and naturalism are false. Such declarations ignore the gap between a *logical thesis* about entailment relations between normative and descriptive statements (Hume's Law) and *metaphysical theses* about grounding relations between normative and descriptive facts (positivism and naturalism).<sup>3</sup> Second, in §II, we offer the most direct route for bridging this gap between Hume's Law and positivism or naturalism. This route commits to two principles which link grounding to necessitation, and necessitation to entailment. Both bridge principles are defensible, albeit controversial. But if one doesn't take this direct route, they need to supply an alternative way to bridge the gap. And we argue that there is much to recommend the route we offer over the indirect, epistemological one offered by Shapiro.

Finally, in §III we clarify the scope of the problem posed by Hume's Law. Since this point has been neglected, we show how Hume's Law and the two bridge principles would undermine moral naturalism as well as prominent forms of moral non-naturalism, such as one recently defended by Stephanie Leary (2017). Only "immodest" forms of non-naturalism can be compatible with Hume's Law and the two bridge principles.

## References

- Bix, B. (2002) 'Natural Law: The Modern Tradition'. In Jules Coleman & Scott J. Shapiro (eds.), *The Oxford Handbook of Jurisprudence and Philosophy of Law*. Oxford University Press.
- Finnis, J. (1980) *Natural Law and Natural Rights*. Oxford University Press UK.
- Kelsen, H. (1934) *Reine Rechtslehre Einleitung in Die Rechtswissenschaftliche Problematik*. F. Deuticke.
- Leary, S. (2017) 'Non-naturalism and normative necessities', in Shafer Landau, R. (ed.) *Oxford Studies in Metaethics*, 12: 151–174.
- Marmor, A. (2010) *Philosophy of Law*. Princeton University Press.
- Pigden, C. R. (1989) 'Logic and the Autonomy of Ethics', *Australasian Journal of Philosophy*, 67(2):127–151.
- Rachels, J. (2000) 'Naturalism', in *The Blackwell Guide to Ethical Theory*, Hugh LaFollette (ed.), Blackwell.
- Shapiro, S. (2011) *Legality*. Harvard University Press.

---

<sup>3</sup>There is disagreement about how to formulate Hume's Law, positivism, and naturalism. We consider this in §I. There is also disagreement about how to interpret Hume's key passage in the *Treatise*, which we won't discuss. See Finnis 1980, pp. 37-48.

# Human deductive reasoning: Conditional perfection and mental models theory

Juan Esteban Cortés-Aravena  
Universidad de Talca - Chile

## Abstract

An interesting phenomenon related to the use of deductive reasoning and the way how humans process information has been called Conditional Perfection or ‘invited inferences’. In simple words, it happens when a conditional utterance is interpreted as a biconditional utterance, i.e., in a pragmatic situation the material implication (or conditional interpretation) of *If...then* statement is understood as *If and only if... then* material equivalence (or biconditional interpretation). This is when *If p then q* is understood as *If and only if p then q* (or *If p then q and if q then p*). For example, if you *mow the lawn, I’ll give you five dollars*, according to Geis and Zwicky (1971) will invite or suggest the inference of *if you don’t mow the lawn, I won’t give you five dollars*. Conditional Perfection has been seriously studied for more than sixty years, and there are still some discrepancies based on its theoretical foundations and its practical performance. I will argue that the Mental Models Theory (MMT) by Johnson-Laird and Byrne (1991, 2009) is an accurate model to explain this phenomena. Also, the results obtained by Bob van Tiel and Schaeken, W. (2016) related to efficient high-cognitive time-response tasks with conversational implicatures and Couto, Quelhas and Byrne (2017) related to inferences with advice may help to support the idea that the human mind is able to evaluate *If...then* statements according to the meaning and context of the sentences. As a result, it is possible for the mental models theory (MMT) to explain Conditional Perfection (CP) by means of iconic mental representations using few and true possibilities (including alternatives, enablers and disablers) rather than the traditional explanation that classifies it as a scalar implicature depending on the quantity maxim by Horn, L. (2000), Moldovan (2009), van der Auwera (1997) and Geis and Zwicky (1971).

**Key words:** Conditional Perfection, pragmatics, deduction, conversational implicatures, mental models theory.

## Introduction

In this research, conditional inferences in the form *if...then* are considered to be a key factor to explain some reasoning activities and certain characteristics and classifications. Even though all human beings are able to formulate speculations and hypothetical situations using conditional statements, where at least premises and conclusions are connected under special conditions (plans, advice, threatening, promises, facts, laws, counterfactual statements, etc.), it is difficult for users (laymen or naïve users) to be aware about all the possible combinations underlying this deductive process. Conditional Perfection has been seriously studied for more than sixty years, and there are still some discrepancies based on its theoretical foundations and its practical performance related to the linguistic forms and manifestations of fallacies (Geis and Zwicky, 1971; Horn, 2000; López-Astorga, 2016; Noordman, 1979, van der Auwera, 1997).

## Initial Studies

Initially, the study of material implication by logicians indicates that the truth-value of a conditional inference is false only if the antecedent is true and the consequent is false, and all the other conditional conjunctions are possibly true, i.e. *p and q*, *not-p and not-q*, and *not-p and q*. So, it is difficult to evaluate the truth-value of the conditional when the antecedent is false because anything follows. Moreover, in terms of syllogistic reasoning, when *if p then q* is stated, it is possible to obtain valid inferences when the premise *p* is followed by *q* (Modus Ponens) and *not-q* is followed by *not-p* (Modus Tollens), but from *not-p* (Denial of the antecedent) or *q* (Affirmation of the consequent) no conclusion can be drawn. On the other hand, the material equivalence *If and only if p then q* allows only two truth-value combinations: *p and q* and *not-p and not-q* are true, *not-p and q* is false, *p and not-q* is also false. In this case, *q* follows *p*, *p* follows *q*, *not-q* follows *not-p* and *not-p* follows *not-q*; a biconditional relation.

Some of the first psychological studies about conditionals (Matalon, 1962; Peel, 1967) suggested that only children tend to interpret the conditional sentences as biconditional, this means that *not-p and q* is not acceptable in an *if...then* sentence. As Noordman (1979) noticed in those and many other investigations (e.g., Inhelder and Piaget, 1958; Paris, 1975; Roberge and Paulus, 1971) the responses were considered correct just when they follow the truth function of the material implication. The tasks in these investigations were to draw conclusions from premises, indicate and enumerate easy and difficult possible combinations. The materials were pairs of pictures or sentences that expressed logical relations and the focus was the evaluation of *if p then q* when *p* was false and *q* was true. Other experiments found the same results in adults; subject interpreted a conditional as a biconditional, with *not-p and q* as false (Legrenzi, 1970; Taplin, 1971; Taplin and Staudenmayer, 1973). Nevertheless, Wason (1966) and Johnson-Laird and Tagart (1969) casted doubt on the material interpretation and argue that “adults do not treat the conditional in a truth-functional manner: they consider it to be *irrelevant* when its antecedent is false” (Johnson-Laird and Tagart, 1969, p. 367). As a result, in sentences like *if p then q*, the conjunction

conditional *p* and *q* is true, *p* and *not-q* is false and any situation with *not-p* as antecedent will be irrelevant. Currently, this is the basic idea of the Mental Model Theory when interpreting conditionals and biconditionals (Johnson-Laird and Byrne, 1991, 2009; Khemlani, Byrne, Johnson-Laird, 2018).

In philosophy, the studies used to understand the Conditional Perfection (CP) as suspected of being a problem when people tend to obtain a conclusion denying the antecedent (DA) to find some conclusions, mainly because the use of a DA reduces the possibility to draw a direct conclusion due to the fact that there are two possible consequents, i.e., *q* and *not q*. CP is defined by van der Auwera (1997) as a phenomenon that can be characterized as follows: *if* is often understood to mean *only if*. Also, van der Auwera (1997) classifies the problem of CP as a problem in the pragmatics of conditionals. The example given by this researcher is that one used by Geis and Zwicky (1971), a typical promise of first conditional statement as modus ponens: 1) *If you mow the lawn, I'll give you five dollars*. According to this view, in this case, the speaker is not just uttering the literal meaning, but it is uttering some implicit meaning. The process described here is related to a scalar implicature:

- 2) *If you don't mow the lawn, I won't give you five dollars (if no p, then no q).*
- 3) *Only if you mow the lawn will I give you five dollars. (only if p, then q).*
- 4) *If and only if you mow the lawn will I give five dollars. (if and only if p, then q).*

So, in the first step, a sentence of the form *if p then q* invites or suggests the inference of the form *not-p then no-q*. This principle “asserts a connection between linguistic form and a tendency of the human mind- a tendency to “*perfect conditionals to biconditionals*”, in words suggested to us by Lauri Karttunen” (Geis and Zwicky, 1971, p. 562). Van der Auwera (1997) states that case (1) will ‘invite’ the inference of 2), 3), and 4). They have 2 characteristics, they seem fairly general, and they don’t seem to be language-dependant.

In pragmatics, subjects tend to logically infer that a simple conditional, under specific context and in certain inference tasks, can be interpreted as a biconditional. That means that the individuals tend to avoid *p* and *not-q* and *not-p* and *q*. It has been stated that Conditional Perfection (CP) is a conversational implicature (taken from Grice) with a distortion on one of the maxims of cooperation, more specific, the maxim of quantity: *make your contribution as informative as is required, and do not make your contribution more informative than is required*. CP has historically been considered a fallacy, i.e. an argument that involves an invalid, rather than a valid form of reasoning (this is the strictly logical usage). However, as it has been shown by linguistic philosophers (logic and pragmatics), it has a systematic regularity in human reasoning.

Similarly, Moldovan (2009) will appeal to a Gricean pragmatic account of conditionals to support his conclusion. For instance, using the Theory of Analysis, he deals with “the problem of how to identify and reconstruct an argument that is put forward in a text or oral discourse” (p. 2). For this account, in order to analyze the conditional arguments it is important to distinguish two kinds of implications. These propositions with extra meaning are called implicatures,

“implicatures always differ from *what is said*, but may be entailed by *what is said*, or merely suggested by it” (Moldovan, 2009, p. 2). The term *implicature* has been introduced to Philosophy of Language to clarify the difference between these two kinds of implications. According to Grice (1975, 1989) the conversational implicatures are those inferences that are not directly inferred in a conversation, mainly because the speaker could be violating some of the maxims to go beyond the literal meaning. And for Geis and Zwicky (1971) “it seems, then, that what we have called “invited inferences” constitutes a special class of “implicatures”, in the terminology of the philosopher H. Paul Grice, although they are clearly distinct from the “conversational implicatures” (p. 565).

A recent and important cognitive research by van Tiel et.al, (2016) analyses CP as part of conversational implicatures, among others, *scalar inference*, *free choice inference* and *it-clefts*. The main conclusions of this research in relation to CP are that its behavior is different from *scalar inferences*, which defeats the idea of Moldovan (2009). First of all, van Tiel (2016) makes the distinctions between two ways of processing the utterance information; on the one hand, the literal-first hypothesis (LFH) that states that the speaker analyses the literal meaning before the intended meaning as a serial process, and on the other hand, the parallel hypothesis (PH) that refers to the possibility that both meanings become activated and enter into the interpretative process at the same time. Opposite to the scalar implicature, the results for CP is that the time process of the intended meaning is as fast as the literal meaning, so CP might not be considered as having an extra cognitive effort because there is no extra delay in response times, “the interpretation enriched with a conversational implicature was processed at least as quickly as the literal interpretation, thus casting doubt on the literal-first hypothesis” (p. 3). Other researchers have found these same characteristics in sarcasm, indirect questions, metaphors, and hyperbole (Gibbs, 1986; Clack, 1979; Glucksberg, 1982).

## Mental Model Theory

Back to the Mental Models Theory (MMT) by Johnson-Laird (1983) and Johnson-Laird and Byrne (1991), which antecedents can be traced back to C.S. Peirce and Kenneth Craik, it is stated that we reason by means of models and that these models represent the different valid possibilities or alternatives consistent with both the premises and the conclusion of the inferences. All these possibilities correlate with reality, so the context, where these utterances are performed have to be taken into account as well as its semantic content. This theory helps us to understand that individuals use their general and previous knowledge to state a set of mental models by means of iconic mental representations.

The mental models represent only what is true, however, the fully explicit models (FEM) represents what is irrelevant for the model too. The principle of truth states that the construction of the FEM based on *not-p*, calls for deliberation and access to working memory. That is to say that, the model *p* and *q* are *what is true* and they are fairly automatic when analyzed by the system 1, and the other alternatives will require further attention with system 2 (in Khemlani, Byrne and Johnson-Laird, 2018). We conclude that the study of CP is better

explained by MMT because what subjects do is to create the mental models of the conditional which is the same as the biconditional, that is to say *p and q*, but *p and not-q* is false, and when the antecedent is false there should be a further evaluation assessing alternatives which initially are null.

In cognitive science, a good example of a research that has studied pragmatic conditional is done by Couto, Quellas y Byrne (2017). This research can give us some interesting conclusions to understand in a better way the use of conditionals and CP. There are some key discoveries about advice which are important for conditional perfection advice and inducement. Firstly, “Readers discriminate rapidly between tips and promises and between warnings and threats”. It seems to be that people discriminate “by relying on pragmatic cues such as whether the outcome is positive for the listener—to discriminate promises and tips vs. threats and warnings, and whether the speaker controls the outcome—to discriminate further, between promises vs. tips, and threats vs. warnings”. (p. 363). So, when we analyze advice conditionals, there are two questions to answer: Is the outcome positive for the listener? And, does the speaker have control over the outcome?

The use of conditionals in reasoning helps us to resolve problems, so, what is the objective of advice and inducements? This is to encourage others to do something. A possible hypothesis is that in promises like *if you mow the lawn, I'll give you five dollars*, the tendency would be the conditional perfection because it is not possible that she does the job but there is no payment as a reward, but, with tips like *if you study your grades will improve*, the tendency would be to maintain the enablers, blocking the MP and MT. Therefore, it would not be the case that the listener makes a derivative evaluation of the possible alternative combinations or that he relies on the monetary supposition. What really happens is that the listener evaluates at the same time the meaning of the syntax and the lexicon added to the speech act (similar to the parallel processing hypothesis). When evaluating them as a promise or as a recommendation, a tendency should be given: much less alternatives for the promise (like a CP) and more alternatives for the recommendation or tip i.e. more enablers and less MP and MT. So, the stronger and more accurate the action is with the power to deliver a reward or punishment, more likely the conditional is to be perfected.

Taking into account the example above mentioned *If you study more your grades will improve*, it is possible to retrieve a counterexample, in this case, a disabler. What if I don't have aptitude for the subject matter? This is an additional background condition, also consistent with *you study more but your grades do not improve (p and not q)*. Or, it could be an alternative, what if the tests are very easy? It is consistent with *you don't study, and your grades improve anyway (not-p and q)*. If somebody rely on these assertions, it is possible to break with the conditional perfection.

As a summary, when the action is not just sufficient for the outcome, but the action is necessary for the outcome, a sufficient condition is transformed into a necessary condition, this happens with a promise, possible a CP. So, when the situation of a DA or AC are inconsistent with the promise, the listener does not retrieve an alternative for it, so the action becomes necessary for the outcome. That means that there are no alternatives and there is a CP, i.e., *(p and q) and (q and p)*. Thus, the possibility to have more information gives the listener the



chance to suppress or avoid some inferences (Byrne, 1989). That means that a possible alternative to the tip is that *the test is very easy*. If the listener knows or somebody has told him or her that *the test is very easy* and at the same time we have the information that *the grades have improved* it would be difficult to conclude that *the student studied more*. As a result, when we have a conditional and the listener does not retrieve a disabler nor an alternative, it is possible to have a conditional perfection. As a result, there is no difference between the conditional and biconditional regarding the number of disablers. But, there is a difference between the conditional and biconditional regarding the number of alternatives. If there are few alternatives and few disablers then the fully explicit models (FEM) tend to a biconditional, i.e., *p and q*; and *not-q and not p*.

## Discussions

As a conclusion, we can say that the study of the Conditional Perfection could be firstly explained as a conversational implicature, which is a pragmatic inference. We agree with Moldovan (2009) in his argument that CP should not be longer understood just as a material invalid inference that allows the speaker to make logical mistakes. Following this argumentation, CP should be analyzed as a natural phenomenon because people tend to behave with this perfective aspect when facing with certain contextual situations, for example, when promising and threatening, tipping or warning. This means that the alternatives or disabler are modularized according to the previews information and the current context. According to this, the natural tendency is to avoid the difference between conditional and biconditional in a first gaze, except when there are clear alternatives or enabler. But this has been already analyzed and explained by Wason (1966) and Johnson- Laird and Tagart (1969).

## References

- Byrne, R. and Johnson-Laird, P. (2009). ‘If’ and the problem of conditional reasoning. *Trends in Cognitive Science*, 13(7):282–287.
- Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, 31:61–83.
- Couto M., Quelhas A. and Byrne R. (2017). Advice conditionals about tips and warnings: interpretations and inferences. *Journal of Cognitive Psychology*, 29(3):364–380.
- Geis, M. and Zwicky, A. (1971). On invited inferences. *Linguistic Inquiry*, 2:561–566.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics* (pp. 41–59). Nueva York: Academic Press.
- Grice, P. (1989). *Studies in the way of words*. Harvard University Press, Cambridge. Harmondsworth: Penguin.

- Horn, L. (2000). From *If* to *Iff*: Conditional perfection as pragmatic strengthening. *Journal of Pragmatics*, 32:289–326.
- Inhelder, B. and Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. Basic Books, New York.
- Johnson-Laird, P. (1981). Cognition, computers, and mental models. *Cognition*, 10:139–143.
- Johnson-Laird, P. (1983). *Mental Models: towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Johnson-Laird, P. (2004). The history of mental models. In K. Manktelow & M.C. Chung (Eds.). *Psychology of reasoning: Theoretical and historical perspectives*, chapter 8, pp. 179–212. New York: Psychology Press.
- Johnson-Laird, P. and Goldvarg-Steingold, E. (2013). Models of Cause and effect. In *Mental Models Theory of Reasoning: Refinements and Extensions*. Florence, KY, USA: Psychology Press, 2013. ProQuest ebrary. Web. 23 October 2015.
- Johnson-Laird, P. and Khemlani, S. (2014). *Toward a Unified Theory of Reasoning. Psychology of Learning and Motivation*, First Edition, 1-42. Elsevier Inc. All rights reserved.
- Johnson-Laird, P. and Shafir E. (1993). The interaction between reasoning and decision making: an introduction. *Cognition*, 49:1–9.
- Johnson-Laird, P.N. and Tagart, J. (1969). How implication is understood. *Am. J. Psychol.*, 82:367–373.
- Jonhson-Laird, P. and Byrne, R. (1991). *Deduction*, Hillsdale, NJ: Laurence Erlbaum Assoc.ed.
- Jonhson-Laird, P. and Byrne, R. (2002). Conditionals: A Theory of Meaning, Pragmatics, and Inference. *Psychological Review*, 109(4):646–678.
- Khemlani, S., Barbey, A., and Johnson-Laird, P. (2014). Causal reasoning with mental models. *Front. Hum. Neurosci.*, 8:849. doi: 10.3389/fnhum.2014.00849.
- Khemlani, S.S., Byrne, R. M.J. and Johnson-Laird, P.N. (2018). Facts and Possibilities: A Model-Based Theory of Sentential Reasoning. *Cognitive Science*, 1–38.
- Legrenzi, P. (1970). Relations between Language and Reasoning about Deductive Rules. In *Advance in Psycholinguistics*, ed. By G.B. Flores d’Arcais, W.J.M. Levelt (North-Holland Publishing Co., Amsterdam).
- López-Astorga, M. (2016). Logic, Pragmatics, and Types of Conditionals. *TOC: Frontiers of Philosophy in China*, 11(2 Special Issue): 279–297.
- Matalon, B. (1962). *Edude Génétique de l’Iplication, in Etudes d’Epistemologie Génétique XVI. Implication, Formalisation et Logique Naturelle* (P.U.F. Paris).

- Moldovan, A. (2009). Pragmatic Considerations in the Interpretation of Denying the Antecedent. *Informal Logic*, 29(3).
- Noordman, L.G.M. (1979). *Inferring from Language*, Springer Series in Language and Communication 4.
- Paris, S.G. (1975). *Propositional Logical Thinking and Comprehension of Language Connectives* (Moutonm The Hague)
- Peel, E.A. (1967). A method for investigating children's understanding of certain logical connectives used in binary propositional thinking. *Brit. J. Math. Stat. Psychol.*, 20: 81–92.
- Roberge, J.J. and Paulus D.H. (1971). Developmental patterns for children's class and conditional reasoning abilities. *Dev. Psychol.*, 4: 191–200.
- Taplin, J.E. (1971). Reasoning with conditional sentences. *J. Verbal Learning Verbal Behav.*, 10: 209–225.
- Taplin, J.E. and Staudenmayer, H. (1973). Interpretation of abstract conditional sentences in deductive reasoning. *J. Verbal Learning Verbal Behav.*, 12: 530–542.
- van der Auwera, J. (1997). Pragmatics in the last quarter century: the case of conditional perfection. *Journal of Pragmatics*, 27: 261–274.
- van Tiel, B. and Schaeken, W. (2016). Processing conversational implicatures: alternatives and counterfactual reasoning. *Cognitive Science*, 1–36.
- Wason, P. C. (1966). Reasoning. In Foss, B. M. *New Horizons in Psychology*.

# Power to the people. The indispensable nature of the normative vocabulary and the political turn in analytic philosophy

Manuel de Pinedo-García  
Neftalí Villanueva-Fernández

The purpose of this paper is to explore the connection between two different ideas that we hold dear: on the one hand, we take normative vocabulary to be indispensable when understanding the behaviour of entities resulting from an evolutionary process is at issue; on the other, we believe that we are witnessing a genuinely novel political turn in analytic philosophy and we welcome it. We want to defend that the former thesis is far from accessory for the latter development: only by moving from a project of identifying and individuating the mental states that underlie instances of oppression to a project of detecting and intervening on actual oppressive practices can the political agenda be properly carried out.

Regarding the first thesis, we take as our starting point two classical arguments. The first one, which can be traced to Moore's anti-naturalism in meta-ethics, is opposed to both poles of a false dilemma: we reject all forms of reductionism regarding normative vocabulary (including eliminativism and fictionalism), but also the appeal to normative properties as the (spooky) entities that our vocabulary is supposed to keep track of. The second one is Wittgenstein's discussion of rule-following and Kripke's take on it: no amount of straight-forwardly factual or purely descriptive statements is sufficient to justify judging something as correct or incorrect (or, we want to add, as succeeding or failing where an evolutionary process has been involved). As we see it, Gibbard's version of naturalism (see for example Gibbard 2012) is the result of taking on board both arguments: while normative vocabulary cannot be reduced to naturalistic vocabulary, for Moorean and Kripkensteinian reasons, factual statements often have a normative import, but whether they have it or not is itself a normative, rather than a factual matter. For instance, we may think that evidence of plant cognition and communication (Calvo 2016) puts pressure on us regarding giving plants a moral consideration, but we need a prior commitment to the moral relevance of cognition and communication: even on the face of a full description of the mechanisms that enable plants to communicate, we can be fully justified in remaining morally cold towards them or even to use the description in a defence of cannibalism, as some philosophers from Murcia have claimed.

While the historical building blocks mentioned in the previous paragraph targeted our understanding of the behaviour of rational creatures, we believe the arguments work equally well when evolutionary processes, being there natural or artificial, are involved. Our only way to make sense of, say, the infamous Facebook algorithms, is by means of psychological and, hence, normative, vocabulary. In previous work, we have argued against the temptation, present on

some varieties of enactivism and other forms of emergentism, of granting a normative treatment to entities as a direct consequence of the discovery of complex “subpersonal” mechanisms. In contrast, we believe that the only intelligible sense in which the mind is emergent is that social creatures like us need to use psychological vocabulary to make sense of evolutionarily complex behaviour: minds or values are neither mysterious nor natural entities. We are natural and social beings that evaluate.

Now, in order to delimit what we mean by the political turn in analytic philosophy, we need to briefly point out what is not. There is a long history of interest in practical political issues among some of the most prominent analytic philosophers (Russell, Neurath, Dummett, Chomsky). However, although one cannot help to appreciate the philosophical background behind their political arguments, it also seems clear that there is a separation between their theoretical and their practical interests. Secondly, in general terms, the methodological practical ideals of (analytic) philosophy—a clear, precise argumentative style or a willingness to expose one’s ideas to the critical scrutiny of others—have always been applied to the study of theoretical issues within the realm of political philosophy. Bernard Williams, Rawls or Nussbaum could be cases at hand. Finally, even the goal of putting our theoretical tools to work towards the solution of practical political matters is not new either. Some of the results of Stanley’s groundbreaking work on propaganda could be an example of this.

While we admire much of the work just mentioned, we want to argue that what is new is the conviction that our philosophical instruments are as good as the role that they can play in the identification of social and economic inequalities and in supporting the project of intervening to eradicate or alleviate them. Let us give some examples from different disciplines from the analytic tradition. In epistemology, the last 10 years have seen a dramatic departure from abstract, individualistic concerns towards social issues related to the ethics of knowing. The work of Miranda Fricker is a well-known example: her commitment to a form of virtue ethics always seem parasitic on her project of identifying forms of epistemic oppression and injustice that were much less visible, if visible at all, before her work. In the philosophy of language, the objective of bringing to light analogue discursive and linguistic forms of injustice is at the center of recent work on speech act theory by philosophers such as Rebecca Kukla or Saray Ayala. It matters not whether we can accommodate our diagnosis of injustice within the best theory in the market, be it Austin’s or someone else’s, but which are the most fruitful concepts for our practical political concerns. Even in the philosophy of mind, long considered the least political form of philosophy, debates regarding the individualistic or externalistic character of mental states are giving way to an approach that emphasises the structural conditions that determine our attribution of such things as implicit attitudes or prejudices (here we can mention Ayala again, as well as Sally Haslanger or Victoria McGeer).

The last example can serve to point out where does the political turn meet the indispensability of normative vocabulary. In contrast with mainstream in philosophy of mind, where the identification of mental states in materialist or functionalist terms has dominated (something analogous could be claimed about the notion of literal meaning in the philosophy of language, or the search for the internal or external conditions for the attribution of knowledge in epistemology),

the emphasis on the social and normative aspects of specific instances of injustice has come to the fore, as a consequence of a radical change in epistemic goals: questions such as what is a belief, what is meaning or what is justification are being replaced by a focus on detecting hidden forms of injustice and on intervening on them. The assessment of the relative value of our theoretical tools for analysis is thus a normative one, since our new epistemic goals are entangled with practical political interests. Evolution has made us normative by nature and we cannot but be normative when trying to understand other evolved entities.

## References

- Ayala, S. (2016). Speech affordances: A structural take on how much we can do with our words. *European Journal of Philosophy*, 24(4):879–891.
- Calvo, P. (2016). The philosophy of plant neurobiology. *Synthese*, 193(5):1323–1343.
- Gibbard, A. (2012). *Meaning and Normativity*. Oxford: Oxford University Press.
- Kukla, R. (2014). Performative force, convention, and discursive injustice. *Hypatia*, 29(2):440–457.

# Slurs and the semantic-pragmatic divide

Jordi Fairhurst

## Abstract

Slurs are derogatory terms that have been the source of growing interest in the philosophy of language during the past decade. This paper aims to analyze the problems in semantic and pragmatic approaches to slurs in order to argue in favor of the need for a pragmatic-semantic approach. First, I will outline Hom (2008) and Hom and May's (2013) semantic approach and show its insufficiencies. Second, I will point out the problems that stem from various pragmatic approaches (e.g. Hornsby 2001; Whiting 2013; Cepollaro 2015; Lycan 2015; Allan 2016). Third, I will argue in favor of the need for a pragmatic-semantic approach, where the semantic content of a slur is pragmatically determined by the speaker's communicative attitude and his particular use of the slur.

## Key words

Slurs, derogatory terms, semantics, pragmatics.

## Introduction

Slurs have received a growing interest in the philosophy of language during the past decade. They are derogatory terms that are generally conceived as a type of hate speech that targets a particular group of people in order to discredit and damage their reputation due to their race, gender, sexual orientation or other social categories (e.g. "faggot" in reference to homosexuals). In order to account for slurs and the derogation and offense they convey two general strategies have been employed. On the one hand, semantic approaches argue that the meaning of slurs is determined by the semantic difference a slur constitutes in relation to its non-pejorative correlative (hereafter, NPC). On the other hand, pragmatic approaches argue that the meaning of slurs can be accounted for resorting to pragmatic information. The aim of this paper is to analyze how slurs function in order to argue in favor of the need of a pragmatic-semantic approach to these expressions. First I will outline Hom (2008) and Hom and May's (2013) semantic approach and show its insufficiencies. Second I will point out the insufficiencies that stem from various pragmatic approaches. Finally I will argue in favor of the need for a pragmatic-semantic approach.

## Semantic approaches to slurs

Hom (2008) and Hom and May (2013) resort to truth-conditional semantics in order to account for slurs.<sup>1</sup> Hom (2008) proposes Combinational Externalism (CE). Externalism states that the meaning of an expression is dependent on external, social practices of a linguistic community. Therefore, the meaning and the offense of a slur are semantically determined by social institutions of racism that are composed by ideology and a set of practices (Hom 2008: 430–431). The meaning of slurs can be schematized thus:

Ought to be subject to these discriminatory practices because of having these negative properties, all because of being NPC. (Hom 2008: 431)

Meanwhile, Hom and May (2013) argue that the meaning of a slur is determined by applying a PEJ function to its corresponding NPC, i.e. a function that takes the semantic value of the non-pejorative to the semantic value of the corresponding pejorative. Therefore, for the non-pejorative predicate ‘N’, PEJ(N) is the concept of being an appropriate target of negative moral evaluation on account of being ‘N’ (Hom & May 2013: 298–299).

Applying Hom (2008) and Hom and May’s (2013) PEJ function would respectively to (1) results in (2) and (3):

1. There are faggots here.
2. There are individuals here who ought to be subject to a set of discriminatory practices because of having a set of negative properties, all because of being homosexual.
3. There are individuals here who are homosexual and deserve negative moral evaluation simply because of being homosexual.

However, there are two main problems that stem from these semantic approaches to slurs that demonstrate the insufficiencies of resorting only to semantics in order to account for the meaning of slurs. First, the semantic content that differentiates a slur from its NPC cannot account for the offensiveness and the derogation that the slur expresses (Sennet & Copp 2014). Whilst (1) is derogatory and offensive, it is unclear if that is the case for (2) and (3).

Second, both proposals endorse derogative autonomy. Namely, they argue that the derogatory content and the offence of slurs are independent of the attitudes of the speakers that use the terms. Slurs are taboo expressions. Consequently, they disregard the specific use of a slur since they all are derogatory and offensive. They are able to account for appropriated uses of slurs (i.e. when a target group utilizes the slur that targets them in order to subvert its conventional meaning and eliminate the derogation it expresses) resorting purely to semantic information, since the conventional meaning is subverted. However, they are other non-derogative non-appropriated (hereafter, NDNA) uses of slurs that cannot be accounted for, such as the direct or indirect reporting of slurs. Consider (4)-(5):

---

<sup>1</sup>I will not analyze here the non-truth-conditional semantic theories that account for slurs, since they encounter the same issues.



4. Yesterday Chris yelled: “Lee is a faggot”.
5. A football fan has been banned from the stadium due to calling a player a faggot.

These uses are problematic for semantic theories since they interfere with the derogatory autonomy and the taboo they tie to slurs. The speakers who utter (4) and (5) do not intend to express contempt or offense towards homosexuals.<sup>2</sup> Cepollaro (2015: 41) extends these problematic uses for semantic approaches to the pedagogic uses of slurs (e.g. “Homosexual people are not faggots”). Another example that has not been considered previously is the ironic use of slurs. An ironic use of a slur contains derogatory content while the speaker expresses her dissociative and mocking attitude to it (Bianchi 2014), thus cancelling the derogatory content and the offence initially expressed. These NDNA uses are problematic for semantic approaches due to the non- cancellability and scoping-out of derogatory content.

On the one hand, semantic approaches generally take non-cancellability as a salient feature of slurs. Namely, the derogatory content they express is not cancellable. However, the various uses described above (i.e. direct and indirect reporting, pedagogic and ironic uses) demonstrate the contrary. Conventionally derogatory and offensive expressive expressions can be used otherwise. Another clear-cut case is my use of these expressions throughout this text, where I have intended to cancel any possible derogatory content that is usually tied to slurs. “The non-cancellability of slurs’ derogatory content is not as obvious as it is often taken to be” (Cepollaro 2015: 41). On the other hand, semantic approaches generally take scoping-out as a salient feature of slurs. It refers to the fact that the derogation and offense conveyed by slurs scopes-out of its embedded position. Consider (6)-(9):

6. Chris is a faggot.
7. Chris is not a faggot.
8. Chris is stupid.
9. Chris is not stupid.

Whilst the derogation and offense conveyed by “stupid” in (8) is not retained in (9), the derogation and offense conveyed by “faggot” in (6) is retained in (7). However, this fails to account for the NDNA uses described above, since the offense and derogation conveyed by “faggot” in (6) is not retained (4) and (5).<sup>3</sup> In sum, semantic approaches cannot account various uses of slurs since they consider that all slurs (excepting appropriated uses) express derogatory content, regardless of the speakers attitude that use the terms.

---

<sup>2</sup>However, as argued by Allan (2016: 223), the speakers must clearly indicate whether they subscribe to the slur or not.

<sup>3</sup>Hom (2012: 397) provides a possible solution. Resorting to the difference between derogation and offense, he argues that the derogatory content is not embedded in certain uses of slurs. For instance, in (7) the derogatory content is not attributed to any subject, there is no scoping-out. Any feeling of offense that may surface is just a psychological effect in the audience. Nevertheless, this account still maintains that NDNA uses express derogatory content—albeit it is not attributed to any subject. Hence, it still fails to account for certain uses of slurs and

## Pragmatic approaches to slurs

The problems outlined have led to an array of pragmatic approaches that attempt to account for slurs by focusing on their use in the specific context within which they are uttered. The derogation and offence of slurs are accounted for resorting to pragmatic information. For instance, Popa-Wyatt and Wyatt (2017: 3) argue that derogation is part of the speaker's communicative intentions, whilst offence is an achieved effect on audience members. Meanwhile, Allan (2016: 212) states that a slur is not necessarily the lexical form in a language expression 'e', but instead the perlocutionary effect of 'e' as a constituent of an utterance 'u'; the said perlocutionary effect can only be determined from 'k', the context of utterance. A few other pragmatic approaches are the following: presuppositions (Cepollaro 2015), conventional implicature (Whiting 2013; Lycan 2015) and silentism (Hornsby 2001).

Nevertheless, when resorting to a pragmatic approach (specially with far-side pragmatic theories) it is necessary to proceed with caution since there is a common issue that seems to extend to these various approaches. Consider (6) and (10):

6. Chris is a faggot.

10. Chris is a homosexual.

According to the aforementioned pragmatic approaches the difference between (6) and (10) is not semantic, it is pragmatic. That is, they operate on the assumption that there is a clear distinction between what is said and what is meant. The semantic content of a slur and its respective NPC does not differ in any way or form. In order to account for this difference we must resort to pragmatic elements. But, nevertheless, semantics is not dependent, in last instance, on pragmatics. Whilst this may not seem problematic at face value, this issue is accentuated when we take into account truth-conditions. Since the only difference is pragmatic, a slur and its correspondent NPC have the same truth-conditions. For instance, (6) and (10) would have exactly the same truth-conditions. This claim is problematic, at the very least, since it entails that derogatory expressions can be taken as equivalent to their NPC at an epistemological level.

## A pragmatic-semantic approach

The problems outlined in the two previous sections demonstrate that the semantic and pragmatic approach encounters issues when attempting to provide an explanation for how slurs function. Whilst semantic approaches are unable to account for the specific use that is carried out by a speaker when uttering a slur and the consequences this use has, pragmatic approaches are unable to show how pragmatic elements may affect semantic content or the truth-conditions of a proposition. The strengths and weaknesses of both approaches demonstrate the need for an alternative: a pragmatic-semantic approach. Namely, it is necessary to study how pragmatic information determines semantic information. It is essential to see how the communicative attitude of the speaker and his use

of a slur are semantically expressed. On the one hand (and against semantic approaches), this allows us to differentiate the various uses of slurs, without endorsing derogatory autonomy or conceiving non-cancellability and scoping out as salient features of slurs. On the other hand (and against pragmatic approaches), studying how pragmatic elements are semantically expressed allows for slurs and their NPC to have different truth-conditions. In sum, both semantics and pragmatics are necessary to account for how slurs function. However, what pragmatic-semantic theory fulfills this requirement is still open to debate and requires further investigation.

## References

- Allan, K. (2016). The reporting of slurs. In Capone, A.; Kiefer, F.; Lo Piparo, F. (Eds.) *Indirect Reports and Pragmatics: Interdisciplinary reports* (pp. 211–232). Cham: Springer.
- Bianchi, C. (2014). Slurs and appropriation: an echoic account. *Journal of Pragmatics*, 66:35–44.
- Cepollaro, B. (2015). In defense of a presuppositional account of slurs. *Language Sciences*, 52:36–45.
- Hom, C. (2008). The semantics of racial epithets. *Journal of Philosophy*, 61:343–377.
- Hom, C.; May, R. (2013). Moral and semantic innocence. *Analytic Philosophy*, 54:293–313.
- Hornsby, J. (2001). Meaning and uselessness: How to think about derogatory words. *Midwest Studies in Philosophy*, 25:128–141.
- Lycan, W.G. (2015). Slurs and lexical presumption. *Language Sciences*, 52:3–11.
- Popa-Wyatt, M.; Wyatt, J. L. (2017). Slurs, roles and power. *Philosophical Studies*, 1–28.
- Sennet, A.; Copp, D. (2014). What kind of mistake is it to use a slur?. *Philosophical Studies*, 172(4):1079–1104.
- Whiting, D. (2013). It's not what you said, it's the way you said it: Slurs and Conventional Implicatures. *Analytic Philosophy*, 54:364–377.

# Act-based propositions and a posteriori necessity: Linguistic meaning for modes of presentation<sup>1</sup>

Ainhoa Fernández

Universidad del País Vasco/Euskal Herriko Unibertsitatea

## Introduction

There are few notions in philosophy as central as *proposition* (McGrath, 2012) and few distinctions as crucial as those drawn between the *necessary* and the *contingent* and between the *a priori* and the *a posteriori*. Generally, propositions are considered to be the main bearers of truth conditions, objects of belief and other propositional attitudes (i.e., what is believed, what is doubted, etc.), references to that-clauses and the meanings of sentences (Gaskin, 2008; Moltmann, 2014). Performing one or more of the above functions, propositions play a central role in cognitive science, linguistics and philosophy. However, there is currently no overall agreement on what kind of entities they are. The fact that the propositions are the meanings of the sentences implies that, depending on which notion of proposition is being defended; different analyses can be made of the necessary or contingent character of certain statements, as well as offering different explanations as to whether we can know their semantic content *a priori* or *a posteriori*.

According to Kripke (1971, 1972/1980) there are necessary statements which can be *a posteriori*.<sup>2</sup> The canonical case presented by Kripke is the statement

(1) Hesperus is Phosphorus

Soames (2006) argues that (1) is not a genuine case of necessary *a posteriori*, showing how its semantic content is knowable *apriori*. However, this argument does not affect other statements of *a posteriori* necessity or statements that include natural kind terms, such as

(2) Water is H<sub>2</sub>O

---

<sup>1</sup>The research carried out for this work has been made possible due to the support of projects FFI2014-52196-P and IT769-13, and a predoctoral contract from the University of the Basque Country (UPV/EHU) (PIF-2016). I am particularly grateful to Manuel Pérez Otero and Agustín Vicente for their valuable comments that have undoubtedly improved this paper.

<sup>2</sup>I will use indistinctly the terms ‘necessary statement’ or ‘necessary truth’ instead of ‘necessarily true statement’; ‘contingent statement’ instead of ‘contingently true statement’; ‘*apriori* statement’ instead of ‘*apriori* knowable statement’ and ‘*aposteriori* statement’ instead of ‘*aposteriori* knowable statement’. The same applies to propositions.

So, the Kripkean thesis that there are necessary *a posteriori* statements can be maintained. In the argument against (1), a rudimentary, neutral notion of proposition is employed. More recently, Soames (2010, 2014, 2015) defended a theory of propositions whereby propositions are abstractions of types of cognitive acts. In this new position, propositions have incorporated modes of presentation. Thus allowing a very different explanation to be given of the necessary *a posteriori* statements and to “rehabilitate” case (1) as a genuine case in such a way as to have an analysis tool applicable to cases such as (1), and to natural kind statements of type (2).

In this paper, I shall argue that the analysis proposed by Soames (2015) is unsatisfactory and does not provide a convincing answer to the question of whether certain statements, such as (1) and (2), are necessary *a posteriori*. In section 1, I shall present Kripke’s argument on the genuine case of the necessary *a posteriori* statement (1) and the objection raised by Soames; in section 2, I shall explain Soames’ new theory of propositions and its application to cases of types (1) and (2), and, in section 3, I shall argue why his solution to the analysis of such cases is inadequate, and suggest an alternative that is compatible with Soames’ new notion of proposition.

## The canonical case of the necessary *a posteriori* statement and Soames’ objection

Kripke (1972/1980, 102) presents the following as an example of a necessary *a posteriori* statement:

- (1) Hesperus is Phosphorus

Suppose an individual observes a heavenly body in the evening and names it ‘Hesperus’, another individual observes a heavenly body in the morning and names it ‘Phosphorus’, and neither of them notices that it is the same heavenly body seen at different times. In other words, they have tagged the same object twice but fail to realize. Later, through observation, it is established that Hesperus and Phosphorus are the same heavenly body. But such identity was unable to be established without the help of experience, so the truth of (1) is *a posteriori*. Should we then conclude that this identity is contingent? No, if two objects are identical, they are identical out of necessity. That is, if we use ‘Hesperus’ and ‘Phosphorus’ as rigid designators of the same heavenly body and say that Hesperus might not have been Phosphorus, we are saying that an object might not have been identical to itself. But self-identity is a property that no object can lack; it is an essential property. Thus, a situation in which Hesperus is not Phosphorus is impossible. Although there is a sense that Hesperus might not have been Phosphorus, this sense is not metaphysical, it is epistemic. Therefore, the statement is necessary and *a posteriori*.

Although, in this case, it is a statement of identity in whose flanks there are proper names, the category of what is necessary *a posteriori* also includes statements in whose flanks there are propositions concerning the origin and composition of certain substances, such as ‘Nixon could not have been the child of

other parents', and also theoretical identities where terms designating natural kinds intervene, such as

(2) Water is  $H_2O$

Kripke (1972/1980, 140-144) provides an explanation of why statement (1) may seem contingent. First, he considers the possibility of a trivial explanation whereby the statement seems contingent because it is really contingent, but he discards this possibility and then offers a specific explanation of the contingency illusion. According to this explanation, since we accept that the reference of a name can be fixed by means of a description, Kripke asks us to imagine that a speaker fixes the reference of the name 'Hesperus' by means of the description: 'the heavenly body seen in such and such position in the sky in the evening' and the reference of the name 'Phosphorus' by means of the description: 'the heavenly body seen in such and such position in the sky in the morning'. Statement (1) has the form  $R1=R2$ , where  $R1$  and  $R2$  are rigid designators and the statement. If we consider the following statement

(1') The heavenly body seen in such and such a position in the sky in the evening is the heavenly body seen in such and such a position in the sky in the morning

We can see that (1') has the form  $D1=D2$ , where  $D1$  and  $D2$  are non-rigid designators, but accidental descriptions. This second statement is contingent and, according to Kripke, when we ask about the necessity of (1), what happens is that we think of (1'), which is a statement that is epistemically very close to (1) but which is, nevertheless, contingent. Thus, what we do is attribute the contingent character of (1') to (1). However, statement (1) is not contingent, it is necessary.

### ***Soames' objection***

Soames (2006) develops an objection that questions whether statement (1) is really *a posteriori*. Consider the following statements:

(1) Hesperus is Phosphorus

(3) Hesperus is Hesperus

(1) has the form ' $a=b$ ' and (3) has the form ' $a=a$ ', where  $a$  represents the name 'Hesperus' and  $b$  represents the name 'Phosphorus'. Both statements express the same proposition even though they contain different names, "they say" the same thing. Both statements express the same proposition, since  $a$  and  $b$  are two rigid designators, more specifically proper names, and the only contribution that a proper name makes to the truth conditions of the sentence or statement in which it appears is the object itself and, since  $a$  and  $b$  are rigid designators of the same object,  $a=b$  and  $a=a$  express the same proposition. On the other hand, for something to be knowable *apriori*, it is suffice that there is a way to know it *apriori*, and, indeed, there is a way to know it *apriori*; that is,  $a=a$  is knowable *apriori*. Thus, for Soames, (1) is not a genuine example of the necessary *a posteriori*.

Should we then conclude that there is no such thing as necessary *a posteriori* statements? No, we should not. There are genuine examples of necessary *a posteriori* statements that Kripke offers, such as (2).

According to Kripkean theses, theoretical identities where natural kind terms appear (the same for proper names) are true if they are true, they are necessarily true because they contain two rigid designators. Statement (2) expresses a scientific discovery, it is *a posteriori*, but this does not mean that the statement is contingent. We identify water by its macroscopic properties, but these properties do not provide the meaning of the term, they only serve to fix its reference. That is, if there were a substance that had an atomic structure different to that of water but had the same contingent macroscopic properties, we would not say that there is water that is not H<sub>2</sub>O, because being H<sub>2</sub>O is an essential property of water. The properties that science discovers with respect to natural kinds are essential properties to those kinds and the essential properties are necessarily possessed. Statement (2) is therefore necessary and *a posteriori*.

## Cognitive Act-based Propositions

Soames (2010, 2014, 2015) and Hanks (2015) recently argued that propositions are types of cognitive acts, cognitive acts of predication.<sup>3</sup> For Soames, entertaining the proposition that *o* is *P* is to predicate the property *P* of *o*. To predicate *P* of *o* is to perform an action whereby we represent *o* as *P*. This happens, for example, when we think of *o* as *P*, or perceive *o* as *P*. Unlike the traditional way of understanding this process, as a relationship between a person and a proposition; predication, as Soames understands it, is a multi-stage relationship between people, objects and properties. As in traditional conception, Soames argues that predication is neutral. Acts of predication are also the basic explanatory bearers of truth conditions and, therefore, of representational properties. Propositions are types of such actions. The proposition that *o* is *P* is the type of minimum act where a subject predicates *P* from *o*. The proposition that *o* is *P* represents *o* as *P* because it is the minimum type of act of predicate *P* of *o*, and in any conceivable token of this type, a subject represents *o* as *P*. However, Soames rejects the traditional idea that propositions have intrinsic truth conditions. According to his proposal, truth conditions are obtained through the instantiated or possible tokens pertaining to each type of act of predication.

### *A posteriori Necessity and Modes of Presentation*

In his current theory of propositions, Soames admits that some propositions have associated modes of presentation. A mode of presentation is a concrete way of knowing, presenting or referring to an object, as stated in the classic Fregean notion of presentation mode. However, the modes of presentation to which Soames appeals (2015, 68–93) are not Fregean modes of presentation but Millian;

---

<sup>3</sup>Soames' and Hanks' theories differ on several points. The arguments I put forward herein only concern Soames' version, since Hanks' version does not have the problem I will later point out.

as they are not present in all the propositions nor are they representatively relevant, although they do cause the propositions to which they are associated to be cognitively different to one another.

Bearing in mind that propositions, under this perspective, are cognitive acts of predication, let us see how two different modes of presentation make cognitive differences depending on how we interpret a given proposition in a context of first-person cognition. Taking into account the following statements:

(4) Ainhua is on fire

(4') I am on fire

(4) is the act of predication of what is burning, with the target of predication *Ainhua*, while (4') incorporates the first-person mode of presentation, the exercise of which constitutes in itself a sub-act of identification of the target of predication and which changes according to the speaker. It is in this sense that modes of presentation imply cognitive differences between propositions.

Incorporating this notion of mode of presentation, Soames has offered a different explanation of statements such as (1) and (2). It should be noted that this explanation does not affect ordinary proper names, but only specific names such as those used in (1) and natural kind terms statements, such as (2).

In the case of

(1) Hesperus is Phosphorus

(3) Hesperus is Hesperus

'Hesperus' and 'Phosphorus' are names denoting the same object, namely Venus. Those well enough informed to employ the names are expected to know that speakers who use them typically presupposes that 'Hesperus' stands for something visible in the evening and 'Phosphorus' stands for something visible in the morning. One who mixed this up would misunderstand, or at the very least not fully understand the names.

With this in mind, consider an utterance of (1) in a context in which two speakers share the presupposition that both understand the names. The speaker asserts not only the proposition semantically expressed by (1), but also what Soames calls "the linguistically enhanced proposition" which incorporates the two modes of presentation of the object, i.e., the access to Venus via the two names entertainable only by those who correctly identify the predication targets via those names.

## **An alternative: One single proposition can be necessary and *a posteriori***

If we ask whether (1) and (2) are necessary *a posteriori* statements, what answer do we get from the cognitivist theory of propositions that Soames defends? If we ask Soames whether the proposition expressed by (1) is necessary *a posteriori*, the answer is no. As we have seen, the semantically expressed proposition is



necessary *a priori*, since it is the same as that expressed by (3) and this proposition is what Soames is taking into account in his analysis, the semantically expressed by the statement (1).

Now, if the question we are asking is about the linguistically enhanced proposition, then we can say that such a proposition, in the cases of both (1) and (2), is necessary *a posteriori*. Since the speakers who claim (1) are proficient in the use of the names ‘Hesperus’ and ‘Phosphorus’. When they claim (1), they are also claiming what Soames calls “the linguistically enhanced proposition” which incorporates the two modes of presentation of the object, i.e., the access to Venus via the two names. It could be represented as follows

$$[(MP1=Hesperus) \text{ Venus}, (MP2=Phosphorus) \text{ Venus}]$$

So, when now faced with the question: *are (1) and (2) necessary a posteriori statements?*; the new answer is *no, the proposition semantically expressed by (1) and (2), respectively, is not necessary a posteriori, although another proposition associated with it, the linguistically enhanced one, is indeed*. So, Soames’ current theory of propositions, with some changes, can still provide one single proposition which is necessary and *a posteriori*, a point that Soames (2006) himself considered as an advantage in his explanation of the essentialist route to the necessary *a posteriori*.

Furthermore, we have seen that, for Soames, propositions are abstractions of cognitive types of predication, in this case linguistic acts. When we perform an act of predication, any act of speech, we refer to the object or objective of predication in a certain way and, if propositions are what Soames says they are, and ultimately depend on the cognitive capacity of the agents performing such acts, then the mode of linguistic presentation, which allows the understanding of the names, must be included in the abstraction of the act, that is, in the corresponding proposition. That is to say, since the mode of presentation provides the distinctive element of a given proposition where they intervene, it is crucial to have modes of presentation being reflected in the extracted proposition.

What Soames calls the semantically expressed proposition can be extracted from different statements, where different modes of presentation intervene and therefore make a difference between one proposition and another. This semantically expressed proposition takes up the classical notion of *semantic content*, but not *linguistic meaning*, which is at stake when it comes to modes of presentation. Linguistic meaning is what imposes conditions for the successful communication and understanding of speakers. So, if what we want to explain is the successful communication and understanding of speakers, we must prioritize linguistic meaning over semantic content in the analysis of propositions involving modes of presentation. Finally, one reason why Soames is prioritizing the semantically expressed proposition instead of the linguistically enhanced could be that he is employing a classical model of communication. This sort of model requires that speakers share the same proposition in order to have a successful understanding among them. However, maintaining this model seems inconsistent with Soames’ theory of propositions, because some modes of presentation are not shareable, as occurs in the case of first-person mode. So, if the modes of presentation are necessary for correct understanding among speakers, then a classical model of communication cannot be maintained. A coordinational model, in which the mo-

des of presentation can be computed seems much more coherent with Soames' theory of propositions. Recanati (2016) has elaborated on this kind of model, but that is a problem for another paper.

## Conclusion

We have seen how Kripke defends the existence of necessary *a posteriori* statements. We have also seen how Soames has objected to Kripke's classic example of the necessary *a posteriori* statement and how now, with his new theory of proposition, that example can be rehabilitated (1) as genuine. However, I have also note why Soames' explanation does not provide us with a satisfactory answer to the initial question as to whether statements such as (1) and (2) are necessary *a posteriori* and, finally, I have shown a way to get a direct answer to that question by prioritizing to the notion of linguistic meaning rather than semantic content, obtaining a single proposition which is both necessary and *a posteriori*. This process shows how the new act-based theories of propositions have great potential to solve dilemmas considered already classic, but still under debate in contemporary analytical philosophy.

## References

- Gaskin, W. 2008. *The Unity of the Proposition*. Oxford: Oxford University Press.
- Hanks, P. 2015. *Propositional content*. Oxford: Oxford University Press.
- Kripke, S. 1971. Identity and Necessity, in Munitz (ed.), *Identity and Individuation*, 1971, New York, New York University Press.
- Kripke, S. 1972/1980. Naming and Necessity, in Davidson y Harman (eds.); *Naming and Necessity*. Oxford: Blackwell.
- McGrath, M., Propositions, *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2014/entries/propositions>. Accessed 30 March 2018.
- Moltmann, F. 2014. Propositions, Attitudinal Objects, and the Distinction between Action and Products. *Canadian Journal of Philosophy*, 43:5–6, special issue 'Propositions and their Grasp or Understanding', edited by D. Hunter and G. Rattan, 2014, 679–701.
- Recanati, F. 2016. *Mental Files in Flux*. Oxford: Oxford University Press.
- Soames, S. 2006. The Philosophical Significance of the Kripkean Necessary Aposteriori, in Soames, S. 2006. *Philosophical Essays (vol. 2). The Philosophical Significance of Language*. Princeton and Oxford: Princeton University Press.
- Soames, S. 2010. *Philosophy of language*. Princeton, NJ: Princeton University Press.

- Soames, S. 2014. Why the Traditional Conceptions of Propositions can't be correct, King, J. C., Soames, S.,y Speaks, J. (eds.). 2014. *New thinking about propositions*. Oxford: Oxford University. Press, 25–44.
- Soames, S. 2015. *Rethinking Language, Mind, and Meaning*. Princeton, NJ: Princeton University Press.

# Causal theory and reference borrowing

**Luis Fernández Moreno**  
Complutense University of Madrid

A theory of reference provides the answer to the question of how expressions get connected (that is, refer) to an entity or to certain entities. In such a theory a distinction can be made between a theory of reference *fixing*, which explains how the referent of a term is determined for the speakers who introduced it, and a theory of reference borrowing, which explains how the reference of the term is determined for the rest of the speakers.

In Devitt/Sterelny (1999), which I will treat as the backbone of my considerations, these authors support a *descriptive-causal* theory of the (ostensive) reference fixing for proper names and natural kind terms, but a *purely causal* theory of their reference borrowing.

At the basis of the quandary that led these authors to sustain that a theory of reference fixing must contain descriptive components underlies the ambiguity of the ostension to the object or the sample of objects involved in the reference fixing. The *disambiguation* of the reference requires in the case of proper names one descriptive component and in the case of natural kind terms two, but the first of these two components—the only one I will take into consideration—is the same as the one required for proper names: a general categorial term.

By *categorial term* Devitt and Sterelny understand sortal terms in the broad sense, that is, terms that convey a criterion of *identity*. When they use the expression “general categorial term” they allude to highly general terms of that type. On this matter it is worth mentioning that they give the example of the term “animal” as the general categorial term involved in the grounding of a name for a cat. And as an example required for the grounding of the term “gold” we could propose the general categorial term “metal”.

As we have said, Devitt and Sterelny hold a *purely causal* theory of reference borrowing. The basic idea of their theory for (paradigm) proper names is that by virtue of the causal-perceptual link between a name and an object, the grounder, and other speakers at the grounding, acquire the *ability* to use the name to refer to the object. Those speakers will use the name in conversation with others and these in turn will acquire the name and borrow its reference from the former speakers—the *lenders*—by acquiring from them that sort of ability on the basis of the perception of the lenders’ use of the name; thus the acquisition of the *borrowers’* ability to refer to an object by a name also involves a causal process. In this way the name will be transmitted through the linguistic community at the same time as the abilities to use the name to refer to the object are passed on, and as new links are added to the causal chains involving the uses of the name. However, the properties that borrowers could associate with a name *do*

*not determine* the reference of the name as they use it, since they needn't be possessed by the referent of the name.

Devitt's and Sterelny's theory of reference borrowing, first put forward for proper names, is similar to the one concerning natural kind terms. Likewise, it is *not necessary* that the properties that borrowers could associate with a natural kind term be possessed by the entities of the natural kind.

Devitt and Sterelny have made several explicit assertions concerning the absence of descriptive requirements in the reference borrowing. For example, they assert that "borrowers do not have to associate the *correct categorial term*" (1999: 80; emphasis added). In a later writing Devitt asserts that "the theory of reference borrowing shows how a person can be linguistically competent with a word despite being *largely* ignorant, or even wrong, about its referent. People can be competent with the name 'Catiline' despite knowing very little about Catiline" (Devitt 2006: 139; emphases added).

In (1999) Devitt and Sterelny claim that the competence with a term consists in the ability acquired in a grounding or reference borrowing to use a term to refer to an object or to samples of objects. According to them *all* reference borrowers of a term are competent with it, but even accepting this claim for the sake of the argument, the question arises whether the borrowers' linguistic competence with a word is compatible with great ignorance or error about its referent, and in case that for this competence it is required to know "very little" about the referent, what is the descriptive component that competent borrowers have to associate with the term. The aim of this paper is to make a proposal on this matter.

Although Devitt and Sterelny do not pay attention to this fact, there are several *causal theorists* who have not rejected, or have even explicitly admitted that there are descriptive requirements in a theory of reference borrowing. I will take into consideration two of them, Keith Donnellan and Hilary Putnam.

In Donnellan (1972), he does not dispute the claim that it may be a necessary condition—although not a sufficient one—for an entity to be the referent of a term as used by the *borrowers* that such an entity satisfy some description that they associate with the term. In this regard, he does not find the claim objectionable "that our use of the name ['Aristotle'] is such that being a human being or not living in modern times, etc. are *necessary* for being the referent of the name" (Donnellan 1972: 367).

Nonetheless, since in the case of various individuals the second kind of properties can be very different, the most comprehensive unquestioned property is that of being a type of individual or entity; in that example the property of being a human being, where the term "human being" is a general categorial term.

Let's take Putnam into consideration. In (1973) Putnam holds more definitely the requirement that the borrower must associate some descriptive components with the borrowed term. He claims that "unless one has some *beliefs* about the bearer of the name that are *true or approximately true*, then it is at best idle to consider that the name refers to that bearer in one's idiolect." (Putnam 1973: 203; emphases added). Concerning this, he gives the following example: "I do not see much point, for example, in saying that someone is referring to Quine

when he uses the name ‘Quine’ if he thinks that ‘Quine’ was a Roman emperor, and that is all he ‘knows’ about Quine” (ibid.). However, this is compatible with the claim that the speaker is to associate with the term “Quine” some “minimal linguistic information [...], namely that it is a person’s name” (1973: 201), and the term “person” is a general categorial term.

I agree with a similar claim to the first put forward by Putnam above, according to which some of the descriptions or properties that users of a term and especially *borrowers* associate with it must be true or *approximately true* of the entity that constitutes its referent for those speakers to refer to that entity. In this respect, it is relevant to make at least two remarks. Firstly, in the formulation of that claim Putnam is *not* speaking explicitly about competence, but only about users of terms (to refer), who include reference borrowers—although according to the last passage by Putnam a competent speaker regarding the name “Quine” is to associate with it the general categorial term “person”. Nevertheless, since according to Devitt and Sterelny reference borrowing entails competence, I will extend the first claim by Putnam to that group of competent speakers. Secondly, the term “approximately true” is not in contexts of this sort susceptible of a precise analysis, but although Putnam does not say so, the aim of introducing the nuance “approximately true” is in my view to block the arguments from the ignorance-and-error type. To those considerations underlies the view that the borrowers cannot be *completely ignorant* or *wrong* about the properties of the entity they refer to. In a passage already quoted Devitt claims that “[borrowers can be] largely ignorant, or even wrong, about its referent [the referent of a word]” (2006: 139), but that does not amount to being *completely* ignorant or wrong about it, and he concedes in the same passage that borrowers can know “very little” about the referent. My claim concerning what that “very little” may consist in is: at least the property of being the type of entity that the referent is, which is expressed by some general categorial term and the latter has to be true or approximately true of the referent.

It is noteworthy that in (1975b) Putnam speaks more explicitly of *competence*, this time concerning natural kind terms, and according to Putnam’s view all competent speakers will have to associate with a natural kind term, implicitly or explicitly, the syntactic markers, the semantic markers and the stereotype of the term. The most relevant of these factors for this talk are the last two. However, in the cases in which the stereotype is different from the semantic markers, the main feature to distinguish the second from the first is that the semantic markers are “category-indicators of high centrality” (Putnam 1975b: 268) and hardly revisable. However, not in the case of the term “water”, but in many other cases, the stereotype will coincide with the semantic markers. Putnam gives as an example the term “molybdenum”, where the stereotype and the semantic marker are the same, i.e., that molybdenum is a metal.

Since we are interested in the question of whether there are descriptive components involved in the reference borrowing, in this case of natural kind terms, the answer will be affirmative if at least the properties contained in the *semantic markers* (or at any rate, properties approximately identical to them) are involved therein, and they will be expressed by general categorial terms.

I already mentioned an assertion in Devitt/Sterelny (1999) and in other works by Devitt questioning the necessity of including descriptive components in a

theory of reference borrowing or at least somewhat reluctantly conceding that the reference borrowing may comprise some *small* descriptive components, although Devitt avoids entering into this question. Thus, assuming that words express concepts, Devitt asserts that “the theory of reference borrowing places *very little* epistemic burden on the linguistically and conceptually competent [...] There is, of course, room for argument about *just how little* an epistemic burden should be placed on the competent, but *we need not join this argument*” (Devitt 2006: 139; first and last emphases added). And he hesitantly gives as an example of the descriptive component required for the reference borrowing of a word, or of its corresponding concept, that of the *type* of entity the referent is: “Perhaps there is some small epistemic burden on the person’s conceptual competence [...]; for example, perhaps the concept *⟨Aristotle⟩* has to be associated with the concept *⟨human⟩*” (Devitt 2006: 40).

According to the assertions by Putnam, partially by Donnellan, and more hesitantly by Devitt—despite Devitt’s and Sterelny’s asseverations in (1999) to the contrary—it is plausible that competent borrowers will have to associate some descriptive component which is true, or at least approximately true, of the referent of the terms, proper names and natural kind terms, which they borrow. In this regard the least questionable descriptive component is very general, i.e., the one concerning the *type* of entity referred to, which will be expressed by some *general categorial term*—and hence by the indefinite description formed with it. If the speaker is *completely* ignorant or wrong about the type of entity the referent is, it can be questioned that the borrower be a competent speaker.

Of course, the latter claim depends on what is required to be a competent speaker. As already said, in Devitt’s and Sterelny’s view in (1999) reference borrowing entails competence, but we can leave aside that specific view of competence and assume a more theory-neutral view of competence in a language, which they characterize as “the ability to produce and understand sentences with the sounds and meanings of that language” (Devitt/Sterelny 1999: 188; Devitt 2006: 201).

According to that theory-neutral view, our judgment on (lexical) competence depends on our conception of understanding and meaning. The authors who adopt a view of meaning strongly relying on a causal theory of reference will support a purely causal theory of competence. A view of that sort is proposed by Devitt and Sterelny, who identify the *sense* of a proper name mainly with “the property of designating its bearer by a certain type of causal link between name and bearer” (Devitt/Sterelny 1999: 67). Although hardly anyone else has shared that view of sense, this should be mitigated by Devitt’s claim already quoted according to which “the theory of reference borrowing places very little epistemic burden on the linguistically and conceptually competent” (Devitt 2006: 139), but “very little” is still something. And although he avoids dealing with the question about what that “very little epistemic burden” should consist in, in the example he hesitantly gives, as indicated above, that “burden” concerning the proper name “Aristotle” is expressed by the general categorial term “human”, which conveys the type of entity that Aristotle is. Another view of that sort, but different from the one held by Devitt and Sterelny, is the one embraced by advocates of the direct reference theory. However, even some of them also concede hesitantly that a competent speaker regarding the term “water” has to

associate with this term the property of being a liquid (see Soames 2005: 184).

In fact, it is plausible that a competent speaker concerning the word “water”—i.e., who understands that word—associates with it at least the property expressed by the general categorial term “liquid”, one regarding the word “Aristotle”, the property expressed by the general categorial term “human being”, etc., or some approximately identical properties. If such a view is accepted one should also admit a certain sort of epistemic component in the notion of competence, and that component is constituted by the property expressed by some general categorial term that conveys a general property of the referred entity, the type of entity it is. This view of competence, however, gives rise to some questions and, in particular, the two following ones. Firstly, since there are many general categorial terms that can be associated with a term which express properties that are true or approximately true of its referent, the question arises regarding what to say about a speaker who associates with the term some of those properties, but not others. Let’s assume that a speaker knows that Quine is a human being, but not a philosopher. From my point of view this speaker is competent insofar as he is knowledgeable about a general property that is true of Quine, although he is not as competent as other speakers that know that Quine is a human being and a philosopher. Competence, at least according to an epistemic view of it, is a matter of *degree*. Secondly, the question could be raised as to the necessity or sufficiency for the competence concerning a term of properties expressed by general categorial terms. On this matter my view is quite modest: the minimum necessary and sufficient condition for the competence about a term is expressed by some general categorial term that is true or approximately true of its referent.

At this point, it can be argued as follows. Speakers can be divided into different sorts. On the one hand, the grounders of a term, who associate descriptive components with the term to fix the reference of proper names and natural kind terms; on the other hand, the *competent borrowers* of the term, who associate with the term some general categorial terms that express very general properties that are true—or approximately true—of the referent, and that convey the type of entity referred to. Lastly, those speakers who, although having borrowed the term, are completely ignorant or wrong about the properties, even the most general ones, possessed by the referent of the term. Only the first two sorts of speakers are *competent*. Concerning the latter it could be claimed that even if they were to refer to an entity by the use of a term according to a purely causal theory of reference borrowing, they would have no idea as to what they refer to. Accordingly, although they could be borrowers regarding a term, they are not competent speakers concerning it.

Thus, my contribution to the debate concerning the theory of reference borrowing is that, adopting a moderately epistemic view of competence, at least the descriptive component required for the reference fixing of proper names and the first one required for the reference fixing of natural kind terms in Devitt/Sterelny’s theory in (1999), i.e., *some general categorial term*, is also a requisite for the competence of reference borrowers. Therefore, as long as causal theorists consider borrowers to be competent speakers, they should maintain a *descriptive-causal theory of reference borrowing*, which involves causal chains in addition to some general categorial term, which is true or approximately true



of the referent of the term.

## References

- DEVITT , M. (2006): *Ignorance of Language*. Oxford: Clarendon Press.
- DEVITT, M. & K. STERELNY (1999): *Language and Reality. An Introduction to the Philosophy of Language*. Cambridge, Mass.: MIT Press, 2<sup>nd</sup> revised and enlarged edition; 1<sup>st</sup> edition, 1987.
- DONNELLAN, K. (1972): “Proper names and identifying descriptions”. In D. Davidson and G. Harman (eds.), *Semantics of Natural Language*. Dordrecht: Reidel: 356–379.
- PUTNAM , H. (1973): “Explanation and reference”. In G. Pearce and P. Maynard (eds.), *Conceptual Change*. Dordrecht: Reidel. Reprinted in H. Putnam (1975a): 196–214.
- PUTNAM , H. (1975a): *Mind, Language and Reality. (Philosophical Papers, Vol. 2)*. Cambridge: Cambridge University Press.
- PUTNAM , H. (1975b): “The meaning of ‘meaning’ ”. In H. Putnam (1975a): 215–271.
- SOAMES , S. (2005): *Reference and Description. The Case against Two-Dimensionalism*. Princeton: Princeton University Press.

# The value of vagueness

David Lanius

Almost since the beginning of the philosophical debate, vagueness has been considered something bad. It leads to the Sorites paradox, borderline cases, the violation of the logical principle of bivalence, and generally to problems of justification in reasoning with language. Most theories of vagueness give up on certain logical principles in order to accommodate the vagueness of natural language and save logic from paradox and inconsistency. This is a *prima facie* reason to see vagueness as a problem.

Nevertheless, there is an increasing number of people who claim that vagueness has a value. The most prominent proponents of this view are Bertrand Russell, Max Black, Ludwig Wittgenstein, Friedrich Waismann and, more recently, Prashant Parikh, Manfred Pinkal, Renate Bartsch, Nora Kluck, Manfred Krifka and Kees van Deemter. They argue that we often use vague terms when we could have used more precise ones instead because their vagueness itself has a positive function. In my talk, I will discuss a series of arguments given in favor of this view and try to refute them. There are (at least) five types of structurally similar arguments that we can find in the literature.

First, there is the *Argument from Perception and Memory*. It has been argued that vagueness is useful because it more adequately represents our perceptual experience. Also, human memory retains only a vague gist of information due to evolutionary economization of memory space. There is simply no need for more precision because we perceive and remember the world only vaguely anyway.

Unfortunately, however, this argument does not show what it purports to do. Maybe our perception and memory are such that we cannot but think and speak vaguely. Maybe vagueness is a necessary feature of human thought and language. But this does not entail that vagueness fulfills a positive function. Even if the argument is successful, vagueness would still be a defect, albeit an unavoidable one.

Second, there is the *Argument from Cognitive Costs*. It is, arguably, easier to pronounce the term “many people” than “324,542 people.” It is, arguably, also easier to hold the thought in mind or to learn the term. Examples like this one are often used in an attempt to demonstrate that vagueness is valuable because it makes it easier for us to use language.

However, an imprecise term such as “many people” is not only vague, but also general, i.e., “many” covers much more cases than “324,542.” A more coarse-grained degree of granularity allows for more generality—a property of language which clearly does make information processing easier. In contrast, there is no reason to believe that vagueness—the property to allow for borderline cases—is functional here in any way.

Third, there is the *Argument from Language Change and Learning*. Vagueness is supposed to make language more adaptable. In particular, it has been argued that the gradual widening of a term's extension along a Sorites series over time and by a significant group of speakers can facilitate a change in meaning which satisfies the communicative needs of a language community.

While there might be some small contribution of vagueness to language change, other properties of language are clearly more efficient to facilitate long-term adaptability—as, for instance, the figurative use of words. Context-dependence is much more useful for this end than vagueness, which at best can smooth the transition actuated by other processes. There is thus no positive argument for the claim that vagueness allows for gradual change in meaning. Neither Sorites susceptibility, tolerance nor strategies of over- and under-generalization entail that vagueness plays any interesting role in supporting language change or learning.

Fourth, there is the *Argument from Communicative Success*. Some philosophers have argued that vagueness may be necessary and sometimes a flaw of language, but it does not preclude successful communication and is thus not a defect by itself. They infer from this that vagueness is valuable.

This is, however, a non-sequitur. Most prominently, Prashant Parikh used a simple game theoretic model to supposedly prove the value of vagueness. His model and also all other examples used in the argument either show only that it is better to use vague language than to stop communication at all or they face the same difficulties as the examples in the *Argument from Perception and Memory* and the *Argument from Cognitive Costs*. Either another property of the terms (such as generality) improves communicative success by making judgments more stable or the gains provided by ease of expression outweigh the risks due to possible borderline cases. In neither case, however, is it vagueness that serves a positive function.

Fifth, there is the *Argument from Inefficiency*. It is often said that the use of precise terms requires unnecessary and inefficient measurement. Vagueness, in contrast, makes ordinary language terms fit for everyday use. Without vagueness we would have to measure all day long the number of grains, the height of people and the wavelength of light when all we want is to talk about heaps, tall people and blue books. In a nutshell, vague language is efficient.

It is correct that measurement is necessary if the case at hand is close the sharp boundary of a precise term and we want to be certain about its application to it. In contrast, if the case is close to the fuzzy boundary of a vague term, we are necessarily ignorant about its application. Both precision and vagueness thus result in ignorance if we are close to the (sharp or fuzzy) boundary of the term's extension. Note, however, that precision has the additional advantage that ignorance can be resolved by measurement. Precision does not require measurement but makes it possible.

Consider the term “many” again. As pointed out above, it is not only vague but also more general than “324,542.” Moreover, it is relative to a contextually valued standard and can thus vary in its extension in a way “324,542” or even “more than 1,000” cannot. Nor do these precise expressions lean themselves towards figurative or other pragmatic uses as do terms such as “many.”

Furthermore, if a speaker calls a number of people “many,” she might not only say something about the earthquake but give also a value judgment. She could provide information about her values and assessment of the state of affairs. As Frank Veltman put it, the speaker adds an opinion. She can pragmatically imply that the number is high relative to other contexts that she considers comparable.

Such pragmatic functions have hardly anything to do with vagueness. Surely, vague terms can be precisified differently by different persons in different contexts. But before they can precisify them, they need to determine a context of application or precisification. This requires them to settle on a particular sense of “tall,” “blue” or “many.” The impossibility to foresee in which ways context affects the interpretation of these terms stems from their property of being context-sensitive—not from their vagueness.

In summary, there are no conclusive arguments for the claim that vagueness has a value, once we properly distinguish it from other linguistic phenomena. I will now conclude with a general argument for the contrary claim that vagueness is systematically overrated and nearly irrelevant for most purposes and in most situations. It runs as follows.

First, borderline cases are rare. We typically use vague terms for clear cases and we typically avoid borderline cases wherever possible. The reason lies in our communicative aims, which require (at least a minimum of) clarity.

Second, there is hardly ever clarity or agreement about whether a particular case is borderline. There is a structural second-order unclarity with regard to vagueness (which is independent from higher-order vagueness). The reason is that before we can determine whether a case is borderline or not, we need to determine the communicative content of the utterance in which the vague term occurs.

Philosophers and linguists are often misled about how easy it is to agree on borderline cases, since the scenarios discussed in the debate are highly idealized. Contextual details and other forms of indeterminacy are stipulated away. These, however, are precisely the variables that make it so difficult to lie hands on vagueness. In any ordinary language interchange, the communicative content of the utterances can be determined in multiple ways. This makes it not only hard to find out whether there is a borderline case or not, but, most strikingly, Gricean maxims generally force (1) the speaker to make utterances and (2) the listener to interpret them such that no borderline cases occur. There is a structural reason to avoid (recognizing) borderline cases—even when they were there.

Thus, the (second-hand) unclarity of borderline cases cannot properly be used positively. Neither speaker nor listener will know in advance what the (illocutionary and perlocutionary) effects of the vague utterance will be, since it will always (to some non-negligible degree) be uncertain whether any given case is borderline. Other forms of indeterminacy are much easier to foresee and are thus much better suited to be used strategically. For instance, there can be stable agreement about whether some utterance is ambiguous or not—even if people continue to disagree about how to resolve the ambiguity.

The discussion of the arguments for the value of vagueness above make it clear

that vagueness is frequently and easily confused with other linguistic phenomena such as polysemy, standard- relativity, generality, and several pragmatic forms of indeterminacy. But once we properly distinguish it from these other phenomena, it becomes clear that vagueness is (at the very least) systematically overrated and much less relevant than commonly assumed.

# A defense of the common ground view about assertion

Josep Macià

In this talk I will examine Abbott's (2008) claims that Stalnaker's (1978, 2002, 2009) widely influential views about both presupposition and assertion (that many of us make use of) are wrong. She calls those views the *common ground view of presupposition* and the *common ground view of assertion*. Abbott claims that the common ground view is not able to distinguish presupposition from assertion, and that it is subject to many different kinds of counterexamples.

Here we will focus specifically on Abbott's criticisms of the common ground view of assertion. According to this view (e.g. Stalnaker 2002) conversation takes place on the basis of a set of assumptions about the beliefs that all the participants in the conversation share (or are disposed to behave, for the purposes of the conversation, as if they share). Abbott claims that according to the common ground view about assertion:

- (a) Speakers "intend their utterances to reduce the set of possible worlds consistent with these beliefs" (Abbott 2008, p. 531), and that
- (b) "To make an assertion is to reduce the context set (i.e. the common ground) in a particular way" (p. 532).

Abbott points out that there are a number of clear counterexamples to the Stalnakerian view of assertion. For instance, when uttering (1) in an university elevator

- (1) The term will soon be over

the speaker will typically have no intention of being informative, but just to avoid an uncomfortable silence in the elevator; (1) is not providing any new information and so neither (a) nor (b) will hold in this case.

Consider also:

- (2) Our speaker tonight is Noam Chomsky. (Abbott 2008, p. 532)  
[when uttered by the person who introduces the speaker; everyone attending the talk, though, does of course already know who the speaker is;  
(2) does not provide any new information that can be added to the common ground; still, though, this utterance of (2) is completely adequate and felicitous, contrary to what the Stalnakerian view would seem to predict]

Other examples that could be used to make related points are (3)–(6): (3–5 are taken or adapted from Abbott 2008)

- (3) (Reminder) You have a dentist appointment next Tuesday, but we've just agreed that I'll pick Sue up.  
[in a case where it is common knowledge that the addressee already knows that she has a dentist appointment]
- (4) We all know that Bill will never come clean of this.  
[if it is true that we all know it, then, it would seem that (a) and (b) cannot hold, and the utterance should be predicted to be infelicitous]
- (5) Cats are cats.
- (6) (Repetition) We will fight for our rights. We will fight for our rights.

I will argue that the Stalnakerian view of assertion is perfectly compatible with the different facts that these very interesting examples bring into notice. Some key points in my argument will be:

- (I) Abbott partially misdescribes the Stalnakerian view of assertion. The defining feature of assertion is not that it reduces the common ground, but rather that (c) it prescribes that its content be added to the common ground in the way that Stalnaker describes (when making an assertion the speaker is prescribing to the audience that they add the assertion's content to the common ground, if no one objects, the content will be added; if they do, it will not).
- (II) The fact that conversations are subject to Grice's Cooperative Principle (*make your contribution as it is required by the accepted purpose of the talk exchange*) when combined with (c) allows us to explain why usually (but, crucially, not always) assertions do reduce the common ground and are inappropriate if they do not: often the, or at least, one of the, purposes of a conversation is to convey information; to make an assertion (that will require that the audience adds to the common ground some content) would not serve the purpose of the conversation (when that purpose is to convey information) if it so happens that that content is already part of the common ground.  
It might be, though, that at some specific point in a talk exchange the purpose of that talk exchange is not to convey information (as in the elevator situation, where the purpose of the utterance is just to avoid the uncomfortable silence).
- (III) As already pointed out by Stalnaker and others, making an assertion will change the common ground not just by prescribing to the audience that they add its content to the common ground but, also, because it will cause that some new fact becomes part of the common ground: that the speaker uttered a certain sentence with certain content.
- (IV) Beliefs can be activated or just implicit (for instance, 1 minute ago the reader already believed that there are no flying elephants; now this belief is activated, 1 minute ago, most likely, it was not). Sometimes the point of making an assertion might just be to activate its content (by means of the mechanism of prescribing the audience to try to add that content to the common ground). Depending on the purposes of the conversation, therefore, acting so as to activate a content (something that it is achieved

by the prescribed assertion operation) will completely fulfill the purposes of the conversation and, so, be completely appropriate.

- (v) We should distinguish between it being the case that some content is known by all the participants in a conversation, and it being the case that that content is part of the common ground.

In the oral presentation I will argue that we can account for each of (1)–(6) on the basis of the Stalnakerian conception of assertion and (i)–(v); and I will, therefore, conclude that the significant challenge that Abbott (2008) poses for the common ground view of assertion can be answered in a satisfactory way.

## References

- Abbott, B. (2008). Presuppositions and common ground. *Linguistics and Philosophy*, 21, 523–538.
- Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics, volume 9: Pragmatics*. (pp. 315–322). New York: Academic Press.
- Stalnaker, R. C. (2002). Common ground. *Linguistics and Philosophy*, 25, 701–721.
- Stalnaker, R. C. (2009). On Hawthorne and Magidor on Assertion, Context, and Epistemic Accessibility. *Mind*, cxviii, 470 (April 2009), 399–409.



# Inferentialism, representationalism and moral responsibility

Llanos Navarro-Laespada<sup>1</sup>

María José Frápolli

University of Granada

The application of Brandom's inferentialism to the study of the meaning of ethical discourse (Price 2013, 31) is systematically accused of preventing the justification of the truth of some ethical beliefs and the falsehood of some others. This talk assumes, as a general principle, that a semantic-pragmatic theory aims to explain what speakers do with their words—not the truth value of their utterances—and defends two thesis. Firstly (section 3.1.), that the objection assumes a solipsistic understanding of knowledge, which is excluded by inferentialism. Secondly (section 3.2.), that the view from which the objection is addressed—the representationalist theory of meaning—is unable to explain how ethical knowledge is possible.

## The inferentialist view

Inferentialism is the thesis that the propositional content of a claim must be individuated according to its inferential relations (Brandom 1994, 91), that is to say, according to what follows from it in both a theoretical and a practical sense. Thus, the inferentialist explanation of the meaning of ethical claims differs from the representationalist view, according to which propositional contents are determined by the representation of states of affairs (Brandom 1994, 69). For inferentialism, a proposition is anything that can act as a premise or a conclusion in an inference. And given that ethical claims act in fact as premises and conclusions in inferences—as the Frege-Geach argument (Geach 1960, 223) shows—these claims have propositional content, that is, they can be true or false.

## Objection

According to a recurring objection, the inferentialist notion of proposition makes it possible for any ethical claim to be true, because it is always possible to accept the set of premises that makes it true. It might happen that a certain ethical claim turns out to be true given the belief system of one speaker and false given the belief system of another. And since no one can judge from another perspective

---

<sup>1</sup>This work has been funded by the *Ministerio de Educación, Cultura y Deporte* (FPU16/04185).

than that provided by her own belief system, in case of disagreement no one would be in a position to determine who is right.

## Response

### 1

The Brandomian notion of knowledge excludes the possibility that all propositions are equally likely to be accepted as premises of our reasoning.

Within the framework of inferentialism, in order to explain the concept of knowledge we have to focus on the function performed by ascriptions of knowledge. Brandom himself has held that we do three things when we ascribe to someone the knowledge that  $p$ : (i) we attribute the commitment that  $p$ , which plays the part of belief in the classical tripartite definition of knowledge as justified true belief, (ii) we attribute an entitlement to that commitment, which plays the part of justification, and finally (iii) we undertake that very commitment ourselves (Brandom 2000, 119), which is the part of truth. When we say that somebody knows that  $p$  we are, among other things, saying that  $p$  is true. An ascription of knowledge involves an expression of support on the part of the speaker who makes the attribution to the claim made by the speaker who receives the attribution. The concept of knowledge is used to attribute normative statuses to speakers, which in Brandom's picture are commitments and entitlements.

From this point of view, knowledge can only be understood as a collective undertaking, in which the corpus of knowledge a community sticks to is the result of an exchange of reasons that generates a well-established belief system. Affirming is engaging in the game of giving and asking for reasons, which is essentially communal. The propositions that we accept as premises are those that are settled in the course of this social game, so that solipsism and relativism are discarded. In order for a proposition to be accepted by a group of speakers, it is necessary either that the proposition fits with their previous belief system or that there are reasons which motivate its change. In order to assert a proposition, it is not enough for it or for the propositions from which it follows to be thinkable. On the other hand, from the fact that individual speakers are situated subjects does not follow that they have an impervious point of view. The exchange of reasons modifies the belief systems of individual speakers and allows correcting the limitations of their respective perspectives. When someone makes an ethical claim that we think is false, we can ask for reasons and offer our own.

### 2

When it comes to deciding which ethical claims are true, the representationalist theory of meaning does not place us in a better position than inferentialism. According to representationalism, there are ethical truths because there are ethical facts—facts involving ethical properties—which make true the claims which represent them. However, Moore's open-question argument (1903/1993, 67) shows

that if ethical predicates refer to properties, these must be non-natural properties. Thus, the empirical means that would allow for assessing whether a claim whose predicates refer to natural properties is true would not be applicable in the case of ethical claims. It would be impossible to make a connection between the truth conditions of ethical claims and any relevant event connected with the practices of the subjects who are supposed to have ethical knowledge, so it is difficult to see how that knowledge would be possible (Benacerraf 1973, 673). Postulating a moral intuition that allows grasping non-natural properties is an *ad hoc* solution and is of no help to mediate in cases of ethical disagreement. The representationalist theory of meaning offers an account of the semantic of ethical discourse that allows maintaining that there are ethical truths, but it does so at the price of sacrificing the possibility that humans know them. If ethical knowledge depends on a correspondence with facts that we cannot verify, then no one is in a position to confirm or discard any ethical claim. One can only suspend judgement.

## Conclusion

Contrary to the usual objection, representationalism—and not inferentialism—is the theory that, by its very nature, prevents the justification of the truth of certain ethical claims and of the falsehood of others.

## References

- Benacerraf, P. (1973). “Mathematical Truth”. *The Journal of Philosophy*, 70(9):661–679.
- Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- Brandom, R. (2000). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.
- Geach, P. (1960). “Ascriptivism”. *The Philosophical Review*, 69(2):221–225.
- Moore, G.E. (1903/1993). *Principia Ethica*. Cambridge: Cambridge University Press.
- Price, H. (2013). *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.

# La representación mental del sonido: ¿formalismo lingüístico o realidad cognitiva?

Nuria Polo  
UNED

La realidad fónica, en cuanto entidad fisiológica, es continua y universal, pero para poder hacer uso de ella se tiene que codificar de forma discreta, arbitraria y simbólica en la mente de los hablantes. El objetivo de la disciplina fonológica es precisamente ese: tratar de averiguar cómo se almacena y se procesa el sistema de sonidos de una lengua en la mente de un hablante. Esto implica descubrir las abstracciones cognitivas que sustentan los sonidos físicos, o los signos en las lenguas signadas, y entender qué combinaciones arbitrarias de elementos están permitidas en cada lengua. Es decir, su objetivo último es poder describir el conocimiento implícito y tácito que los hablantes tienen sobre los sonidos de sus lenguas y que, por tanto, los hablantes de una misma comunidad lingüística deben compartir para poder entenderse entre sí.

El sistema de elementos estructurado que conforma la cognición fonológica, es decir, los sonidos del lenguaje en la mente, se suele denominar *representación fonológica*. La mayoría de los autores están de acuerdo en que la naturaleza de la representación es universal, en la medida en que todos los seres humanos comparten la misma cognición, y en que las unidades específicas de esa representación varían según las lenguas; no obstante, existe una gran polémica sobre cuáles o cómo son esas unidades.

Tradicionalmente se ha creído que en la memoria a largo plazo se almacena un grupo muy reducido de unidades, que se procesan en la memoria a corto plazo o de trabajo a partir de instrucciones ordenadas en serie (Chomsky y Halle, 1968). Además, estas unidades son de dominio específico, es decir, no pueden interpretarse por otros sistemas de computación mental aparte del fonológico. Con la irrupción del conexionismo en la teoría fonológica, la ordenación de reglas en serie se reinterpreta como restricciones que actúan en paralelo y la abstracción máxima de unidades desaparece (Smolensky y Prince, 1993/2004). Tanto si es en serie como en paralelo, el procesamiento de unidades sirve para transformar la representación fonológica en las emisiones reales del sonido.

Sin embargo, el conexionismo actual propone un enfoque mucho más radical que los postulados previos al negar la existencia de la representación fonológica, ya que esta solo tiene razón de ser en un sistema que necesite unidades sobre las que operar. En este sentido se asume que todas y cada una de las emisiones que un hablante oye o produce se almacenan en la memoria a largo plazo y cuando se han almacenado suficientes ocurrencias en la mente de un individuo,

se fija un ejemplar de esa unidad, con lo que el almacén fonológico se encuentra en constante redefinición (Pierrehumbert, 2001). Además, se considera que la fonología es un tipo más de cognición, sujeta a las mismas restricciones de procesos mentales más generales como la asociación intermodal, la automatización neuromotora, la esquematización o la categorización (Bybee, 2001). Como consecuencia, los elementos fonológicos pasan de ser considerados universales, e innatos, a ser considerados como parte de un sistema emergente que se construye a partir de las relaciones entre el hablante y el oyente.

El debate actual gira en torno a demostrar o rechazar mediante argumentos lingüísticos, que pueden ser empíricos o no, la existencia de las unidades que conforman dicha representación. Por ejemplo, en español es muy común que la [d] intervocálica caiga en los participios de la primera conjugación (*compra'o* por *comprado*). Este rasgo se ha explicado tradicionalmente mediante la postulación de una regla que transforma la consonante oclusiva de la representación profunda en una consonante espirantizada en la estructura superficial y a su vez una segunda regla que eliminaría la consonante debilitada. ¿Por qué cae la [d] intervocálica? Porque en español existe una regla fonológica de debilitamiento y ¿por qué se sabe que existe esta regla? Porque cae la [d]. Para Ohala (1990) este argumento es circular y parece que gran parte del formalismo usado tradicionalmente en la teoría fonológica se construye *ad hoc*. Es decir, se incurre a menudo en este tipo de descripciones que en muchas ocasiones carecen de valor predictivo o explicativo. Para evitar esto, una rama de la fonología actual postula que se almacenan todas y cada de las formas, percibidas y producidas, con información muy detallada, tanto lingüística como no lingüística, y se construyen las nuevas formas mediante analogía con formas previas similares. Este mismo mecanismo analógico explicaría también los procesos diacrónicos de elisión de [d] que ocurren con una distribución similar (Estrada, 2018).

Por el contrario, Albright y Hayes (2003) proponen que se representarían las dos formas, *compra'o* y *comprado*, y la [d] se eliminaría en más casos según la frecuencia de uso de las palabras, es decir, esto ocurriría más veces con el verbo *comprar* que con el verbo *brindar*. Si este comportamiento se puede explicar a partir de reglas, estas se deberían aplicar sin excepciones independientemente del contexto, pero si se utiliza como recurso explicativo la analogía de un ejemplar a otro, el contexto de la palabra o de la oración puede influir en la comparación y afectar a la cognición. Estos autores utilizan un modelo estadístico (*Minimal Generalization Algorithm*) para testar empíricamente la teoría fonológica descrita formalmente.

Lo que pone de manifiesto este ejemplo es que cualquier fenómeno fonológico se puede describir desde cualquiera de las concepciones existentes de representación mental del sonido, asumiendo que existen unidades de procesamiento o sin la necesidad de postularlas. Así que, a pesar de los avances de los últimos años, la teoría fonológica no ha podido resolver ni cómo es la representación mental de los sonidos, en caso de que exista, ni cómo es la naturaleza de su procesamiento. Por ello, en la presente charla se revisarán argumentos como los del ejemplo anterior para tratar de aportar luz sobre este problema epistemológico y en última instancia invitar a la reflexión y discusión sobre la ciencia fonológica entre filósofos y lingüísticos.

## Referencias citadas

- Albright, A. y Hayes, B. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition*, 90, 119–161.
- Bybee, J. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Chomsky, N., y Halle, M. 1968. *The sound pattern of English*. Nueva York: Harper and Row.
- Estrada, A. 2018. Factors involved in the evolution of intervocalic /d/ in the Iberian Peninsula: The case for frequency. En Recasens, D. y Sánchez Miret, F. (eds.) *Production and Perception Mechanisms of Sound Change*. Munich: LINCOM, 145–156.
- Ohala, J. J. 1990. There is no interface between phonetics and phonology. A personal view. *Journal of Phonetics* 18, 153–171.
- Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. En J. Bybee y P. Hopper (eds.). *Frequency effects and the emergence of lexical structure*. John Benjamins: Amsterdam, 137–157.
- Prince, A. y Smolensky, P. 1993/2004. *Optimality Theory*. Oxford: Blackwell.

# “La aproximación fácil” a la ontología de Amie L. Thomasson: pragmática y pragmatismo entrelazados

M<sup>a</sup> Uxía Rivas Monroy  
Dpto. de Filosofía y Antropología  
Facultad de Filosofía  
Universidad de Santiago de Compostela  
uxia.rivas@usc.es

## Introducción

Esta presentación parte de la propuesta de Amie Thomasson, que ella denomina “aproximación fácil” a la ontología, y pretende completarla con las aportaciones del pragmatismo en relación con la posición que ella mantiene de que las preguntas ontológicas internas relativas a un marco lingüístico son solo preguntas relativas a la semántica o pragmática de las expresiones. La consideración de Thomasson pone el énfasis, principalmente, en el uso de los términos del lenguaje, de ahí que las preguntas ontológicas internas se puedan entender como relativas a las reglas que determinan los usos de los términos implicados en las mismas, especialmente, el término “existir”. El objetivo final de este breve escrito es poner de relieve que lo que Thomasson asume como una cuestión interna de tipo semántico/pragmático podría también entenderse desde un punto de vista pragmatista, lo que daría profundidad filosófica a lo que queda implícitamente asumido bajo los conceptos generales de uso, regla o condiciones de aplicación. Esta manera fácil de responder a las preguntas ontológicas propugnada por Thomasson deja, en general, un poso de insatisfacción, pues preguntas de tanto calado filosófico como ¿qué hay?, ¿existen los números o las proposiciones? son resueltas a la manera carnapiana, simplemente atendiendo a las condiciones de aplicación de los términos. Si bien esta es la posición que explícitamente defiende Thomasson, y por ello las respuestas a las preguntas ontológicas son fáciles de contestar, creo que la posición pragmatista, sin contradecir la posición de Thomasson, podría proporcionar una respuesta filosóficamente más satisfactoria, si bien también más discutible.

El desarrollo del trabajo comenzará con una breve exposición de la aproximación fácil a la ontología defendida por Thomasson, destacando principalmente el aspecto en que su posición está en consonancia con la distinción de Carnap de preguntas de existencia internas y externas a un marco lingüístico y, que como ella misma manifiesta, se puede también comprender como la distinción uso/mención, lo que le permitirá justificar una respuesta deflacionaria a las preguntas de existencia. A continuación, indicaré que es en el *uso* de los términos

donde la ontología “dura” implícitamente se esconde, porque el lenguaje mismo proyecta el compromiso ontológico justamente en la manera en la que lo usamos. La aproximación fácil a la ontología es fácil precisamente porque da por supuesto esta perspectiva ontológica subyacente al uso del lenguaje, apoyándose simplemente en que es ese uso del lenguaje el que conlleva el compromiso ontológico. Cualquier posición filosófica ontológica dura, en concreto, cualquier epistemología ontológica que pretenda dar cuenta de la asunción de compromisos ontológicos, presenta problemas insalvables, según Thomasson, que impide tomar una posición clara por una u otra de las respuestas ontológicas presentadas. Mi reflexión se inclina a favor de una posición filosófica como el pragmatismo, que daría cuenta del porqué el lenguaje y el uso del mismo llevan implícito esos compromisos ontológicos. Creo que es una posición que no contradice la respuesta fácil a las preguntas ontológicas, sino que trata de explicar filosóficamente por qué esta respuesta es así. Por eso, pragmática y pragmatismo podrían verse estrechamente vinculados.

## La “aproximación fácil” a la ontología de Amie L. Thomasson

En este apartado presentaré, sucintamente, algunas ideas de la perspectiva que Thomasson denomina “la aproximación fácil”<sup>1</sup> a la ontología, al tiempo que muchas otras serán obviadas por cuestión de espacio. Estas ideas proceden del libro que la autora publicó en 2015 *Ontology made easy*<sup>2</sup>.

El marco principal de la propuesta de Thomasson es la distinción que realizó Carnap en su artículo “Empirismo, semántica y ontología” (1950/1956)<sup>3</sup> entre cuestiones de existencia internas y externas a un marco lingüístico. Por ello, no es de extrañar que el rótulo escogido por Thomasson al comienzo del libro, en la Introducción, sea el siguiente: “La aproximación fácil olvidada”, aludiendo precisamente a la posición que Carnap mantuvo en el artículo arriba mencionado y que se centraba en la pregunta de si era posible dar una respuesta acorde con el empirismo, que siempre caracterizó a su filosofía, en relación a la cuestión de la existencia de entidades abstractas como los números, las propiedades o las proposiciones, entre otras, imprescindibles en la elaboración del conocimiento científico. La respuesta de Carnap a estas preguntas ontológicas se resolvió con la distinción entre cuestiones internas y cuestiones externas a un marco lingüístico, y es esta la línea que va a retomar Thomasson para contestar la pregunta ontológica general acerca de qué hay<sup>4</sup>. Una de las consecuencias que se siguen, entre otras, de esta ontología fácil es la de asumir una perspectiva deflacionaria

---

<sup>1</sup>A partir de ahora, y para simplificar la denominación, llamaré “ontología fácil” a la “aproximación fácil a la ontología” propuesta por Thomasson, que ella también emplea en alguna ocasión en su libro.

<sup>2</sup>Amie L. Thomasson, *Ontology made easy*, Oxford: Oxford University Press, 2015.

<sup>3</sup>Cfr. R. Carnap, “Empiricism, Semantics, and Ontology”, en R. Carnap, *Meaning and Necessity*, Chicago: The University of Chicago Press, 1956, 2<sup>nd</sup> ed., pp. 205–222.

<sup>4</sup>En la elaboración de su propio marco teórico, deudor de la ya indicada distinción carnapiana, Thomasson enfrenta su posición a los neo-quineanos. Para Thomasson este ha sido el modelo predominante en las últimas 60 décadas a la hora de tratar las cuestiones ontológicas, habiendo quedado la propuesta de Carnap prácticamente relegada al olvido.



con respecto a las discusiones y debates ontológicos y, también, la de comprometerse con un realismo simple<sup>5</sup> (o básico) con respecto a las entidades bajo discusión.

Una de las principales afirmaciones de la ontología fácil de Thomasson es que las preguntas ontológicas son fáciles de contestar haciendo uso de nuestra competencia lingüística y también apoyándonos en investigaciones empíricas. De tal forma que las preguntas ontológicas pueden contestarse empleando razones conceptuales o respuestas empíricas. Esta afirmación es deudora de la distinción carnapiana ya indicada en el siguiente sentido: para Carnap las cuestiones de existencia acerca de las entidades son preguntas internas a un marco lingüístico, ya que el marco lingüístico introduce un predicado general, por ejemplo, ser un número, y también nombres para las distintas entidades que tienen la propiedad general, estas últimas son los valores de las variables de las afirmaciones existenciales, en este caso, números como 1, 2, 3, etc.; y también expresiones para propiedades específicas de estas entidades, como “ser un número par” o “ser un número primo”, y también expresiones relacionales para indicar relaciones entre números como “ser divisible por”, “ser múltiplo de”; e, igualmente, expresiones para funciones, cuantificadores, etc. Y, obviamente, reglas tanto para definir los significados de los términos como para para formar oraciones correctas a partir de esas expresiones. Este último es un punto crucial para Thomasson, como se verá más adelante. Así pues, las preguntas internas son preguntas que se pueden contestar a partir del marco lingüístico, por ejemplo, ¿hay un número primo mayor que 100?, etc. Las preguntas externas son para Carnap pseudo-preguntas, pues no tienen respuesta a partir del propio marco lingüístico; son, por consiguiente, preguntas que no procede realizar a partir del marco lingüístico concreto elaborado; estas son preguntas relativas a la existencia del sistema de entidades como un todo, del estilo “¿existen los números?” y para Carnap no deben ser interpretadas como aserciones o creencias en la realidad de estas entidades, que se puedan valorar en términos de verdadero o falso, pues estas preguntas no son teóricas o cognitivas. A lo sumo, son preguntas externas al propio marco y, por ello, son preguntas de tipo práctico relativas a las razones que nos llevan a aceptar el marco lingüístico, a saber, su conveniencia, su eficacia instrumental, su éxito en relación con los objetivos, etc. Para Carnap la distinción entre cuestiones de existencia externas e internas a un marco lingüístico es equivalente respectivamente a la distinción que hay entre aceptar un sistema de entidades y realizar una aserción relativa al marco lingüístico en el que se habla de las entidades particulares del sistema.

Thomasson, siguiendo a Carnap e insistiendo en los aspectos vinculados al uso de las expresiones, afirma que las preguntas internas son fáciles de responder y que las preguntas externas son pseudopreguntas<sup>6</sup>, pero que una manera más clara de interpretar esta distinción carnapiana es, en su opinión, hacerla en términos de la clásica distinción uso/mención. Según esta interpretación, entonces, las preguntas de existencia internas son las que se preguntan haciendo *uso* del marco lingüístico, es decir, desde el interior del mismo, usando los términos a partir de las reglas de uso que el propio marco lingüístico introduce para ellos,

---

<sup>5</sup>“Easy ontology leads to realism about the questioned entities in affirming that there are properties, propositions, numbers, and so on, *in the only sense these terms have.*” A. L. Thomasson, op. cit., p. 153.

<sup>6</sup>Cfr. A.L. Thomasson, op. cit., p. 74.

mientras que, para Thomasson las cuestiones externas son relativas a la mención de los términos, en las que lo que resulta relevante son preguntas de tipo práctico, tales como si se deberían adoptar ciertos términos con sus respectivas reglas de uso o no<sup>7</sup>. Thomasson concreta esta importancia que ella concede al uso de los términos, vinculados a sus respectivas reglas de uso, en la formulación de las condiciones de aplicación<sup>8</sup>. Las condiciones de aplicación son “ciertas reglas de uso básicas que están entre aquellas que son constitutivas del significado [‘meaning- constituting’] para el término”<sup>9</sup>, son, pues, según Thomasson, reglas semánticas de uso que se adquieren cuando se aprende el lenguaje, equiparándolas a aquellas reglas que servirían para introducir el término, siendo semejantes a reglas gramaticales, y siendo también similares a ellas en el sentido de que es posible que un hablante competente no sea capaz de hacerlas explícitas. El punto esencial de las condiciones de aplicación —una vez más en consonancia con las preguntas de existencia internas de Carnap— es que ellas no precisan recurrir a la existencia previa de las entidades a las que los términos se refieren. De esta forma, las preguntas de existencia son para Thomasson fáciles de contestar simplemente haciendo uso de estas condiciones de aplicación, que reflejan la competencia conceptual de los hablantes y, en ocasiones, haciendo uso también de investigaciones empíricas.

Otro punto muy relevante y que, junto al anterior, va a jugar un papel fundamental en su defensa de la ontología fácil y que ya estaba presente en Carnap, es la respuesta a las preguntas de existencia interna de tipo general que se realizan haciendo uso de las reglas deductivas usuales, pues se puede proporcionar una respuesta fácil a una pregunta interna general de existencia realizando un proceso inferencial simple a partir de la verdad de aseveraciones más básicas<sup>10</sup>.

---

<sup>7</sup>Uno de los muchos párrafos a través de cuales Thomasson insiste en que las preguntas ontológicas son fáciles de responder porque solo se necesita usar los términos según las reglas de uso que los gobiernan es el siguiente: “But if we are using those terms [‘number’, ‘property’, ‘proposition’, ...] according to the rules of use by which they come to be introduced to the language, then those rules enable us to resolve the questions straightforwardly (through analytic or empirical means),...: the question is an internal question. So, if the external question is *not* supposed to be so straightforwardly answerable (so it is not an internal question), then it must be aiming to use the terms in question *without* their being governed by the standard rules of use.” (Ibid., p. 39).

<sup>8</sup>Estas condiciones de aplicación proporcionan un criterio formal para las preguntas sobre la existencia, que se presenta bajo la forma de un esquema general; así pues, el esquema de las condiciones de aplicación para un término general “K” es el siguiente: “E: Ks exist iff the application conditions actually associated with ‘K’ are fulfilled”. (Ibid., p. 86). Este esquema tiene para Thomasson la ventaja de que permite moverse en las dos direcciones del uso y la mención, a saber, desde la mención de los términos y la evaluación de si sus condiciones de aplicación se satisfacen, hasta el uso de los mismos al preguntarnos si las entidades por ellos referidas existen o no.

<sup>9</sup>Ibid., p. 89. Traducción mía.

<sup>10</sup>Esta idea es expresada con gran claridad en el siguiente texto: “So admitting variables that range over entities of a certain type commits you to employing the *concept* of number. Once we employ, that is, *use* that concept, however, its associated rules of use entitle us to make trivial inferences, for example, from ‘five is a number’ to ‘there is a number’, and thus to make inference to claims such as ‘there are numbers’. Given the rules of use for the concept *number*, once we admit the concept, we accept claims like ‘there are numbers’ or ‘numbers exist’. But in doing so we are not committing ourselves to a deep Platonistic metaphysics and cannot be accused of ‘illegitimate hypostatizations’, we are simply making the trivial, internal existence claim.” Ibid., p. 56.

## Pragmática y pragmatismo entrelazados

A partir de los aspectos que se destacaron de la ontología fácil propuesta por Thomasson, queda claro que las respuestas ontológicas son fáciles porque se basan en el uso que los hablantes hacen de los términos, junto con las pertinentes investigaciones empíricas. La dimensión más destacada por Thomasson en relación con las preguntas ontológicas internas es la semántica/pragmática, que subyace a las condiciones de aplicación de un término. En mi opinión, aunque esta pueda ser una respuesta apropiada, presenta algunas dificultades, una de ellas es claramente reconocida por Thomasson y consiste precisamente en determinar cuáles son las reglas de uso para los términos ordinarios. Pero al margen de estas dificultades, la propuesta de Thomasson, ligada al deflacionismo ontológico y a su posición de sospecha con relación a la ontología dura —aquella que establece debates sustantivos acerca de lo que hay—, parece no haber rozado más que la superficie del problema. Y esto se puede observar cuando afirma que no es necesaria teoría alguna que se ocupe de lo que es la verdad o la referencia, o que explique en consiste la existencia, ella prefiere evitar las controversias que, inevitablemente, se siguen cuando hablamos sobre conceptos, su existencia y su naturaleza, y por eso prefiere hablar de los términos y de sus condiciones de aplicación.

A mi modo de ver, el recurso al uso de las expresiones y a su dimensión pragmática para establecer los compromisos ontológicos o, en la terminología de Thomasson, apelar a las condiciones de aplicación de los términos para contestar fácilmente a las preguntas ontológicas, puede ser adecuado, pero filosóficamente interesa también una explicación de por qué esto es así, esto es, una explicación de por qué el uso de las expresiones lleva implícito ese compromiso ontológico, y para ello es importante explicar cómo el lenguaje y los sistemas de representación realizan esa función. Esta explicación, desde el momento que atañe al significado y uso de las expresiones, va a tener que lidiar con la dimensión semántica/pragmática de las mismas, introduciéndose en cuestiones ontológicas, ligadas a la referencia de los términos, a la noción de verdad y de realidad. Tanto la filosofía analítica como el pragmatismo han tratado de dar explicación a este hecho, aunque algunos de los más reconocidos representantes de la filosofía analítica lo hicieron postulando directamente entidades abstractas y comprometiéndose con un platonismo difícil de asumir, lo que daría la razón a Thomasson sobre la dificultad de llegar a una respuesta convincente y fácilmente asumible en relación con los debates sustantivos que propone la ontología dura, centrada en criterios de existencia.

En general, el pragmatismo mantiene que no podemos decir que es lo que hay independientemente de la mediación que un sistema de representación signico, marco lingüístico o lenguaje establece, donde no solo el lenguaje sino también el mundo y la mente (o los individuos que hablan un lenguaje) juegan un papel. En mi opinión, este punto de vista pragmatista, breve y vagamente aquí presentado, podría asumirse que está implícitamente aludido en “las condiciones de aplicación de un término”, el núcleo teórico explicativo de la aproximación fácil a la ontología de Thomasson, pues son esas condiciones de aplicación de un términos las que permiten una respuesta fácil a las preguntas ontológicas. El pragmatismo, más que la filosofía analítica, que también trató estos temas, pre-

senta una solución más acorde con lo que mantiene Thomasson cuando afirma que al usar el término “objeto”, según las reglas de un determinado lenguaje, no estamos queriendo decir que los objetos existentes dependan de ese lenguaje. Esta explicación tan básica y poco elaborada, con la que Thomasson zanja la cuestión sustantiva de que se entiende por objeto, es la que el pragmatismo elabora con mayor profundidad, en concreto, esto puede observarse en la propuesta de Peirce con su recurso a las categorías cenopitagóricas y a la semiótica o teoría de los signos. Peirce, partiendo de la semiótica y del estudio de la semiosis, esto es, el proceso en el que algo funciona como signo, explicó filosóficamente la noción de realidad, apoyándose en la mediación que los sistemas de signos, esto es, los lenguajes realizan sobre lo dado, lo que le llevó a afirmar que el mayor grado de realidad se obtiene a partir de la noción de signo<sup>11</sup>. Y también, habría que incluir las contribuciones del último Putnam<sup>12</sup>, en su defensa del realismo pragmático, al que Thomasson acusa de antirrealista, pero que ya resulta imposible desarrollar aquí.

---

<sup>11</sup>Cfr., *Semiotic and Significs: The Correspondence between Charles S. Peirce and Victoria Lady Welby*, C. Hardwick (ed.), Bloomington: Indiana University Press; 2<sup>a</sup> edición, Elmhurst, IL: The Press of Arisbe Associates, 2001, carta de Peirce a Victoria Welby del 12/10/1904.

<sup>12</sup>Cfr., para una discusión más detallada en favor del realismo pragmático de Putnam, que no cometería el antirrealismo al que se refiere Thomasson, F. González García & M.U. Rivas Monroy, “The Pragmatic Realism of Hilary Putnam”, *Following Putnam’s Trail: On Realism and Other Issues*, Rodopi: Amsterdam, pp. 223–242, 2008.

# The dynamics of value ascription

Andrés Soria Ruiz

Evaluative expressions are expressions of natural language that figure paradigmatically in *ascriptions of value*, that is, utterances of sentences containing such expressions whereby speakers communicate a positive or negative appreciation. However, sentences in which evaluative expressions appear are not always used to make ascriptions of value. They can also be used to make *judgments of fact*, that is, statements whereby only descriptive information is transmitted (Hare 1952, Umbach 2016). Our purpose is threefold: first, we aim to establish the aforementioned contrast between evaluative and descriptive uses of evaluative expressions. Secondly, a dynamic model that distinguishes these two uses is proposed. Finally, a definition of value ascription based on this model is offered.

Consider the contrast between the following contexts in which ‘*good sports car*’ is used:

- (1) a. (Context: Samir and Nora do not share a standard for a good sports car; they are both in front of a white Mercedes and have studied all its specs carefully):  
Samir: ‘This is a good sports car’
- b. (Context: Samir and Nora share a strict standard for good sports cars: they have to be white Mercedes of such-and-such specs. Samir sees a car and calls Nora on the phone; Nora hasn’t seen the car):  
Samir: ‘This is a good sports car’

We hold that, in context (1a), Samir is making a **pure ascription of value**, while in (1b), he is making a **pure judgment of fact**. The contrast might not be immediately apparent, but it can be further assessed by noting the following: (a) Samir’s claim in (1b), but not in (1a), can be resisted by saying things like ‘*that can’t be*’, ‘*how do you know?*’; (b) if Nora in (1b) goes on to discover that the car lacks the properties that satisfy their common standard—for instance if she discovers that the car isn’t white—she can accuse Samir of *lying* to her, although that accusation could hardly make sense in (1a) (Hare 1952, p.113); (c) if Nora does reject Samir’s claim, the disagreement between them would look *faultless* in (1a) (Kölbel 2003). Not so in (1b), where such a disagreement would seem to turn either on the car’s specs or on aspects of their (presumably shared) standard, about which either ought to be mistaken. And finally (d), Nora’s uptake of Samir’s claim in (1a) would alter her *practical attitudes* towards sports cars (as she adopts Samir’s standards), but not so in (1b) (Williams 1981 *a.m.o.*).

How can we model this contrast? The view advocated here allows for one and the same utterance containing a value expression to have a *different communicative profile* depending on features of the context in which it is made. This view takes inspiration from metalinguistic (stemming from Barker 2002; see Sundell & Plunkett 2013, Umbach 2016 as well) and expressivist approaches to normative

language (Gibbard 2003, Yalcin 2012). We proceed in four steps: first, enter *hyperplans* (Gibbard 2003). Just as possible worlds are maximally decided states of affairs, hyperplans are *maximally decided states of planhood*; a plan that tells you what to do in every conceivable situation. Secondly, define a relation of support [rejection] of an object of evaluation  $\alpha$  by a hyperplan  $h$  in the following way:

- (2) (*Support [rejection]*): A hyperplan  $h$  supports [rejects]  $\alpha$  iff the agent of  $h$  orients their action towards [against]  $\alpha$  in all relevant situations.

Thirdly, enrich the dynamic view of context (Stalnaker 1978/2002, Heim 1982/2002) with a hyperplan parameter (Yalcin 2012). Contexts are now  $n$ -tuples of parameters containing at least a set of possible worlds  $W$  and a set of possible hyperplans  $H$ , representing interlocutors' *shared plan of action* (Portner 2004, 2012 for related proposals). Finally, adopt the following dynamic entry for *good*:

- (3)  $\llbracket \text{good} \rrbracket^c = \lambda \alpha \lambda c. \forall w \in W_c, \alpha \text{ meets a standard } s_c \text{ in } w \ \& \ \forall h \in H_c, h \text{ supports } \alpha$

(3) says that, for an object  $\alpha$  to be called *good* is for the context to be updated in a way such that, at every world in the context, the properties of  $\alpha$  satisfy a certain, contextually determined standard; and that every hyperplan in the context supports  $\alpha$  (in virtue of satisfying such standard).

This allows us to model evaluative and descriptive uses of *good* as different types of update: a pure evaluative update—(1a)—requires a context where factual information about the object under evaluation is presupposed (a context whose world parameter entails all relevant information about the object under evaluation):

- (4) (*Pure evaluative update*): For all contexts  $c'$  s.t.  $\forall w. \alpha$  meets  $s_{c'}$  in  $w \rightarrow w \in W_{c'}$ ,  
 $c' + \llbracket \alpha \text{ is good} \rrbracket = \langle W_{c'}, \{h \in H_{c'} : h \text{ supports } \alpha\} \rangle$

By contrast, a pure descriptive update—(1b)—requires a context where practical information is antecedently shared (a context whose hyperplan parameter entails support for objects like the one under evaluation), so that the use of *good* ends up operating on the possible world parameter of the context:

- (5) (*Pure descriptive update*): For all contexts  $c'$  s.t.  $\forall h. h$  supports  $\alpha \rightarrow h \in H_{c'}$ ,  
 $c' + \llbracket \alpha \text{ is good} \rrbracket = \langle \{w \in W_{c'} : \alpha \text{ meets } s_{c'} \text{ in } w\}, H_{c'} \rangle$

As will be argued, this proposal offers a neat explanation of the particular features of evaluative uses observed at the outset.

Barker, C. (2002). The dynamics of vagueness • Gibbard, A. (2003). Thinking how to live • Hare, R. M. (1952). The Language of Morals • Heim, I. (2002/1983). On the Projection Problem for Presuppositions • Kolbel, M. (2003). Faultless disagreement • Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste • MacFarlane, J. (2014). Assessment sensitivity • Plunkett, D. and T. Sundell (2013). Disagreement and the semantics of normative and evaluative terms • Portner, P. (2004). The semantics of imperatives within a theory of clause types • Portner, P. and A. Rubinstein (2012). Mood and contextual commitment • Silk, A. (2017). Evaluational Adjectives • Stalnaker, R. (2002/1978). Assertion • Umbach, C. (2016). Evaluative propositions and subjective judgments • Yalcin, S. (2012). Bayesian Expressivism

# El sentido común y la filosofía del lenguaje informal

Javier Vilanova Arias<sup>1</sup>

UCM

vilanova@filos.ucm.es

Desde sus inicios, la Filosofía se ha desmarcado del sentido común. La razón es clara. En buena medida la filosofía (junto con todas esas otras cosas que hoy en día llamamos ciencias) nace gracias a un “principio de sospecha”, según el cual las apariencias son solo eso, apariencias, que más bien ocultan que revelan la verdadera naturaleza de las cosas. Pero la opinión común se basa en experiencia, y la experiencia es solo experiencia de las apariencias, por lo que recurrir a la opinión compartida es confundir el envoltorio con el regalo.

Esto se puede ver nítidamente en el Fedro de Platón. La doxa, al fundarse en la verosimilitud en vez de la verdad, no puede producir verdadero conocimiento sino, al contrario, sirve tanto para defender una cosa como su contraria (el filósofo será el que sabe ir más allá de lo que se ve y se toca y alcanzar el ámbito de las esencias necesarias, donde mora la certeza). Otro buen ejemplo nos lo da Descartes. Al no estar sujeto al control de la crítica racional, al no ser el resultado de la aplicación de un “método”, para el francés el sentido común resulta inútil para producir genuino conocimiento, o cuando menos para producir conocimiento “de calidad”. En sus propias palabras, es pura *cognitio*, una apercepción animal de lo que pasa y no *scientia*, conocimiento reflexivo y bien fundado. En el el siglo XX se mantiene este desprecio por el sentido común, que se asocia al prejuicio y la opinión irreflexiva (en algunos casos como el de Sellars incluso se ve como algo que debe ser directamente reemplazado por el conocimiento científico).

Una excepción bien conocida a este general menosprecio por el sentido común lo constituye la llamada “Escuela de Sentido Común” (a veces llamada “escuela escocesa de sentido común” debido al hecho de que la mayoría de sus integrantes provienen de ese país). Ya en 1724 Claude Buffier, un monje jesuita, en su *Traité des premières vérités* ya replica Descartes señalando que *es el sentido común, y no los logros de la filosofía escolástica, el que provee la regla principal para juzgar cuestiones fundamentales*<sup>2</sup>. E inmediatamente después de la contundente ofensiva de Hume, Thomas Reid contra-ataca alegando que incluso la lógica o las más metodologías más científicas o más excelsas han sido producidas usando nuestras facultades naturales, por lo que al ir contra el sentido común el filósofo cae en un absurdo. Como alega en las conclusiones de su primera obra importante, *Inquiry into the Human Mind: Nos sirven directamente en los asuntos*

---

<sup>1</sup>Este trabajo ha sido llevado a cabo al amparo del proyecto de investigación *Cognitive Vulnerability, Verisimilitude and Truth* (FFI2017-84826-P) así como de los grupos de investigación complutense 930664 y 930370.

<sup>2</sup>P. Buffier 1843, I.v.33, 15.

*cotidianos de la vida, en los que nuestra facultad de razonamiento nos dejaría en la oscuridad. Son parte de nuestra constitución, y todos los descubrimientos de la razón se basan en ellas. Constituyen lo que se llama “el sentido común de la humanidad”, y todo lo que claramente contrario a ellas es lo que llamamos “absurdo”. Su fuerza es el “buen sentido”, que a menudo nos encontramos en personas que no son muy inteligentes. Una desviación notoria de ellas, surgida de un desorden en la constitución de la persona, es lo que llamamos “locura”, como cuando un hombre cree que está hecho de cristal. Cuando un hombre se permite ser sacado de los principios del sentido común por argumentos metafísicos, podemos llamar a esto “locura metafísica”, que se diferencia de otro tipo de locuras en ser intermitente en vez de continua: es apta para el paciente en momentos solitarios y especulativos; pero cuando el vuelve a la compañía de otros el sentido común recobra su autoridad<sup>3</sup>.*

Esta corriente filosófica más bien minoritaria y con escaso protagonismo, que es la que suele identificarse como “Escuela del Sentido Común”, y que con altibajos se mantiene hasta nuestros días en la obra de autores no demasiado conocidos como James Beattie, James Oswald, Dugald Stewart o Sir William Hamilton y, más recientemente, Noah Lemos, Stephen Boulter o Lynd Ferguson<sup>4</sup>.

En mi opinión, la razón por la que la escuela del sentido común ha fracasado en su empeño por reintroducir el sentido común en la reflexión filosófica es la noción de sentido común con la que trabajan. No es una buena idea, y esto a su vez provoca que estos filósofos tampoco hayan entendido bien, en mi opinión, la manera correcta de usar el sentido común en la práctica filosófica.

En efecto, acorralados por las demandas de certeza de Descartes, Kant y los que vendrán después, los filósofos de esta escuela defienden el sentido común como una fuente de verdades universales y necesarias. Terminan así hablando de él como algo parecido a un oráculo que uno puede consultar por el simple expediente de cerrar los ojos y concentrarse en la cuestión, lo que a su vez hace que el sentido común devenga en algo trascendente (o, si se entiende mejor así, previo) a nuestra experiencia y a nuestras humanas prácticas. De ahí que en el trasfondo funcione siempre una teoría innatista del origen del sentido común:

Aquí, por lo tanto, entiendo por sentido común aquella disposición que la naturaleza ha colocado en toda la humanidad, o manifestamente en la mayoría de ella, en orden a llevarles, una vez que han alcanzado el uso de la razón, a un juicio común y uniforme.<sup>5</sup>

“La naturaleza ha colocado en toda la humanidad” suena aquí demasiado a “infuso en el alma humana por la divina providencia” (y de hecho Reid dice abiertamente, justo antes del párrafo antes reproducido, que *they are put into our minds by God*). Esta coartada teológica, que en la época de Reid todavía conservaba su grano de pertinencia en el debate filosófico, resulta impropio en el panorama laico y completamente aseglarizado de hoy en día. Lo que no impide que, de otra manera, no encontremos esta versión del sentido común como “un don” con un nuevo ropaje. En efecto, la gran mayoría de los filósofos actuales que favorecen al sentido común optan por una explicación de corte

---

<sup>3</sup>T. Reid 1764, cap. 7).

<sup>4</sup>Véanse S. Boulter 2007, L. Ferguson 2007 y N. Lemos 2004.

<sup>5</sup>T. Reid, *Ibidem*, cap. 1.



científico que sustituye a “Dios” por la “Evolución”. El “argumento evolucionista” que encontramos en autores contemporáneos más o menos simpatizantes con la escuela del sentido común como Lynd Forgusson o Stephen Boulter así como entre los psicólogos y lógicos que defienden los modos de razonamiento comunes en las *rationality wars*. Aunque muchos de sus proponentes siguen hablando de “creencias” y adscribiendo algún papel a la experiencia en su actualización, la tendencia general es considerar al sentido común como una facultad innata en el ser humano que la naturaleza, ya que no Dios, ha puesto en nuestros genes, ya que no en nuestra alma, al modo que ha puesto la capacidad de orientarse mediante ondas sonoras al murciélago o la de producir telarañas a algunos artrópodos. En mi opinión, este *Deus sive Natura* es incapaz tanto de darnos una idea empíricamente adecuada del sentido común como una justificación de su pertinencia para la práctica filosófica.

Aunque sé que voy a simplificar un poco demasiado, resumamos esta línea de pensamiento así:

- las creencias de sentido común son de origen natural (estamos biológicamente predeterminados a tener esas creencias),
- lo que creemos por naturaleza no podemos dudarlo y por lo tanto es aquello de lo que tenemos más certeza,
- en consecuencia, si el filósofo desea alcanzar conclusiones ciertas, debe partir de las premisas del sentido común y, jamás, pretender que cree seriamente algo que va contra ellas.

El primer error proviene del hecho de que la propia concepción del sentido común con que se trabaja ya hace muy problemática la idea de una “filosofía del sentido común”. Si las cosas del sentido común son sabidas e indudables para todo el mundo, ¿para qué necesitamos pensar más sobre ellas? Parece que al filósofo cabal solo le quedara la tarea de “pasar a limpio” los apuntes que la naturaleza ha inscrito en nuestra cabeza, y ocasionalmente señalar con el dedo al filósofo que en sus especulaciones osara salirse de tiesto. La certeza deja, qué duda cabe, poco campo de maniobras. Pero el error de bulto, en mi modesta opinión, está en esa confianza en la inquebrantable compulsión de lo que en nosotros hay de innato. No se trata de negar que nazcamos con alguna predeterminación biológica a creer y hacer ciertas cosas. Y desde un punto de vista personal el punto de partida del sentido común puede ser genético, pues recién nacidos nuestra dotación biológica es lo único que traemos en la mochila. Pero por mucho que nazcamos equipados con un sentido “natural”, alguna adaptación cognitiva o mental al medio en el que vamos a vivir y otros maravillosos equipamientos de serie, después vienen la educación, los padres, la escuela, las malas influencias, los desengaños amorosos y tantas otras peripecias que van reconfigurando nuestra forma de pensar y actuar. Y en el caso del sentido común ocurre exactamente lo mismo. Hay algo que ya compartimos todos los hombres biológicamente, eso que Aristóteles llamaba *koiné aithésis* y que corresponde investigar a las ciencias. Y es obvio que eso que ya compartimos “como especie” es algo que contribuye a configurar y posibilita que podamos compartir algunas de las cosas que caen dentro de lo que denominamos sentido común. Pero no es menos obvio que hay otro tanto que compartimos culturalmente, que no viene en los genes y que se trasmite de generación en generación exo-somáticamente, es decir, por fuera del

cuerpo, a través de la educación y el lenguaje. En el caso del sentido común, además, hay importantes variaciones históricas y geográficas (es obvio que “lo que es de sentido común” cambia en el espacio y el tiempo), con lo cual, sin dejar de reconocer el factor genético, dado que la cultura tiene la última palabra (por la simple razón de que viene después), lo menos confundente es decir que el sentido común es cultural.

Necesitamos, pues, un concepto distinto de sentido común y al mismo tiempo una manera distinta de entender su papel en la investigación filosóficos. Si examinamos la manera en que apelamos al sentido común en la vida cotidiana (lo que podríamos llamar la concepción de sentido común del sentido común), veremos caracterizado el sentido común como una propiedad de algunos hechos, dichos o pensamientos humanos que se distinguen por (i) su obviedad, (ii) su carácter implícito, (iii) su familiaridad, (iv) su no necesidad de prueba y (v) su orientación práctica. Lo más característico del conocimiento de sentido común es, pues, su generalidad. Es “común” o general en todos los sentidos. El sentido común es el conocimiento común sobre los hechos comunes en torno a las cuestiones comunes. O lo que es lo mismo, es un conocimiento compartido por la gran mayoría de las personas sobre los hechos más generales en torno a las cuestiones más básicas. Casi podría definirse así: conocimiento común sobre los hechos comunes en torno a las cuestiones comunes<sup>6</sup>.

Ahora bien, ¿cuál es el depositario del sentido común? ¿Dónde lo encontramos instanciado? Ya que no en nuestra alma, en nuestra mente o nuestros genes, como he argumentado, es necesario buscar en otro lugar. Un simple examen de los lugares en que utilizamos la expresión “de sentido común” (que omito aquí por falta de espacio) basta para ver que, primariamente, lo adscribimos a HECHOS y DICHOS humanos (lo que hizo Pepito es de sentido común, lo que dijo Juanito es de sentido común). Acciones y actos de habla: ¿hay un *genus* común a ambas cosas? Para responder a la pregunta daré un pequeño rodeo.

Se le ha dado distintos nombres, “Filosofía del Lenguaje Informal”, “Pragmatismo Lingüístico”, “Fenomenología Lingüística”, pero el que se ha hecho más popular es “Filosofía del Lenguaje Ordinario”. Se coincide en cifrar los poco más de 100 kilómetros que separan la universidad de Cambridge de la de Oxford como su cuna y las décadas de los treinta y cuarenta como las de su lenta germinación. Lenta porque los que se consideran sus pioneros, Ludwig Wittgenstein y John Austin, eran remisos a publicar sus pensamientos prefiriendo, de acuerdo a su propia concepción de la filosofía y el lenguaje, difundir su mensaje de viva voz. Como los filósofos analíticos que les precedieron, creían que la mejor manera de resolver problemas filosóficos era examinando el propio lenguaje en el que se planteaban. Lo que los distinguía radicalmente de ellos era la concepción del lenguaje. Si los filósofos anteriores consideraban las palabras como un medio para describir la realidad, Austin y Wittgenstein la ven como instrumentos para hacer todo tipo de cosas. *Cómo hacer cosas con palabras* es el revelador título del libro en el que Austin investiga todas las formas posibles en las que podemos hacer cosas con el lenguaje, desde prometer, pedir y elogiar hasta divorciar, promulgar o devaluar. A Wittgenstein, por su parte, le gustaba comparar las

---

<sup>6</sup>Análisis de los rasgos del sentido común más o menos cercanos al que efectúo yo aquí pueden encontrarse en el muy recomendable libro de N. Rescher 2005, y en los artículos respectivos de H. Parrett 1987 y A. Stroll 1987.

palabras con herramientas. Una palabra es como un martillo, cuyo sentido radica en su función y su conocimiento en su modo de uso, y como el martillo puede ser usado para muchas cosas y de modos muy diferentes dependiendo de las circunstancias.

Una consecuencia que se saca inmediatamente de esta concepción pragmática del lenguaje es que sus reglas y conceptos son, según este enfoque, irresolublemente vagos (aunque quizás sea más justo decir que son flexibles). Como las herramientas, se adaptan a las circunstancias y necesidades del momento, y por lo tanto pueden ser usadas de manera muy distintas. Por esa misma razón para Austin y Wittgenstein las auténticas reglas de una práctica lingüística nunca están “escritas”. Aunque los diccionarios y las gramáticas pueden ser útiles para aprender una lengua o para uniformizar las variantes dialectales, como todo aprendiz de una lengua extranjera sabe no te sirven para aprender a hablar como los nativos. Para ello debes ver como ellos “usan” las palabras, debes ver las palabras en acción en cuantas y más variadas situaciones mejor. Es el uso el que determina la regla y no al revés. Así que el verdadero sentido de las palabras y las oraciones, piensan estos filósofos, está en el lugar donde las palabras son usadas, en su hábitat natural, es decir, en la vida corriente y moliente de la gente corriente moliente que tantas veces resuelve sus problemas “hablando”.

Comparemos. Lenguaje ordinario: práctico, vago, cotidiano, mayoritario... Sentido común: práctico, vago, familiar, compartido por todos...

En última instancia, si se entiende bien lo que es el lenguaje, como un conjunto de herramientas que usamos para los más variados fines, o como una colección de convenciones que nos permiten hacer cosas con palabras, una buena manera de captar la dimensión social del sentido común es remarcando su naturaleza simbólica. Una vez se identifica el lenguaje con un fenómeno histórico, inseparable de la comunidad que lo alberga y fundamentalmente pragmático, la mejor manera de caracterizar la naturaleza del sentido común es aplicarle la etiqueta de “lingüístico”.

## Bibliografía

- Austin, John L. (1961). *Philosophical Papers*. Oxford: Oxford U.P.
- Boulter, Stephen (2007). *The Rediscovery of Common Sense Philosophy*. New York: Palgrave.
- Buffier (1843). “Traité des premières vérités et de la source de nos jugements”. In: *Oeuvres philosophiques du Père Buffier*, Paris: Adolphe de la Haye, 1843.
- Cavell, Stanley (1969). *Must we mean what we say?* New York: Charles Scribner’s Sons.
- Forguson, Lynd (1989). *Common Sense*. Routledge, London.
- Holthoon, Frits van and Olson, David R. (1987) (eds). *Common Sense. The Foundations for Social Science* University Press of America.
- Lemos, Noah (2004). *Common Sense*. Cambridge U.P. New York.

- Lycan, William G. (2007). "Moore's Common Sense". In: Susana Nuccetelly y Gary Seay (eds), (2007). *Themes from G. E. Moore*, Oxford U.P., Oxford.
- Moore, George E. (1922). *Philosophical Studies*. London: Routledge & Kegan Paul.
- Parrett, Herman (1987). "Common Sense: From Certainty to Happiness". In: Holthoon and Olson (eds.) p. 17–31.
- Rescher, Nicholas (2005). *Common Sense. A New Look at an Old Tradition*. Milwaukee: Marquee U.P.
- Reid, Thomas (1983). *Inquiry and Essays*, ed. by Beanblossom, R. and Lehrer, K. Indianapolis, Hackett.
- Stroll, Avrum (1987). "Foundationalism and Common Sense". In: Holthoon and Olson (eds.) 1987, 35–54.
- Wittgenstein, Ludwig (1953). *Philosophical Investigations*.
- Wittgenstein, Ludwig (1953). *On Certainty*. Oxford: Basil Blackwell.

# FILOSOFÍA DE LA MENTE Y EPISTEMOLOGÍA

# The extension of cognitive abilities

**Gloria Andrada de Gregorio**  
UAM

Is our cognitive character extended to the artifacts we interact with? To answer this question we must look at a current trend in epistemology concerning the epistemological implications of extended cognition (Clark and Chalmers 1998, Clark 2008), and particularly the relation between extended cognition and virtue epistemology.

Various arguments have been given for thinking that knowledge is sometimes the product of cognitive virtues or abilities that are based on extended cognitive processes, that is cognitive processes whose material realizers are scattered across brain body and world (Pritchard 2010). But does the extension of abilities entail the extension of an agent's cognitive character? Palermos (2014) argues that it does. According to his view, not only cognition is extended but so is an agent's cognitive character: when properly coupled with an organism, external resources such as technological artifacts, become integrated into an agent's cognitive character, thereby extending it. Menary (2012), on the other hand, rejects the artifact-extension view of cognitive character. He argues that cognitive character is extended by the integration of cognitive practices (a subset of cultural practices) which transform her cognitive abilities by extending what she is able to do, but not by coupling with an external resource. This means that system outwits character.

In this paper I argue that character is extended by a process of enculturation, not by coupling with external resources. However, some elements in our epistemic environment are constitutive of our cognitive abilities, and this is something that is shared among some accounts of virtue epistemology (Sosa 2010) and extended cognition. The view I shall defend occupies a middle ground between those put forward by Palermos (2014) and Menary (2012), which in my view are too strong and too weak, correspondingly.

My point is twofold. First of all, I argue that external resources are constitutive of our cognitive abilities, but my argument does not rest on extended cognition but on the structural properties of abilities as dispositions. Second of all, I defend that character is extended by acquiring new cultural practices that transform what an agent is able to do. Once this is accepted, an interesting implication is revealed: we must appreciate the collective aspect of our cognitive character, which pushes virtue epistemology out of its traditional individualistic framework.

My argument goes as follows:

Abilities are dispositions to perform well (Sosa 2010) and as such they are structured in the way of dispositions generally. This entails that they are triggered in a particular environment. In the environment we can locate several resources

that constitute an agent's full competence. Consequently, external resources located in an agent's epistemic niche are in this sense constitutive of her cognitive abilities. But ability (full competence) outwits cognitive character. This leads to my second point. Building on the view developed by Menary (2012), I defend that the extension of cognitive characters does not involve causal couplings with artifacts, but integration of cognitive practices that transform what an agent is able to do. This includes transformation of her cognitive capacities by patterned practices (Roepstorff et al. 2010). This view of extended cognitive characters allow us to distinguish genuine extension from cognitive off-loading or outsourcing, i.e.: extension versus structural enablers.

In this regard, an extended cognitive character includes an agent's cognitive traits, her epistemic identity (i.e: whether she takes good care of her epistemic artifacts, whether she is meticulous in her inquiries, etc), and the cognitive practices she acquires, which include cultural norms. It does not include the external elements themselves, given that when decoupled from the external resource she retains part of her competence, to the extent that when in a situation such that the external resource is present (and she is in good shape) she can perform well. This differentiates my account of extended character from different possibilities such as automatic integration of technology (Carter et al 2018), cybernetic implants, pharmacological enhancement or genetic engineering. It is a phenomenon that can be instantiated in real-life scenarios and that requires learning and training.

Interestingly, there is something novel that extended cognition brings to the fore and that has not received proper attention: the fact that cognitive characters are extended and transformed by the acquisition of cultural practices entails that cognitive traits (including some virtues and vices) do not rest solely on the individual but on a wider epistemic community where an agent is enculturated. This is something that calls for a change in the way we have traditionally theorized about the nature of character. Also, extended cognitive characters—if properly integrated usually by a process of learning and training—remain constant across a variety of situations, thus retaining its stability as well as other attributes.

## References

- Alfano, M., Skorburg (forthcoming), "The Embedded and Extended Character Hypotheses", to appear in Adam J. Carter, Andy Clark, Jesper Kallestrup, S. Orestis Palermos, and Duncan Pritchard (eds.). *Extended Epistemology*. Oxford: Oxford University Press.
- Clark, A. and Chalmers, D. J. (1998). The extended mind. *Analysis*, 58(1):7–19.
- Carter, Clark, Palermos (2018). *New Humans, Ethics, Trust, Extended Epistemology*. Oxford University Press.
- Brogaard, B. (2014). A Partial Defense of Extended Knowledge. *Philosophical Issues*, 24:39–62.

- Doris, J. M. (2002). *Lack of Character: Personality and Moral Behavior*. New York: Cambridge University Press.
- Harman, G. (1999). Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error. *Proceedings of the Aristotelian Society*, 99:315–331.
- Harman, G. (2000). The Nonexistence of Character Traits. *Proceedings of the Aristotelian Society*, 100:223–226.
- Menary, R. (2012). Cognitive practices and cognitive character. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 15(2):147–164.
- Palermos, S. O. (2014). Knowledge and cognitive integration. *Synthese*, 191(8):1931–1951. doi:10.1007/s11229-013-0383-0.
- Pritchard, D. (2010). Cognitive Ability and the Extended Cognition Thesis. *Synthese*, 175:133–151.
- Roepstorff, A., J. Niewo hner, and S. Beck. (2010). Enculturating brains through patterned practices. *Neural Networks*, 23(8-9):1051–1059.
- Sosa, E. (2010). How Competence matters in epistemology. *Philosophical Perspectives*, 24, Epistemology.



# El fiabilismo y la justificación de las creencias colectivas

Moisés Barba Magdalena

## Resumen

¿Bajo qué condiciones está epistémicamente justificada una creencia colectiva? Goldman ha propuesto aplicar el fiabilismo acerca de la justificación a esta cuestión. Según su propuesta, una creencia colectiva está justificada sólo si es el resultado de un proceso que tiende a llevar al colectivo a formar creencias colectivas verdaderas. El propósito de esta comunicación es argumentar que la propuesta de Goldman no es satisfactoria, porque la fiabilidad del proceso de formación de las creencias colectivas no es condición suficiente para que las creencias colectivas estén justificadas. Para desarrollar esta tesis, tomaré como punto de apoyo teórico la teoría de Gilbert acerca de las creencias colectivas como compromisos conjuntos de cierto tipo, y desarrollaré dos argumentos: primero, que la fiabilidad del proceso de formación de las creencias de un colectivo es compatible con que su implementación se deba a la suerte. Segundo, propondré un principio, que llamo *Accountability Colectivo-Miembro*, como condición necesaria de la justificación de las creencias colectivas, y mostraré que hay casos en los que esta condición no se da, pero sí la fiabilidad del proceso de formación de las creencias del colectivo.

## Las creencias colectivas

De forma habitual, atribuimos creencias a los colectivos. Decimos que un partido político cree que cierta medida es la mejor, que un jurado cree que el acusado es inocente, o que un equipo de investigación cree haber encontrado la explicación para un fenómeno. Se ha debatido ampliamente acerca de la naturaleza de las creencias colectivas; yo asumiré, en este texto, la conocida como teoría del compromiso conjunto, elaborada por M. Gilbert (Gilbert, 2014). Esta explica la normatividad de la acción colectiva. Actividades como pasear con otra persona o conversar son paradigmas de este tipo de acciones. En ellas hay una normatividad involucrada, de modo que si, por ejemplo, uno abandona el paseo o la conversación sin más explicación, el otro tendrá razones para protestar. Según Gilbert, esta normatividad radica en que, en la acción colectiva, cada participante asume ante los otros ciertos compromisos. Estos llevan a la formación de un “sujeto plural”, esto es, al reparto del trabajo entre los participantes de modo que se persiga conjuntamente un fin como si se tratara de un solo sujeto. Cuando abandonamos sin más una acción de este tipo, estamos faltando a nuestro compromiso de contribuir a dar vida al sujeto plural, y por ello el otro puede

recriminárnoslo. Según Gilbert, las creencias colectivas pueden analizarse según este modelo como un tipo de acción colectiva. Así, que un colectivo crea que *p* consiste en que sus miembros asumen el compromiso conjunto de contribuir mediante sus acciones a que el colectivo se comporte, en conjunto, como un individuo que creyera que *p*.

Podemos preguntarnos por la justificación epistémica de las creencias colectivas. Recuperando los ejemplos anteriores, podríamos decir del partido político, del jurado y del equipo de investigación que están o no justificados en sus creencias, en función de si, por ejemplo han procesado la información disponible de manera correcta. ¿Cómo entender más exactamente la justificación de las creencias colectivas?

## La propuesta de Goldman

Goldman ha propuesto abordar este problema desde el fiabilismo (Goldman, 2014). Es preciso señalar que Goldman no se comprometió, en este artículo, con ninguna formulación particular de las creencias colectivas, sino que se limitó a seguir a List y Pettit en que las actitudes proposicionales colectivas están fundamentadas en, sin reducirse a, las de los miembros del colectivo (List y Pettit, 2011). Ahora, esto es compatible con la teoría del compromiso conjunto. Desde dicha teoría, las actitudes proposicionales de los colectivos son analizadas como dependientes de las de sus miembros. Así, el marco elegido para entender las creencias colectivas no plantea tensiones iniciales con la propuesta de Goldman, de modo que su concepción de la justificación colectiva debería ser aplicable a las creencias colectivas entendidas como compromisos conjuntos de cierto tipo.

Goldman propone traducir la teoría de la justificación epistémica del nivel individual al colectivo. Puesto que ya disponemos de sólidas teorías acerca del concepto general de la justificación epistémica al nivel individual, bastaría con hallar el modo correcto de traducirlo al nivel colectivo. La teoría de la justificación defendida por Goldman es el fiabilismo. A grandes rasgos, que una creencia esté justificada consiste, según esta teoría, en ser el resultado de un proceso que tiende a producir creencias verdaderas (Goldman, 1979). Traducida al nivel colectivo, esta teoría arrojaría que las creencias colectivas están justificadas si son el resultado de un proceso por el cual el colectivo en cuestión tiende a formar creencias colectivas verdaderas. Mi propósito, en este texto, es argumentar que, para que las creencias colectivas estén justificadas, no es suficiente con que sean producidas de modo fiable.

Hay una cuestión adicional que debo explicar antes de proseguir. En su artículo, Goldman trata la fiabilidad como una condición *necesaria* de la justificación de las creencias colectivas. Mi objetivo, sin embargo, es mostrar que no se trata de una condición *suficiente*. ¿No estaría faltando a la propuesta de Goldman, tratando de refutar algo que él no ha afirmado? Creo que no es así. La razón es que Goldman trata la fiabilidad sólo como una condición necesaria porque asume que cualquier teoría de la justificación epistémica debe incluir condiciones para acomodar los *defeaters*; no obstante, sí parece asumir que todo cuanto es necesario, respecto de la producción de las creencias colectivas, para que estas

puedan considerarse justificadas es que sean producidas de forma fiable. Esto es lo que me propongo mostrar como falso.

Me serviré, para mis fines, de dos argumentos. El primero consiste en presentar un caso en el que un colectivo forma sus creencias a través de un proceso fiable, pero que es usado por suerte. Asumiendo la tesis de que la suerte excluye la justificación (Pritchard, 2005), así como una concepción particular de la misma, en nuestro caso habrá, al mismo tiempo, suerte y fiabilidad, por donde se concluye que la fiabilidad no es condición suficiente para la justificación colectiva. El segundo consiste en proponer una condición necesaria de la justificación colectiva. Esta condición es la de que, si un colectivo C cree justificadamente que p, entonces el que C crea justificadamente que p constituye una razón para que los miembros de C formen, individualmente, la creencia de que p. Esto se deriva de entender a los colectivos a la vez en tanto que dotados de una subjetividad y en tanto que “tecnologías”, esto es, valiosos en la medida en la que permiten a los individuos alcanzar un bien. Puesto que la fiabilidad del proceso de formación de las creencias de un colectivo puede estar presente a la vez que esa condición está ausente, se concluye que la fiabilidad no es suficiente para la justificación colectiva.

## Fiabilidad y suerte

La fiabilidad del proceso de formación de las creencias de un colectivo no puede ser considerada como una condición suficiente de la justificación de las mismas si podemos señalar casos en los que un colectivo forma sus creencias a través de un proceso fiable, pero en los que, de todos modos, dichas creencias no están justificadas. Un tipo de caso posible sería aquél en el que el proceso en cuestión es empleado sólo por suerte, en la medida en la que asumamos que la suerte excluye la justificación.

¿Qué significa que la suerte excluya la justificación? Intuitivamente, no diríamos que, por ejemplo, alguien conoce la respuesta a una pregunta cuando elige qué creer tirando a una moneda. Aunque el resultado de la tirada le lleve a la creencia correcta, ese sujeto no ha llegado a ella a través del ejercicio apropiado de sus facultades cognitivas, y, por ello, la creencia que forme no estará justificada. Ahora bien, existen numerosas concepciones de la suerte. En lo que sigue, voy a asumir como correcta una de dichas concepciones, a saber, la de la suerte como falta de control (Riggs 2009, Broncano-Berrocal 2015). Según esta concepción, un evento es por suerte para un sujeto si, y sólo si, su ocurrencia está más allá del control del sujeto. Además, entenderé por control la capacidad tanto para ejercer control efectivo sobre un evento como para monitorizarlo (*track*), actuando de un modo que nos permitiría alcanzar nuestras metas en relación con la cosa controlada (Broncano-Berrocal 2015).

Así, la estructura del caso a proponer es la siguiente: imaginemos un colectivo C que emplea un proceso  $\Pi$  para la formación de sus creencias colectivas. Dicho proceso es fiable, pero C no tiene control sobre él. Por lo tanto, las creencias que C formando a través de  $\Pi$  serán habitualmente verdaderas, pero sólo por suerte. Así, las creencias de C no estarán justificadas. Por lo tanto, será cierto,

al tiempo, que las creencias de C son fiables y que no están justificadas, con lo que la fiabilidad no es condición suficiente de la justificación de las creencias de C.

Ilustremos esto con un ejemplo. Imaginemos que está por partir la Misión Colonia, un grupo de astronautas que comparten el cometido de colonizar tres lejanos planetas. No se trata de que cada uno vaya a perseguir este fin por su cuenta, sino que, en tierra, se comprometen conjuntamente a ello, formando un colectivo orientado a ese fin. Puesto que la misión debe colonizar tres planetas, la Misión Colonia se divide en tres equipos. Aun así subdivididos, es importante tener en cuenta que siguen formando un único colectivo, ya que siguen compartiendo el fin de colonizar los tres planetas. Así, si cualquier equipo fuera a regresar a la Tierra sin haber cumplido parte y sin mayor explicación, el resto tendría derecho a protestar.

Imaginemos, igualmente, que la Misión Colonia está obligada a formar creencias a lo largo de lo que dure su expedición. Por ejemplo, la Misión debe formar creencias colectivas periódicas acerca de si la colonización se está completando con éxito, acerca de si los planetas en cuestión albergan agua, etc. Pero ocurre que la Misión sabe que, al llegar cada equipo a su destino, le será imposible comunicarse con los demás, porque, digamos, las circunstancias ambientales de los planetas a colonizar estropearán los aparatos de radio. Puesto que una creencia colectiva sólo puede formarse cuando los miembros pueden comunicarse entre sí su compromiso mutuo, a la Misión Colonia le sería imposible formar las creencias colectivas requeridas. Afortunadamente, la Misión cuenta con la ayuda de un extraterrestre de amplios recursos tecnológicos. Este extraterrestre se presta a ayudar del siguiente modo: no transmitirá ningún mensaje de un equipo a otro, sino que se limitará a recabar la información de que cada equipo dispone, la agregará y procesará, y comunicará el resultado a cada uno de los equipos. En otras palabras, el extraterrestre ejercerá de proceso de formación de las creencias colectivas de la Misión. Añadamos, por último, que el extraterrestre es muy fiable, de modo que la Misión gozará de un proceso de formación de creencias colectivas fiable y las creencias colectivas que forme de ese modo tenderán a ser verdaderas.

Ahora bien, ¿diríamos que las creencias de la Misión Colonia, así formadas, estarán justificadas? Mi respuesta es que no. La razón es que no tendrán, *como colectivo*, control sobre el proceso de formación de sus creencias colectivas. Puesto que no pueden comunicarse, no puede tomar la decisión colectiva de afectar al proceso de cualquier modo ni tampoco pueden comprobar su desarrollo ni si las creencias resultantes son verdaderas, ya que, para ello, necesitarían comunicarse. No obstante, se nos podría señalar que, en realidad, la Misión Colonia sí ha ejercido control sobre el proceso. A fin de cuentas, han decidido, estando en tierra, fiarse del extraterrestre y adoptarlo como el proceso de formación de las creencias del colectivo, con la expectativa de que las creencias así formadas serían verdaderas. Por lo tanto, la Misión Colonia formaría creencias verdaderas no por suerte, y nuestro caso no habría servido para ilustrar que la fiabilidad es insuficiente para determinar la justificación de las creencias colectivas.

Juzgo que esta objeción sería correcta, así que es preciso dar todavía un último giro a nuestro ejemplo: digamos que, estando en tierra, los distintos equipos tienen descendencia en los planetas que colonicen, y educar a sus hijos para que sigan

cumpliendo con la misión, explicándoles, por ejemplo, cómo la Misión Colonia forma sus creencias colectivas. Acuerdan, además, darles un sistema para saber si los demás equipos siguen comprometidos y para comunicar si ellos mismos lo están. Por ejemplo, digamos que cada uno de los planetas a colonizar se ve desde los demás, y que cada equipo tiene la capacidad de producir una fuerte luz de color verde o rojo. Así, lo que cada equipo acuerda es enseñar a su descendencia que, si ven que de los planetas en los que están los otros equipos viene luz verde, es que los otros equipos siguen comprometidos con la misión, mientras que, si la luz que llega es roja, significa que están por abandonarla. Así, mientras haya luz verde, la siguiente generación de la Misión Colonia (llamémosla “Misión Colonia<sub>2</sub>”) seguirá constituyendo un solo colectivo, ya que siguen mutuamente comprometidos en la consecución de un fin. Ahora bien, ¿Tendrá la Misión Colonia<sub>2</sub> control sobre el proceso de formación sus creencias? En este caso, la respuesta es no. A diferencia de sus padres, la Misión Colonia<sub>2</sub> nunca tuvo, como colectivo, la posibilidad de acordar o rechazar el uso del extraterrestre como proceso de formación de sus creencias colectivas. A esto se podría objetar con que ambas generaciones de la misión no constituyen dos colectivos distintos sino el mismo, de tal modo que el control que tuvo la primera sobre el proceso de formación de sus creencias lo tiene también la segunda. Podemos contestar de dos modos: primero, invirtiendo los términos: es obvio que la Misión Colonia<sub>2</sub> no ha podido ejercer el mismo control que sus padres, y por lo tanto se trata de otro colectivo. Segundo, debemos individuar los colectivos atendiendo no sólo a sus fines y estructura, sino también a qué miembros los componen. De otro modo, dos colectivos distintos que compartiese fin y estructura, pero cuyos miembros estuviesen en competición, nos parecerían el mismo colectivo.

## Fiabilidad y *Accountability*

Otro modo en el que podemos mostrar que la fiabilidad no es suficiente para que las creencias colectivas estén justificadas es postulando una condición necesaria de la justificación colectiva y pensando casos en los que esta no se da, pero sí la fiabilidad del proceso de formación de las creencias del colectivo. Voy a proponer una condición de este tipo. Para entender qué razones habría para aceptarla, reparemos en dos fenómenos que toda teoría de la justificación epistémica colectiva debería reconocer y conciliar: por un lado, nuestra teoría debería permitir casos en los que el colectivo esté justificado en creer algo, pero no lo esté al menos alguno de sus miembros; puede haber casos en los que un colectivo implementa un buen proceso de formación de creencias colectivas agregando la evidencia que tienen sus miembros, de tal modo que la creencia colectiva resultante estaría justificada, pero en los que la evidencia de que disponen alguno de ellos todavía hace justificado, para ellos, no formar una creencia individual sobre la base de la formada por el colectivo. Por otro lado, no deberíamos contar como creencias colectivas justificadas las de colectivos que formen creencias que, aun bien apoyadas por la evidencia colectivamente procesada, vayan sistemáticamente en contra de las creencias justificadas de sus miembros.

Creo que esto último es intuitivo: nos costaría alabar epistémicamente a colectivos cuyos miembros no puedan beneficiarse de las creencias colectivas formadas.

Esta intuición puede explicarse así: debemos entender a los colectivos no sólo como sujetos de cierto tipo, sino también como “tecnologías”. Típicamente, los colectivos se forman para servir a un fin y, por ello, su valor es el de un instrumento en manos de los individuos. Así, no deberíamos alabar a los colectivos que, llegando a creencias bien formadas, no puedan llevar a sus miembros a formar creencias justificadas sobre esa base, y deberíamos esperar de una teoría de la justificación epistémica colectiva que sólo cuente como casos de creencias colectivas justificadas aquellas formadas por los colectivos cuyos miembros puedan beneficiarse de su membresía.

Así, podemos ofrecer el siguiente principio:

**Accountability Colectivo-Miembro:** “Es condición necesaria de la justificación epistémica de las creencias colectivas que, si un colectivo C cree justificadamente que p, entonces el que C crea justificadamente que p da una razón a los miembros de C para creer individualmente que p.”

Este principio acomoda los fenómenos expuestos: permite casos de colectivos que forman creencias justificadas sin que sus miembros, justificadamente, la compartan, al disponer de razones para no hacerlo más poderosas que la que les da el que el colectivo forme esa creencia; y prohíbe casos de colectivos que formen creencias justificadas que sus miembros nunca estén justificados en compartir, conservando, así, el valor instrumental de los colectivos.

Ahora, podemos imaginar casos de colectivos sin *Accountability* en los que el se implementa un proceso de formación de creencias colectivas fiable. Así, podemos imaginar un colectivo cuyos miembros desconfían justificadamente del proceso de formación de sus creencias colectivas, el cual, a pesar de lo que crean, resulta muy fiable, aunque a ellos les resulta imposible comprobarlo. Así, tenemos casos de colectivos cuyas creencias colectivas son fiables, pero en los que no se cumple *Accountability*. En consecuencia, la fiabilidad no es condición suficiente de la justificación de las creencias colectivas.

## Referencias

- Broncano-Berrocal, Fernando (2015). Luck as risk and the lack of control account of luck. *Metaphilosophy*, 46:1–25.
- Gilbert, Margaret (2014). *Joint Commitment. How We Make the Social World*. Oxford University Press.
- Goldman, Alvin. (2014). Social Process Reliabilism: Solving Justification Problems in Collective Epistemology. In Lackey, J. (ed.) *Essays in Collective Epistemology*. Oxford University Press.
- List, C. y Pettit, P. (2011). *Group Agency*. Oxford University Press.
- Pritchard, Duncan (2005). *Epistemic Luck*. Oxford University Press.
- Riggs, Wayne (2009). Knowledge, luck, and control. In Haddock, A., Millar, A. & Pritchard, D. (eds.). *Epistemic value*. Oxford University Press.

# The bounds of agency: a review of the group person possibility

**Julia Blanco**

First person privileged access to one's mental states has played a central role in the debate around personal identity. Is it possible to maintain a psychological criterion of personal identity without attributing any role to this first person feature? This is the proposal Rovane (1997) brings up: the possibility of finding a criterion of personal identity that is not grounded in consciousness and at the same time allows us to uphold the three main thesis proposed by Locke (1690): 1) A person is an entity with a first person point of view, 2) personal identity consists of the unity and continuity of this point of view and 3) this point of view does not need to be based on a spiritual substance or a biological life. Rovane's theory is based on three central points: an ethically grounded notion of person, a normative analysis of rationality and the rejection of the radical distinction between interpersonal and intrapersonal relationships. Instead of considering the first-person point of view as the phenomenological unified access to one's own thinking, it will be considered as the point of view from which a person deliberates. The purpose of this deliberation will be to make judgements that will be referred to as 'all-things-considered' judgements: judgements about what would be better to think and do taking everything in a person's rational point of view into account. Rovane will call overall rational unity to the state a person would reach if she achieves this goal of deliberation. The idea is that every person is committed to achieving this goal everytime she deliberates.

In the formulation of her condition of personal identity, Rovane (1997) refers to what she calls coordinated activities. This term is used to frame two types of activities with a complex structure: joint activities, or activities carried out by more than one person in order to achieve a common project; and long-term activities: activities carried out by an individual that involve several actions over a period of time, which requires a certain degree of foresight and consideration of our future selves. This communication will show the close link between inter and intrapersonal relationships showing that the distinction between these two kinds of activity is not radical but a matter of degree. The two central ideas will be the following: on the one hand, there is no metaphysical reason for denying the possibility that a joint activity would involve a group of human beings committing themselves to achieve overall rational unity together: it would simply be a case of joint activity with a maximum degree of demand. In the case of long-term activities, the central argument is that the relation established between the present self and the future self is analogous to the one that occurs in a process of interpersonal communication.

Now we can present the condition of personal identity: there must be a set of intentional episodes such that: 1) The intentional episodes within the set are

in the appropriate rational relations to carry out coordinated activities. 2) The set must include a commitment to unifying projects that require the kind of coordinated activities that 1) makes possible. 3) The commitment to carrying out those unifying projects implies the commitment to achieving overall rational unity within the set. Given this definition and the analysis of coordinated activities, there is no reason to reject the possibility of group-persons: a group of human beings who share the same set of intentional episodes and fulfill the condition of personal identity. My objection, however, is aimed at the defense of the possibility of group persons understood as a case of group agency with ‘maximum degree’ requirements over rational deliberation and that obtain over time. It is not the goal of this review to show that group persons are not a metaphysical possibility, but showing that there is an insuperable difficulty for a group person to hold their existence over time. We will further argue that this objection shows that the condition of personal identity, as it was set out before, has a problem to justify temporal continuity of a person, which is the main reason to propose a personal identity criterion. The argument of this objection will consist on showing a case where two human beings—A and B—form a group person. At a given moment, A witnesses a traumatizing image and it deeply alters several of her desires and believes. During the time that goes by until the moment that A shares this information with B, each one of them will have deliberated, not only considering different intentional episodes, but with a different overall rational unity goal. Thus, for each of them, this unity is grounded on different sets of mental states. To sum up, during the period of time between M1 and M2, A and B do not meet the condition of personal identity and therefore we could say that the group-person has been ‘interrupted’. However, it seems like the goal of finding a condition of personal identity is precisely to find a criterion that can explain why we can say of a person that she keeps being the same person over time. Finally, it is my aim to assess the scope of this problem and outline an alternative that allows us to keep the ethical criterion of the notion of person and a condition of personal identity based on agency. This possibility implies returning to the roots of this problem and reevaluating if the question about personal identity is a metaphysical or an epistemic question.



# The role of emotions in philosophical explanations of intentional action

Marta Cabrera Miquel

In this talk, I reflect on the part that emotions play in philosophical explanations of intentional action. Of all the mental states that we traditionally refer to in such explanations, emotions tend to play an extremely insignificant role. Nonetheless, an insightful analysis of emotions and their functions within human psychology seems to indicate that the treatment they have so far been given in our explanations of action entails relevant losses that affect our understanding of human behaviour. Given the capacity that emotions have in guiding and inclining subjects to action, it is surprising how superficial the attention they have received in this area has been. Bearing in mind the way in which philosophers of mind and action of recent years have explained our actions as agents, my aim is to explore how emotions contribute to these explanations and how they relate to the various states and mental dispositions involved in them. At first, I thought that the absence of emotions in the explanations of intentional action was mainly due to the fact that, in certain circumstances, emotions can play a very negative role in belief formation and in rational action. However, a more careful reading has made me see that in recent decades, rather than being ignored, emotions have tended to be reduced into two types of intentional states: beliefs and desires. In this sense, they are not completely absent when it comes to explaining intentional action; one question we will then have to ask ourselves is whether reference to beliefs and desires is enough to make the action derived from an emotion intelligible.

I also aim to achieve a clear understanding of the difficulties of including emotions in the explanations of intentional action. For this purpose, I will briefly examine Donald Davidson's (1980) well-known causal theory of intentional action, paying special attention to his conception of propositional emotions. For Davidson, emotions are reducible to the intentional states that belong to the group of favourable attitudes that, in combination with instrumental beliefs, generate the action of an agent. This particular conception allows him to include emotions in the rationalizations through which we explain or make intelligible the agent's behaviour.

With the aim of highlighting the problems arising from Davidson's treatment of emotions and his way of including them in the explanations of action, I will introduce the criticisms that his model, based on beliefs and desires, has received from three philosophers. First, I will analyse the role that Richard Wollheim (1999) attributes to emotions in the generation of action, as well as the consequences of his reconsideration of the notions of satisfaction and frustration of desires. Secondly, in order to understand the relationship that exists between the beliefs, desires and emotions that we include in our explanations, I will dwell

on the distinction in which Carlos Moya (2001) insists on between internal and constitutive relations and external and causal relations. This distinction will show us what are the difficulties faced by any theory that seeks to understand intentional action in terms of rationality and justification and that aims to give an account of emotions in terms of their relation to beliefs and desires. Thirdly, I will explore the way in which Peter Goldie (2000) conceives the intentionality of emotions, as well as his idea that, instead of talking about rationality or intelligibility, when we try to explain the actions that derive from an emotion we should rather invoke the notions of ‘appropriateness’ and ‘proportionality’.

The review of the objections of Wollheim, Moya and Goldie will help me bring to light the significant weaknesses of the traditional theory of intentional action. Once the three points of view have been considered, I will offer a diagnosis based on the ideas that I believe to be most relevant. Briefly, the main conclusions are the ones that follow:

- (1) The explanation of emotions in terms of beliefs and desires is problematic because there are cases in which, instead of these paradigmatic intentional states, our emotions involve imagination, perception, fantasy, etc., so it is appropriate to expand the notions of satisfaction and frustration of a desire as proposed by both Wollheim and Moya in order to include other mental states in our analysis of the intentional action derived from emotion.
- (2) The attribution of propositional content to emotions faces the harsh criticism to which cognitivism is exposed: by identifying emotions with beliefs or judgments, cognitivism is unable to account for the presence of emotions in young children and animals. Nevertheless, as Goldie points out, there is a sense in which there are conceptual relations between beliefs and emotions, so it is not so clear that we should attribute a brute causality to the impact that emotions have on our beliefs, desires and ultimately our action.
- (3) The phenomenology or felt quality of emotions has no place in the picture that Davidson uses to explain intentional action: the standpoint from which the action is explained is neutral and impersonal to the felt experience.
- (4) The use of the notion of rationality in the explanations of actions in which emotions are involved generates a significant loss to our understanding of such actions.

# A neo-Carnapian ontological commitment for science

Xavier de Donato Rodríguez

José L. Falguera<sup>1</sup>

University of Santiago de Compostela, Spain

**Short abstract:** The present contribution is a defence of a deflationary approach to the ontology of science based on ideas that were presented by Carnap (1950). We argue against the aspirations of the so-called scientific realists in establishing what really exists (according to scientific theories) by concluding that traditional realist arguments, based on the inference to the best explanation and the no-miracle argument, are clearly non-conclusive to the effect of revealing the real ontology of the world. Our aim is not to argue in favor of idealism nor in favor of any other substantive position regarding ontology that commit us with an answer to external questions (independent from any linguistic framework). Rather our aim is to apply Carnap's ideas on ontology to the problem of how to interpret the ontological commitment of a scientific theory and how to respond to the issue about the reference of theoretical terms. We offer different arguments according to which theoretical terms primarily refer to abstract objects (in the version of Thomasson's abstract artefacts).

**Keywords:** ontological commitment; abstract objects; abstract artefacts; scientific realism; theoretical terms; neo-Carnapianism; deflationary approach; ontology of science; metaontology.

## I

To what entities or kind of entities, if any, are we committed when we accept a scientific theory? Is it possible to give criteria for the sort of ontological commitment involved in scientific discourse? One usual way of answering to these questions is to say that scientific theories are committed exactly to those entities whose existence is necessary to presuppose in order that the sentences of the theory, particularly those containing theoretical terms, can be considered to be true. This is the position of a scientific realist, because scientific realists say, among other things, that theories should be interpreted at face value, that

---

<sup>1</sup>This work was supported by the Spanish Ministry of Economy and Competitiveness under project FFI2017-82534-P and by the National Agency of Scientific Promotion and Technology of Argentina under project PICT 2014-01741. ORCID ID of Xavier de Donato: [orcid.org/0000-0002-8464-9960](https://orcid.org/0000-0002-8464-9960); ORCID ID of J.L. Falguera: [orcid.org/0000-0003-2631-1563](https://orcid.org/0000-0003-2631-1563).

is, in terms of a realist semantics. In order for this to be possible, it is clear that the terms used in a true theory (in particular its theoretical terms) should have reference in the real world. Consequently, if one wants to maintain that a given theory (a theory that is currently accepted by scientists in face of the evidence) is true or approximately true, it seems that one must be committed to the existence of theoretical entities. As scientific realists want to maintain that our most successful and well confirmed scientific theories are (approximately) true, they must say that the acceptance of a theory constrains us to postulate the existence of certain theoretical entities. It is a supposition of all this the claim according to which one of the roles of scientific theories is to establish what really exists (as part of the world) and that scientific theories can in fact serve to do this job.

But, although being a possible answer to our initial questions, the realist view is not the only one. Carnap is famous for having believed that scientific theories, even the best currently accepted, do not serve to the purpose of establishing what really exists. It is relatively common to find in the current literature on ontological commitment that Quine's criterion based on first-order quantification is what realists have in mind when they consider the issue of what entities are we committed to when accepting a scientific theory (those entities which our best theories actually quantify over). Leaving aside the question whether Quine himself was a scientific realist, the fact is that his criterion has been used by realists in order to design a framework to determine the ontology of a theory. Quine (1948) was famously reacting against Carnap (1950/56), who seemed to be committed to the existence of numbers, properties, propositions and other "dubious" entities. But Carnap never boarded any form of Platonism nor any form of ontological realism in any field of philosophy. Quite at the contrary, since the times of the *Aufbau*, he is known to maintain a purely neutral, agnostic position as far as traditional (or external) ontological matters are concerned. It is relatively frequent, even nowadays (see Demopoulos 2011 and Lavers 2016) to attribute to Carnap an instrumentalism that would go so far as to reject the existence of theoretical entities. But this is not Carnap's position. Carnap's view is that (1) even our best scientific theories do not serve to the purpose of establishing what really exists or what the world is really made of, or, put in other words, scientific theories do not serve for the general aim of traditional metaphysics or ontology, and (2) we cannot be justified in our claims about the existence of theoretical entities neither about their inexistence (at least if these claims are understood in terms of what realists or antirealists want to establish). In the present paper, we defend a neo-Carnapian approach to ontological commitment according to which the commitment involved in accepting a scientific theory involves just a deflationary or "internal" acceptance of theoretical entities, rejecting a metaphysical or "external" commitment to their existence (as well as an antirealist claim of metaphysical inexistence). In this regard, our view argues in favor of a (deflationary) commitment to abstract entities for theoretical terms (as Carnap 1950/56). Though, unlike Carnap, these abstract entities are not mathematical entities but what we can call "fictional objects" (considered as abstract artifacts).

First, we will synthesize the arguments against the realist view of scientific theories and, in particular, against the pretension that scientific theories can help to establish the real ontology of the world. The scientific realist typically assumes,

among other things, that our best scientific theories must be (approximately) true (in a realist sense). She typically argues this by appealing to the inference to the best explanation (IBE) and by the so-called “no-miracle argument” (NMA). Recall that IBE is a procedure to choose the theory which could be considered the best to explain some phenomena and NMA is an argument according to which scientific realism is the only approach that allows us to establish that the achievements of science are not a miracle (Putnam 1975, 73). In short, against IBE we can appeal to the fact that this kind of inference is of no effect in face of empirically equivalent theories and that, even in absence of empirically equivalent theories, IBE does not guarantee truth. Furthermore, we have the so-called metainductive pessimism (Laudan 1981). Against NMA, we can reply that this argument relies on a notion of explanation according to which truth is needed in order for the explanation to work and to be accomplished, but we must observe that this approach to explanation is not the only one nor the most satisfactory in the market. At least some alternative approach would not allow for a non-miracle inspired argument (see, for instance, van Fraassen’s 1980 pragmatic approach). Another point against NMA is that theories may well have success in predicting and explaining phenomena even if they are not true. Some other explanation (for instance, adaptationism) could perhaps also explain the success of our theories. Colin Howson (2000) and others have even maintained (through this has been disputed) that NMA is flawed because it commits the base rate fallacy. If all this suffices to dispute the main arguments in favor of scientific realism, it seems that we must conclude that (our best) scientific theories are not appropriate for establishing what really exists in regards to the material world, as we cannot guarantee that our theories are really true. This conclusion would give reason to Carnap when he argued that scientific theories are not useful to answer the questions of traditional ontology.

## II

If we accept this, now comes the question whether this shows that scientific theories do not allow for any form of ontological commitment. And here comes Carnap to help us. In a Carnapian spirit, we can claim that scientific theories *do* involve a particular kind of ontological commitment; it is just that the commitment has not the same sense as it had in the case of the realist view. Carnap (1950/56) carefully distinguished internal and properly meaningful existence questions from external and meaningless existence questions. This perspective presupposes that internal questions depend upon the fact that we have already available a linguistic framework with proper rules governing the use of the terms of that language. Existence questions that go far beyond linguistic frameworks do not have cognitive value at all and, hence, are pseudo-questions. Internal questions, however, do have a cognitive value (logical or empirical) but in return receive a trivial answer within the linguistic framework. As Thomasson (2016) has suggested, one of the main reasons for dismissing Carnap’s views on ontology has been perhaps that an extended interpretation of what he says comes in form of taking Carnap as implying a form of relativism (Eklund 2013) or in terms of a form of anti-realism (Putnam). A crucial point of these interpretations has to do with the fact that Carnap’s view is, commonly but incorrectly,

understood as implying quantifier variance (see Putnam 1987 and 1990, Hirsch 2002, and Fine 2009). In the line of Thomasson (2015, 2016), we interpret Carnap's deflationism as being committed neither to quantifier variance nor to any antirealist metaphysical position. So, quantifiers cannot involve an ontological commitment in the external sense. But, trivially, quantifiers involve ontological commitment to the entities over which they quantify in the internal sense.

Recall that Carnap maintains that any external claim of existence and any assertion about the real events that actually constitute the world are devoid of cognitive content. This is another way of saying that theories cannot help to establish what is the *real ontology* of the world. An important consequence of this is that neither Platonism nor antirealism can be in the end maintained, as the questions they try to answer are not decidable by empirical or analytical means and, hence, they are *pseudoquestions*. So, the only way to accept entities like properties, propositions, numbers and the like is to accept their existence *in relation* to a certain linguistic framework. To admit a new kind of entity is in fact to accept a certain linguistic framework. As Carnap says: "[w]e may still speak (and have done so) of 'the acceptance of the new entities' since this form of speech is customary; but one must keep in mind that this phrase does not mean for us anything more than acceptance of the new framework; i.e., of the new linguistic forms. Above all, it must not be interpreted as referring to an assumption, belief, or assertion of 'the reality of the entities'." (Carnap 1950/56, 214)

Now, in the case of scientific theories, what kind of existence questions are admissible, that is, do in fact have sense or cognitive content? Scientific theories, considered as linguistic frameworks in the way Carnap understands them, commonly involve an internal commitment to three kind of entities: (i) the entities designated by mathematical expressions that are needed in many of our current scientific theories, (ii) the entities that are designated by the observational terms of the theory, and (iii) the entities that are designated by its theoretical terms. Entities of type (i) typically are taken to be abstract entities (but, remember, not in a Platonic sense), because mathematical entities are one of the most paradigmatic examples of abstract objects. Entities of type (ii) are typically individuals, properties or relations that are observable, though our commitment with them should not be an external commitment because they arise from how the world is phenomenologically given to us, not from how the world really is. What Carnap calls "the world of things" is just "the spatio-temporally ordered system of observable things and events" (Carnap 1950/56, 207). When Carnap considers the thing language as a kind of very basic linguistic framework, he points out that the concept of reality occurring in the internal questions made within the thing language is an empirical concept, obviously not a metaphysical one. "To recognize something as a real thing or event means to succeed in incorporating it into the system of things at a particular space-time position so that it fits together with the other things as real, according to the rules of the framework." (ibid.). These entities of type (ii) can be considered as concrete entities. Finally, entities of type (iii) are for Carnap mere mathematical entities, and therefore abstract. This can be surprising. But Carnap (1956, § IV) insists on this consideration by indicating that the entities of the theoretical vocabulary are based on a system of spatio-temporal coordinates, where spatio-temporal points are ordered quadruples of real numbers, the spatio-temporal regions are classes of

points dots, and the physical magnitudes functions of points or spatiotemporal regions to real numbers or tuples of real numbers. Carnap (1958 and 1966) uses the Ramsey-sentence to provide the empirical content of a theory in a way that avoids the use of theoretical terms but does not aim to avoid the commitment to their references, but understanding these as logico-mathematical entities. In fact, Carnap acknowledges, in response to Hempel, that the Ramsey sentence designates theoretical entities, but adds that such entities are not supposed to be physical entities not observed (and not observable) but logical-mathematical entities. In this regard, he points out: “I agree with Hempel that the Ramsey-sentence does indeed refer to theoretical entities by the use of abstract variables. However, it should be noted that these entities are not physical objects, such as atoms, electrons, etc., but rather (at least in the form of the theoretical language that I have chosen [...] purely logical-mathematical entities, eg, natural numbers, classes of such, classes of classes, etc” (Carnap 1963, 963). The reason of Carnap for rejecting that the theoretical terms of an accepted theory designate non-observable entities, and therefore concrete, we believe may be due to the fact that such a position would not allow to differentiate such entities from those assumed by a scientific realist and would lead to Carnap to assume ontological commitments of external character.

Following Carnap, we can also understand scientific theories as linguistic frameworks, so that any scientific theory (currently considered correct or rejected as incorrect) entails its own internal ontology, though only when a scientific theory is considered correct by someone it entails an (internal) ontological commitment for her. Once assumed that ontological questions are internal questions relative to a linguistic framework, for the case of empirical theories we distinguish among empirical objects, theoretical objects and mathematical objects (where objects can be particulars, substances, properties or relations). We will consider that empirical objects are concrete objects because they can be perceived by someone, and theoretical and mathematical objects are abstract artefacts. But here comes an important difference between our approach and Carnap’s. A peculiarity of our approach is that the objects which theoretical terms directly refer to are abstract objects, though not because they are logico-mathematical constructions, but because they are fictional objects (assuming that fictional objects are a type of abstract artefacts). It seems to us that it is not plausible to take the relevant functions as ranging just over mathematical entities: one thing is to assign numerical values to an entity, another is to *reduce them* to these numerical values. On the other hand, the reason to say that theoretical terms refer to fictional objects is that, as scientific theories are highly idealized, theoretical expressions cannot primarily refer to concrete objects (neither to external concrete objects, because these would have to do with the kind of external ontological questions which are not clarified with the help of scientific theories, nor to internal concrete objects, because they are not empirical objects and other supposed concrete entities (a) cannot properly satisfy many of the idealized theoretical conditions that are established by the theory and (b) would be indistinguishable from putative external theoretical objects). Finally, taking this option does not commit us to Platonism in a double sense. Firstly, as it has been said, our view does not amount to an external claim regarding the existence of abstract objects. And secondly, when we admit fictional (and mathematical) entities as abstract objects we do not need to understand them

as eternal and mind-independent objects. We can appeal to Thomasson (1999) in order to understand them as a particular case of “abstract artefact” (this is not Thomasson’s proposal for the case of theoretical terms but to the case of fictional characters, but we extend her position to the scientific case for the reasons we have given). In relation to all this, we will also show the similarities (as well as the differences) that our view has with respect to another well-known deflationary approach to ontology (and metaontology), that of Thomasson (2015).

It is important to highlight that we don’t reject the possibility that theoretical terms of our best theories *indirectly* (and perhaps approximately) refer to real *concrete* objects. Obviously, we don’t assert it. Our position on this is agnostic. Our view is agnostic as far as metaphysical positions both concerning the existence of external entities and the interpretation of scientific theories (realist/antirealist debate in philosophy of science), just as Carnap’s view was (contrary to many interpretations of his view). Simply, the real existence (or inexistence) of theoretical entities is an issue that scientific theories cannot help to answer. And trying to answer this question is trying to answer external questions which are not resolved by analytical or by empirical ways.

Finally, our approach has the advantage of being able to provide a unique homogeneous analysis for mathematical terms, as well as for theoretical terms of theories that are rejected/accepted. It also makes sense of the referential divergence of theoretical terms of rival theories across radical changes. It also lacks the problems of the realist approach to the ontology of science without committing us to any form of antirealism. It gives sense to the idea that scientific theories contain counterfactual and highly idealized assumptions. And, more importantly, in allowing for the theoretical terms of any theory to be referential, it also allows for a semantics that respects our intuitions regarding different kinds of ordinary sentences containing theoretical terms (and other kind of terms).

## References

- Carnap, R. (1950/56). “Empiricism, Semantics, and Ontology”, in *Meaning and Necessity*, Chicago: University of Chicago Press.
- Carnap, R. (1956). “The Methodological Character of Theoretical Concepts”, in *Minnesota Studies in the Philosophy of Science*, vol. I, ed. by H. Feigl and M. Scriven, Minneapolis: University of Minnesota Press.
- Carnap, R. (1958). “Beobachtungssprache und theoretische Sprache”, in *Dialectica*, XII (English trans.: “Observation Language and Theoretical Language” in *Rudolf Carnap, Logical Empiricist*, Dordrecht: D. Reidel, 1975).
- Carnap, R. (1963). “Replies and Systematic Expositions”, in P. A. Schilp (ed.). *The Philosophy of Rudolf Carnap*, La Salle, Ill.: Open Court, 857-1012.
- Carnap, R. (1966). *Philosophical Foundations of Physics*, ed. by Martin Gardner, New York: Basic Books.



- Demopoulos, W. (2011). “On Extending ‘Empiricism, Semantics, and Ontology’ to the Realism/Instrumentalism Controversy”. *Journal of Philosophy*, 153:647–669.
- Eklund, M. (2013). “Carnap’s Metaontology”. *Noûs*, 47(2):229–249.
- Fine, K. (2009). “The Question of Ontology”, in D. Chalmers & D. Manley and R. Wasserman (eds.): *Metametaphysics*, Oxford: Clarendon Press, 157–177.
- Hirsch, E. (2002). “Quantifier Variance and Realism”, in E. Sosa and E. Villanueva (eds.): *Realism and Relativism, Philosophical Issues* 12, Oxford: Blackwell, 51–73.
- Howson, C. (2000). *Hume’s Problem: Induction and the Justification of Belief*. Oxford: Oxford University Press.
- Laudan, L. (1981). “A Confutation of Convergent Realism”. *Philosophy of Science*, 48:19–48.
- Lavers, G. (2016). “Carnap on Abstract and Theoretical Entities”, in Blatti, S. and S. Lapointe (2016, eds.): *Ontology after Carnap*, Oxford: Oxford University Press, 13–30.
- Putnam, H. (1975). *Mathematics, Matter and Method*. Cambridge: Cambridge University Press.
- Putnam, H. (1987). *The Many Faces of Realism*. La Salle, Ill.: Open Court.
- Putnam, H. (1990). *Realism with a Human Face*. Cambridge, MA: Harvard University Press.
- Quine, W.V.O. (1948). “On What There Is”, in *From a Logical Point of View*, Cambridge, Harvard U.Pr.
- Thomasson, A.L. (1999). *Fiction and Metaphysics*. Cambridge: Cambridge University Press.
- Thomasson, A.L. (2015). *Ontology made easy*. Oxford: Oxford University Press.
- Thomasson, A.L. (2016). “Carnap and the Prospects of Easy Ontology”, in Blatti, S. and S. Lapointe (2016, eds.): *Ontology after Carnap*, Oxford: Oxford University Press, 122–144.
- van Fraassen, Bas C. (1980). *The Scientific Image*. Oxford: Oxford University Press.

# *Musement* o la actividad de la *red neural por defecto*

Antonio Duarte Calvo  
Dpto. de Lógica y Filosofía Teórica  
Universidad Complutense de Madrid

## Resumen

En esta contribución me centraré en analizar la noción de *Musement* introducida por Peirce en 1908 para, paso a paso y siguiendo la descripción del propio Peirce, identificar esta “agradable ocupación de la mente” con los procesos mentales propios del ser humano que surgen a través de la *red neural por defecto*. Esta red constituye un sistema cerebral que ha sido recientemente definido anatómicamente y que se activa en los individuos, preferentemente, en momentos de pasividad, cuando las exigencias de relacionarse con el entorno exterior se relajan.

## Introducción

“Semiotics begins and ends with biology and the sign science and the life science ineluctably imply each other.”  
(Sebeok 1979, p. viii)

La noción de *Musement* fue introducida por Charles S. Peirce en su artículo *The Neglected Argument of the Reality of God* (1908) como “una actividad mental no reflexiva, capaz de dar una especie de respuesta instintiva para producir una fuerte creencia en la realidad de Dios”<sup>1</sup> (Rodrigues 2017, p. 164). En el texto de Peirce se vislumbra que esto es posible porque el *Musement* trae consigo una abducción (Barrena 2008; Nubiola 2004), siendo este tipo de razonamiento por el que los individuos pueden adquirir “fuertes creencias”. Así, en primer término, analizaremos el *Musement* considerando el papel que juegan en él la abducción y la creatividad. Una vez definido y caracterizado el concepto, veremos su estrecha conexión, podríamos hablar prácticamente de una perfecta identificación, con la actividad de la *red neural por defecto*. En la última década, los neurocientíficos han definido y demarcado esta red neural que participa en modos internos de cognición (Buckner et al. 2008).

---

<sup>1</sup>En este artículo muestro mis propias traducciones de las citas cuyos textos originales sean en inglés salvo que se indique una edición traducida al español en las referencias.

## ¿Qué es el *Musement*?

Resulta casi imposible definir *Musement* de manera clara e inequívoca y sin embargo, siguiendo el texto de Peirce, somos capaces de identificar claramente a qué actividad humana se refiere. Como se ha apuntado más arriba, para Peirce, el *Musement* puede traer consigo la idea de “la realidad de Dios” pero, desde luego, no es esta idea o resultado lo que caracteriza al *Musement*. Sin duda, a un escaso número de personas les resultaría fácil reconocer en sí mismos el *Musement* siguiendo la definición de Rodrigues dada más arriba. La clave de la definición reside en “actividad mental no reflexiva” y “respuesta instintiva para producir una fuerte creencia”. En la definición de Nubiola (2004, p. 87), con toda probabilidad, podemos empezar a identificar esta actividad de la mente: *Musement* es “la meditación libre, el juego de la mente que contempla el universo, un peculiar estado de la mente y del corazón que van de una cosa a otra de modo libre, sin seguir regla alguna.”

En cualquier caso, siguiendo paso a paso la descripción de Peirce (CP, 6.458-6.464, 1908) es donde hallamos la plena identificación de esta actividad con los derroteros que toma nuestra mente en muchos de nuestros momentos “pasivos”. A través del texto de Peirce, podemos destacar la siguiente lista de sus rasgos característicos:

*Musement* es esa “agradable ocupación de la mente que...” (Peirce, CP 6.458, 1908):

1. No posee un nombre distintivo. (Peirce, CP 6.458, 1908).
2. No tiene un propósito definido aunque es diferente de la ensoñación. (Peirce, CP 6.458, 1908).
3. Es un Puro Juego de la mente sin reglas, excepto la ley de la libertad. (Peirce, CP 6.458, 1908).
4. Conecta dos de los tres Universos. (Peirce, CP 6.458, 1908).
5. Puede comenzar pasivamente en cualquier momento. (Peirce, CP 6.459, 1908).
6. Es susceptible de convertirse en análisis científico. (Peirce, CP 6.459, 1908).
7. Es capaz de resolver problemas. (Peirce, CP 6.460, 1908).
8. No se restringe a un tipo concreto de razonamientos. (Peirce, CP 6.461, 1908).
9. Es más amplio que un método lógico o de estudio científico. (Peirce, CP 6.461, 1908).
10. Actúa como un diálogo interno de las ideas y las imágenes. (Peirce, CP 6.461, 1908).
11. Está abierto a tomar distintos rumbos. (Peirce, CP 6.462, 1908).
12. Plantea sugerencias plausibles susceptibles de futura verificación. (Peirce, CP 6.464, 1908).

Como Puro Juego, el *Musement* no es una tarea mental pesada o que requiera una especial concentración. Empieza de manera pasiva, en aquellos momentos en los que la mente se encuentra libre y sin ataduras, cuando nos relajamos. Los tres Universos de la Experiencia peirceanos (el de las Ideas, el de la Realidad en bruto y el de los Signos) pueden ser conectados en este juego debido a la misma ausencia de límites, y es en este momento donde el puro juego se convierte en *Musement*. Es diferente de la ensoñación ya que, aunque de manera distante y distraída, suele girar en torno a nuestra propia realidad. En ocasiones, por ejemplo, mediante este puro juego de la mente y lo que parecen ser pensamientos inconexos, nos sorprendemos al acordarnos de alguna tarea pendiente que nuestro cerebro había olvidado que teníamos que realizar. Otras muchas veces, jugando al *Musement*, encontramos espontáneamente la idea clave para solucionar una cuestión que ha ocupado largamente nuestra mente y no habíamos sido capaces de resolver.

Conectando en este punto con la teoría de Peirce de la abducción, después de esta sugerencia del *Musement*, si queremos conocer el valor de esta hipótesis, es necesario llevar a cabo la subsiguiente verificación de la misma. Por tanto, *Musement* es puro juego, sin reglas, excepto la libertad, donde la mente trata de conectar los distintos universos con la búsqueda de la explicación que se demanda (Sebeok y Umiker-Sebeok, 1988, pp. 26–28). Por su parte, Wirth (2014, p. 6) lo resume como “el movimiento mental que oscila entre la libertad y la reocupación de la mente por los límites del autocontrol” y añade, “el pensamiento abductivo y la conjetura son, en un principio, ‘puro Juego’, el cual no está controlado por la razón crítica. [...] El resultado del *Musement* puede ser una abducción creativa”. Este puro juego podría considerarse el motor de ciertas conexiones entre las distintas inclinaciones y conocimientos que guían de manera única y exclusiva a cada uno de nosotros a abducir de una manera particular.

## ¿Cómo jugamos al *Musement*?

El *Musement* se aprecia de manera especial cuando nuestras ideas aparecen mientras realizamos alguna actividad despreocupada y lúdica; incluso muy a menudo somos conscientes de qué tipo de actividades nos ayudan a desentrañar los más intrincados enigmas que, en muchas ocasiones, nuestros más concentrados y profundos pensamientos no son capaces de resolver.

Es significativo encontrar múltiples relatos donde campeones abductivos encuentran la clave que proporciona una hipótesis realmente novedosa en momentos de relajación. Así, por ejemplo, Sherlock Holmes se entrega al *Musement* cuando se enfrasca en la música que sale de su violín; habitualmente lo hace en mitad de una importante investigación. Watson, nos lo narra en diversas ocasiones: “ya muy avanzada la noche pude escuchar el gemir melancólico y apagado de su violín, indicio de que seguía meditando sobre el extraordinario problema cuya aclaración se había propuesto” *Estudio en Escarlata* (Doyle 2008, p. 90). Y no puede ser otra cosa más que puro juego, cuando Einstein resolvía algunos problemas con esta misma afición, según el testimonio de su hijo Hans Albert: “A menudo tocaba el violín en la cocina hasta altas horas de la noche, improvisando melodías mientras reflexionaba sobre complicados problemas [...] Luego, de re-

mente, en plena interpretación, anunciaba con excitación: ‘¡Lo tengo!’” (Isaacson 2009, p. 40). Es fundamental recordar que para tocar un instrumento hacemos uso de la memoria procedimental, por lo que el acceso a estos recuerdos y su utilización se realiza sin la necesidad de una atención o control conscientes, es decir, dejando la mente en un estado libre. Y es uno de estos momentos donde dejamos volar los pensamientos, el que describe V. Grossman cuando desvela el momento de iluminación de Víktor Pávlovich, físico nuclear de la URSS en la época de la Segunda Guerra Mundial, uno de los personajes principales de *Vida y destino*:

Víktor Pávlovich caminaba en la oscuridad por la calle desierta. De repente le vino a la cabeza un pensamiento inesperado. Y enseguida, sin dudar, supo que ese pensamiento era cierto. Tenía una nueva explicación para el fenómeno atómico que hasta ahora parecía no tener explicación y los abismos se habían transformado en puentes. ¡Qué sencillez, qué luz! Aquella idea era sorprendentemente bella. Parecía que ni siquiera la hubiera engendrado él, como un nenúfar blanco que emergiera de la oscuridad serena de un lago. Exclamó admirando su belleza... Qué extraña coincidencia, pensó de repente, que aquella idea se le hubiera ocurrido cuando su mente estaba tan alejada de las reflexiones científicas, cuando las discusiones sobre el sentido de la vida le tenían absorbido. (Grossman 2007, pp. 365–366).

Aunque no todos tengamos esa extraordinaria capacidad abductiva de nuestros protagonistas anteriores, nosotros, también somos capaces de identificar diversas ocasiones en las que en estado de *Musement*, es decir, cuando pensamos que no estamos pensando en nada, hemos sentido una iluminación, que tal vez, a los pocos segundos se haya convertido en una hipótesis. En cualquier caso, este “no pensar en nada” lo reconocemos como uno de nuestros más valiosos momentos. Como decía Borges: “Pensar de tarde en tarde en Sherlock Holmes es una de las buenas costumbres que nos quedan. La muerte y la siesta son otras. También es nuestra suerte convalecer en un jardín o mirar la luna.” Quizás, la libertad que en ese momento sentimos sea precisamente el fruto de que cuando nosotros no pensamos en nada, nuestro cerebro puede “pensar” en todo.

## ***Musement* como actividad mental surgida a través de la red neural por defecto**

“When the brain/mind thinks in a free and unencumbered fashion, it uses its most human and complex parts.”  
(Andreasen et al. 1995, p. 1583)

Siguiendo la revisión de Buckner et al. (2008) describiré muy brevemente el reciente descubrimiento de la red neural por defecto. Habiendo, previamente, desentrañado las características del *Musement*, no será difícil encontrar una perfecta correspondencia con la actividad mental que esta red genera.

La red neural por defecto es un sistema cerebral específico y anatómicamente

definido que actúa cuando la exigencia de relacionarse con el entorno se relaja (*surge pasivamente* [5]). La evidencia de esta red surgió cuando los neurocientíficos empezaron a “medir” la actividad cerebral en reposo como estado de control experimental. Sorprendentemente, estos estudios revelaron que la actividad en ciertas regiones cerebrales específicas aumentaba durante los estados de control en comparación con las medidas del cerebro cuando se proponían en los sujetos tareas dirigidas.

A partir de los trabajos de Raichle, Gusnard y colegas (Raichle et al. 2001; Gusnard et al. 2001) se comenzó a estudiar esta actividad del cerebro como un sistema neurobiológico fundamental con propiedades fisiológicas y cognitivas que lo distinguen de otros sistemas (*no posee nombre distintivo, pero es una ocupación mental definida* [1]). La posibilidad de que la red neural por defecto contribuya a modos internos de cognición es consistente con los subsistemas que conforman su anatomía. La red se entiende mejor como subsistemas múltiples que interactúan. El subsistema del lóbulo temporal proporciona información de experiencias previas en forma de recuerdos y asociaciones, componentes básicos de lo serían las simulaciones mentales. El subsistema prefrontal facilita el uso flexible de esta información. Estos dos subsistemas convergen en ciertos nodos de integración (*conecta dos de los tres Universos* [4]). Cuando se realizaban los experimentos, a menudo, los pensamientos de los participantes giraban en torno a planes futuros, eventos personales recientes o un simple vagar de la mente (*sin propósito definido* [2], *puro juego de la mente sin reglas* [3], *diálogo interno de las ideas y las imágenes* [10]). Una hipótesis de trabajo es que la función principal de esta red es la de construir simulaciones mentales internas que se utilizan de forma adaptativa (ampliando el tema esto estaría íntimamente relacionado con el concepto peirceano de *hábito virtual*, ver West 2016). Estos eventos imaginarios tienden a ser prácticos y libres de fantasía (*no tiene propósito definido aunque es diferente de la ensoñación* [2]). Además, una propiedad definitoria de la red neural por defecto es su flexibilidad. Cuando la red se activa, aparecen procesos centrales comunes, pero el contenido y el objetivo al que se aplican estos procesos son de lo más variado (*puro juego de la mente sin reglas, excepto la ley de la libertad* [3], *no se restringe solo a un tipo concreto de razonamientos* [8], *está abierto a tomar distintos rumbos* [11]).

Por otro lado, en neurociencia, estudios experimentales recientes sobre la creatividad (véase, por ejemplo, Beaty et al. 2014) sugieren una conectividad funcional directa entre la red neural por defecto y la creatividad (*es más amplio que un método lógico o de estudio científico* [9], *es capaz de resolver problemas* [7], *plantea sugerencias plausibles susceptibles de futura verificación* [12]). Los resultados de Beaty y colaboradores apuntan que la habilidad del pensamiento divergente requiere de una mayor cooperación entre las regiones del cerebro asociadas con procesos cognitivos tanto controlados como espontáneos. Como una cualidad propia de los procesos creativos, estas ideas nuevas surgidas mediante la actividad de la red neural por defecto son susceptibles de un posterior desarrollo y estudio (*es susceptible de convertirse en análisis científico* [6]).

Así, el *Musement*, aquella actividad mental definida por Peirce hace más de cien años, encuentra en el siglo XXI su razón neurobiológica. Es más, el creciente número de científicos que, tras el descubrimiento de esta red neural, defienden el ocio y el alejamiento de ambientes multi-tarea (en especial aquellos asocia-

dos con las nuevas tecnologías) como elementos esenciales para el desarrollo de la creatividad humana (ver, por ejemplo, Smart 2014), parecen seguir el primer pensamiento que nos ofreció Peirce sobre el *Musement*: “There is a certain agreeable occupation of mind which, from its having no distinctive name, I infer is not as commonly practiced as it deserves to be.” (Peirce, CP 6.458, 1908).

## Referencias

- Andreasen, N. C., D. S. O’Leary, T. Cizadlo, S. Arndt, K. Rezai, G. L. Watkins, L. L. Boles Ponto y R. Hichwa (1995). “Remembering the past: Two facets of episodic memory explored with positron emission tomography.” *American Journal of Psychiatry*, 152(11):1576–1585.
- Barrena, S. (2008). “El ‘Argumento Olvidado’ de C. S. Peirce cien años después.” *Methodus*, 3:23–45.
- Beaty, R. E., M. Benedek, R. W. Wilkins, E. Jauk, A. Fink, P. J. Silvia, D. A. Hodges, K. Koschutnig y A. C. Neubauer (2014). “Creativity and the default network.” *Neuropsychologia*, 64:92–98.
- Buckner, R. L., J. Andrews-Hanna y D. L. Schacter (2008). “The Brain’s Default Network.” *Annals of the New York Academy of Science*, 1124:1–38.
- Doyle, A. C. (2008)/[1887-1917]. *Todo Sherlock Holmes*. Ediciones Cátedra.
- Grossman, V. (2007)/[1959]. *Vida y destino*. Galaxia Gutenberg. Círculo de Lectores.
- Gusnard, D. A., E. Akbudak, G. L. Shulman, y M. E. Raichle (2001). “Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function.” *Proceedings of the National Academy of Science of the United States of America*, 98:4259–4264.
- Isaacson, W. (2009)/[2007]. *Einstein, su vida y su universo*. Debolsillo.
- Peirce, C. S. (1931-1958)/[1860-1913]. *Collected Papers*, vols. 1-8, C. Hartshorne, P. Weiss y A. W. Burks (Eds.). Cambridge, MA: Harvard University Press. (CP)
- Nubiola, J. (2004). “C. S. Peirce y la abducción de Dios.” *Tópicos*, 27:73–93.
- Raichle, M. E., A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard y G. L. Shulman (2001). “A default mode of brain function.” *Proceedings of the National Academy of Science of the United States of America*, 98:676–682.
- Rodrigues, C. T. (2017). “The Vagueness of the Muse.” *SOPHIA*, 56:163.
- Sebeok, T. A. (1979). *The sign & its masters*. University of Texas Press, Austin.
- Sebeok, T. A y J. Umiker-Sebeok (1988)/[1983]. “‘You Know My Method.’ A Juxtaposition of Sherlock Holmes and C.S. Peirce.” En U. Eco y T. A. Sebeok (Eds.), *The sign of three: Dupin, Holmes, Peirce* (pp. 11–54). Indiana University Press.

- Smart, A. J. (2014)/[2013]. *El arte y la ciencia de no hacer nada*. Clave intelectual.
- West, D. E. (2016). “The Abductive Character of Peirce’s Virtual Habit.” *Semiotics 2016: Archaeology of Concepts*, pp. 13–22.
- Wirth, U. (2014). “Abduction, Wit, Stupidity – from Peirce to Freud.” En M. Bergman y J. Queiroz (Eds.), *The Commens Encyclopedia: The Digital Encyclopedia of Peirce Studies*. Pub. 121219-1823a.



# Hacia una noción de verdad compatible con un pluralismo genuinamente pragmatista o pluralismo radical

Jorge Alejandro Espinosa López  
UNAM

## I Pluralismo

- 1.1. El pluralismo se define como la visión en la que se sostiene la posibilidad de tener diferentes versiones de mundo con distintas ontologías, cada una de ellas con un interés e importancia independiente, sin que exista algún requerimiento o presunción de reducibilidad a una base común (Goodman, 1978, p. 4).
  - 1.1.1. El pluralismo es una visión opuesta al *monismo*, este último afirma la posibilidad de que exista *un* solo tipo de conocimiento válido del mundo y una *única* ontología última y fundamental.
  - 1.1.2. En su versión científicista, el *monismo* se expresa a través de la tesis *reduccionista* en la que se asume el eventual establecimiento de una teoría única, completa y comprensiva del mundo natural basada en un conjunto fundamental de principios (Ruphy, 2013).
- 1.2. El pluralismo tiene diferentes alcances, estos pueden ser epistemológicos (existen distintas formas mediante las cuales conocemos y clasificamos el mundo) u ontológicos (no solamente existen distintas formas de conocer y clasificar el mundo, además cada una de estas formas tiene su propio dominio de objetos, es decir, su propia ontología).
- 1.3. Quienes suscriben las tesis pluralistas asumen que existen distintos niveles y campos de explicación completamente autónomos e independientes (con su propia ontología) que no pueden ser reducidos a un único discurso como en 1.1.2.
- 1.4. Los detractores sostienen que el pluralismo es una forma de relativismo en el que no existen estándares, ni criterios fijos para explicar y regular las prácticas científicas. Nociones como objetividad, y verdad pierden relevancia y valor normativo.
  - 1.4.1. El argumento de 1.4 no solamente es sostenido por los detractores del pluralismo, sino también por filósofos pluralistas que desean conser-

var sus intuiciones realistas más básicas y que temen que su propuesta sea interpretada como un relativismo.

- 1.5. Hay tres estrategias relevantes que detractores y simpatizantes asumen para evitar el relativismo:
- a) Negar el pluralismo. Afirmar que es posible una reducción como en 1.1.2.
  - b) Descripciones elípticas. P. ej.
    - a') The king of Sparta had two votes according to Herodotus.
    - b') The king of Sparta had one vote according to Thucydides.
  - c) Suponer un sustrato ontológico mínimo que ponga coto a nuestras representaciones.
- 1.6. a) no es relevante para mi análisis, pues asumo que el pluralismo es la visión que mejor explica las prácticas científicas. b) y c) son tesis complementarias. b) es una estrategia bastante cuestionada, pues a') y b') no nos dicen nada de Esparta, sino solamente lo que Herodoto y Tucídides dicen sobre ella dentro de un Esquema Conceptual (EC). Así, tenemos dos verdades que no nos dicen nada del mundo que describen, sino solamente sobre aquello que se dice del mundo dentro de un EC. c) también es una mala estrategia pluralista pues contradice los compromisos pragmatistas del pluralismo generando inconsistencias y contradicciones.

## II

### El problema

- 2.1. El presente trabajo tiene un objetivo doble, el primero *crítico* y el segundo *propositivo*.
- (a) En su aspecto *crítico* el objetivo es someter a revisión las tesis b) y c). En otras palabras, demostrar que el pluralismo, para ser una propuesta congruente (que pueda dar cuenta de nociones como corrección e incorrección, objetividad, normatividad, racionalidad y verdad), no requiere de un realismo mínimo. A diferencia de un *pluralismo tradicional* sustentado sobre el paradigma kantiano en el que el sujeto y el objeto se anteponen, un *pluralismo genuinamente pragmatista* nos coloca en línea recta hacia una *auténtica multiplicidad* al margen de cualquier dicotomía, dualismo y presupuestos metafísicos.
  - (b) En su aspecto *propositivo* el objetivo es postular una noción de verdad plenamente consistente con un *pluralismo genuinamente pragmatista* o *pluralismo radical*. Esto permitirá demostrar que, aun en su acepción más radical y a pesar de abandonar cualquier requerimiento mínimo realista, el pluralismo todavía puede dar una respuesta al problema del *relativismo* y la verdad.
- 2.2. Desarrollo de (a).

- 2.2.1. Para los detractores c) es un requisito indispensable. Ya desde 1909, William James se había percatado del recelo que provocan las tesis relativistas del pluralismo, pues los críticos del pluralismo reclaman la idea de que la unidad es un requerimiento indispensable para la racionalidad, en tanto es una idea necesaria para la justificación y el establecimiento de los límites del conocimiento, pues en la multiplicidad no puede haber límites, normas o conocimiento significativo.
- 2.2.2. Un caso ejemplar es el Realismo Interno (RI) de Hilary Putnam. El propósito del RI es:
- A. [...] romper el collar de fuerza que cierto número de dicotomías ponen sobre nuestro pensamiento, destacando entre ellas la dicotomía entre las perspectivas “objetivas” y “subjetivas” de la verdad y la razón (Putnam, 1987).

A pesar de aceptar que el RI tiene una deuda de origen con el idealismo trascendental de Kant, Putnam se desmarca de la noción de cosa en sí para evitar la distinción tradicional entre “propiedades intrínsecas” y “propiedades extrínsecas”, y cumplir con A:

- B. el RI dice que no sabemos de qué estamos hablando cuando hablamos sobre las “cosas en sí”. [...] la cosa en sí y la propiedad que la cosa tiene en sí misma pertenecen al mismo círculo de ideas, y ya es el momento de admitir que lo que el círculo abarca es un territorio sin valor (Putnam, 1987).

Para Putnam la noción de cosa en sí, como requerimiento para poner un límite a nuestras representaciones, es lo mismo que sostener la existencia de propiedades intrínsecas en los objetos. Ambas ideas pertenecen, como lo dice en B, a un territorio sin valor.

Por eso, Putnam apunta:

- C. The way to read Kant is as saying that what Locke said about secondary qualities is true of all qualities—the simple ones, the primary ones, the secondary ones alike (Putnam, 1981: 61).

Lo que Putnam está diciendo en C es que para el RI incluso aquellas cualidades que juzgamos como puramente objetivas y pertenecientes a los objetos mismos, *dependen enteramente* de nuestra percepción subjetiva. En consecuencia, ya no es posible trazar la siguiente distinción:

- D. entre lo que es una proyección y lo que es una propiedad independiente y unitaria de las cosas en sí mismas. Pienso que, al menos epistémicamente, el intento de dibujar esta distinción, de hacer este corte, ha sido un fracaso total. Ha llegado el momento de ensayar la hipótesis metodológica de que no puede hacerse ningún corte como ese (Putnam, 1987: 75).

Por tanto, no podemos atribuir a los objetos, bajo ningún motivo, cualidades intrínsecas que sean independientes de nuestros EC, mente o lenguaje:

- E. In a internalist view also, signs do not intrinsically correspond to objects, independently of how those signs are employed and by whom [...] Objects do not exist indepently of conceptual schemes. We cut

up the world into objects when we introduce one or another scheme of description (Putnam, 1981: 61).

A pesar de lo sostenido en A, B, C, D y E Putnam todavía piensa que no todo depende de nuestros EC, ya que hay cosas que se dan sin nuestra intervención:

F. ‘objects’ themselves are as much made as discovered, as much products of our conceptual invention as of the ‘objective’ factor in experience, the factor independent of our will, then of course objects intrinsically belong under certain labels (Putnam, 1981).

Putnam afirma que el mundo depende *fuertemente* de nuestros Esquemas Conceptuales (EC) y que no puede establecerse una distinción entre las propiedades que los objetos tienen en sí mismos y lo que nosotros proyectamos. Pero a la vez sostiene que este mismo mundo no existiría sin un sustrato independiente que otorgue la materia con la cual llenar esos EC. El problema es grave en tanto que Putnam, para desmarcarse del *relativismo* tiene que decir algo acerca de ese sustrato que dota de contenido nuestros EC, estableciendo un corte entre lo que es una proyección y una propiedad independiente, siendo inconsistente con aquello que había sostenido en A, B, C, D y E.

Guillermo Hurtado tiene muy clara esta disonancia presente en el RI y señala que aunque la objeción pueda parecer débil apunta a un problema profundo. “El problema es cómo podemos aceptar que algunas cosas suceden independientemente de nosotros, sin aceptar, a la vez, que estas cosas dependan –hasta cierto punto– de la realidad independiente” (Hurtado, 1998, p. 42). La propuesta de Putnam es inconsistente desde el momento en que postula la necesidad de un requerimiento metafísico mínimo.

2.2.3 El caso de Putnam permite afirmar lo siguiente: el pluralismo, para ser una propuesta congruente no requiere de un realismo mínimo como propone c) o como sugiere b). Este requerimiento mínimo está en flagrante contradicción con los presupuestos pragmatistas fundamentales del pluralismo.

2.3. Presentación de un genuino realismo pragmatista o pluralismo radical.

2.3.1. Para evitar las inconsistencias anteriores es necesario plantear un pluralismo pragmatista radicalizado que nos coloque en línea recta hacia una autentica multiplicidad al margen de cualquier dicotomía, dualismo o presupuesto metafísico. La expresión de este “pluralismo pragmatista radical” la podemos encontrar claramente en la propuesta pluralista de Nelson Goodman.

2.3.2. ¿Qué principios debe satisfacer un genuino pluralismo pragmatista?

I. El rechazo a cualquier idea de que exista algo “dado” epistemológicamente que sirva de fundamento sólido al conocimiento empírico. Y en su vertiente ontológica, el rechazo de cualquier presupuesto metafísico-monista en el que se admita c).

II. La renuncia a cualquier tipo de dualismo que implique una imagen dicotómica del mundo, p. ej., la distinción entre mente y

cuerpo o la imagen dualista del mundo y sus cualidades primarias, por un lado, y de la mente y sus datos de los sentidos, por otro.

2.3.3. Estos requerimientos se derivan de las formulaciones clásicas del pragmatismo de Peirce y James respecto a la experiencia y la objetividad.

2.3.4. El pluralismo de Goodman satisface los requerimientos fundamentales de una genuina multiplicidad pragmatista pues cumple con I. en tanto sostiene que el requerimiento de una estructura única, permanente y neutral no es ni suficiente ni necesario para explicar los problemas de la epistemología o distinguir entre las versiones de mundo verdaderas de las falsas. Esta estructura es autorrefutable y contradictoria pues implica adoptar cierta postura externalista:

Talk of unstructured content or an unconceptualized given or a substratum without properties is self-defeating; for the talk imposes structure, conceptualizes, ascribes properties (Goodman, 1978).

2.3.5. Al negar la relevancia de ‘lo dado’, Goodman cumple también con II.:

The realist will resist the conclusion that there is no world; the idealist will resist the conclusion that all conflicting versions describe different worlds. As for me, I find these views equally delightful and equally deplorable—for after all, the difference between them is purely conventional [...] In practice, of course, we draw the line wherever we like, and change it as often as suits our purposes (Goodman, 1978).

2.3.6. Goodman sostiene, en un sentido fuerte, que las distintas cosas — materia, energía, ondas, fenómenos— de la que están hechos los mundos surgen junto con los mundos, se hacen con las versiones. Así, percibir el movimiento consiste en producirlo, descubrir leyes implica trazarlas, reconocer patrones es un asunto de inventarlos e imponerlos. Comprensión y creación van de la mano, ocurren en un mismo momento, son indisociables. (Ibíd.)

2.3.7. El establecimiento de las categorías, enunciados proyectables y clases relevantes dentro de una versión de mundo es lo que Goodman llama: *corrección* categorial. Para llevar a cabo la *corrección* categorial se implican una serie de relaciones vinculadas con versiones de mundo previas. La creación de mundos no es un proceso *ex nihilo*, pues los mundos se crean a partir de otros mundos que tenemos a la mano y que asumimos como verdaderos o correctos (ibíd). *Hacer es un rehacer* (ibíd.).

### III

## El problema de la verdad

3. Desarrollo de (b).
  - 3.1. El cuestionamiento es: ¿Qué noción de verdad es la más adecuada para un pluralismo que ha renunciado a cualquier supuesto metafísico y dualista?
  - 3.2. Respuesta: una noción pragmatista en la que la verdad se asuma como un predicado que puede ser aplicado de manera correcta o incorrecta a los enunciados con valor de verdad. El criterio de esta corrección o incorrección dependerá de los distintos usos del lenguaje en contextos concretos. Para atribuir la verdad a un enunciado dentro de un lenguaje específico se deben seguir normas, criterios o reglas pragmáticamente establecidas y justificadas dentro de un contexto.
  - 3.3. Los supuestos:
    - 1) La verdad no es una propiedad.
    - 2) Se conservan las condiciones de verdad pero en un sentido pragmatista.
    - 3) Afirmar que “p” es verdadero consiste en seguir una regla.
    - 4) Significado = uso.
    - 5) Equilibrio reflexivo.
  - 3.4. Características de la verdad:
    - Relatividad
    - Procesualidad
    - Performatividad
    - Vitalidad
  - 3.5. ¿Por qué ocuparse todavía de la verdad?
    - La verdad nos permite evaluar las creencias que tenemos del mundo.
    - La noción de verdad está íntimamente ligada a la noción pragmatista de *creencia como disposición*.
    - La noción de verdad es una intuición básica para la vida humana pues está íntimamente ligada a la normatividad.
    - Porque existe una asociación errónea de la verdad con la necesidad y suficiencia.
  - 3.6. Verdad en sentido pragmático: En tanto es iniciado dentro de las prácticas lingüísticas de una comunidad el hablante va aprendiendo las reglas convencionales para predicar la verdad de un enunciado en el lenguaje de su comunidad. Esto implica que si un agente ha aprendido a manejar un lenguaje ha aprendido también las reglas que le permiten establecer algo como verdadero.

## Bibliografía

- Baker, G.P. & Hacker, P.M. (1988). *Wittgenstein. Rules, Grammar and Necessity*, Volume 2 of an Analytical Commentary on the Philosophical Investigations, London, Blackwell.
- Barnes, B. (2001). *Practice as Collective Action en The Practice Turn in Contemporary Theory*, edit. Schatzki, T.; Knorr K.; Von Savigny, E., London, Routledge.
- Bernstein, R., (2010). *The Pragmatic Turn*, traducción de Pablo Lazo Briones, *El giro pragmático*, Ciudad de México, UAM.
- Bloor, D. (2001). Wittgenstein and the Priority of Practice, en *The Practice Turn in Contemporary Theory*, edit. Schatzki, T.; Knorr K.; Von Savigny, E., London, Routledge.
- Burgess, Alexis G. & John P. (1980). *Truth*, New Jersey, Princenton University Press.
- Ellenbogen S., (2003). *Wittgenstein's Account of Truth*, State University of New York Press.
- Fish, S. (2003). Truth but No Consequences; Why Philosophy Doesn't Matter, en *Critical Inquiry*, Vol. 29, No. 3 (Spring 2003), pp. 389–417.
- Foster J. A. (2010). Meaning and truth theory, en Gareth Evans y John McDowell (eds) *Truth and Meaning. Essays on Semantics*, Claredon Press, Oxford.
- Goodman N. (1953). The New Riddle of Induction, en *Fact, Fiction and Forecast*, Harvard University Press.
- Hurtado, G. (1998). Realismo, relativismo e irrealismo, México D.F., en *Crítica* Vol. XXX, No. 90.
- James, W. (1904). Does 'Consciousness' Exist?, en *Journal of Philosophy, Psychology, and Scientific Methods*, 1.
- James, W. (1907). *Pragmatism: A New Name for Some Old Ways of Thinking*, Dover Publications, New York.
- Kripke, S. (1982). *Wittgenstein on Rules and Private Language*, Blackwell Publishing, Oxford. Trad. Rodríguez Marqueze, Jorge, *A propósito de reglas y lenguaje privado*, Tecnos, 2006.
- L. Kirkham R. (1992). *Theories of Truth. A Critical Introduction*, Massachusetts, MIT Press.
- Lombardi, O. & Ransanz, A. R. (2012). *Los múltiples mundos de la ciencia. Un realismo pluralista y su aplicación en la filosofía de la ciencia*, México D.F., UNAM-Siglo XXI.
- Lynch, M. (2004). *Why Truth Matters*, Massachusetts, MIT.

- Lynch, M. (1998). *Truth in context: An essay on pluralism and objectivity*, Massachusetts, MITpress.
- Macbeth, D. (2007). Pragmatism and Objective Truth, en *New Pragmatists* Comp. Misak, Cheryl, Oxford.
- Misak, Ch. (1991). *Truth and the End of Inquiry: A Peircian Account of Truth*, Clarendon Press, Oxford.
- Misak, Ch. (2007). Pragmatism and Deflacionism, en *New Pragmatists* Comp. Misak, Cheryl, Oxford.
- Peirce, C. (1878). How to Make Our Ideas Clear, en *The Essential Writings* (1972), Prometheus Books, New York.
- Putnam H. (1981). *Reason, Truth and History*, Cambridge University Press.
- Putnam H. (1987). *The Many Faces of Realism*, Open Court, Chicago.
- Putnam H. (1992). *Realism with a Human Face*, Harvard University Press, Edited by James Conant.
- Rorty, R. (1991). Representation, Social Practise, and Truth, en *Objectivity, Relativism and Truth*, Cambridge University Press, Cambridge.
- Rorty, R. (1983). *La filosofía y el espejo de la naturaleza*, Madrid, Cátedra.
- Ruphy, S. (2013). *Scientific Pluralism Reconsidered*, Pittsburg, University of Pittsburg Press.
- Wittgenstein L. (1933-1935). *The Blue and Brown Books*, Trad., Francisco Gracia Guillén, Tecnos.
- Wittgenstein L. (1922). *Tractatus Logico-philosophicus*, Madrid, Tecnos.
- Wittgenstein L. (1953). *Philosophical Investigations*, Oxford, Blackwell Publishing.
- Wittgenstein L. (1969). *On Certainty*, Oxford, Blackwell Publishing.



# The experience and processing of temporal features of perception

Olga Fernández Prat

Daniel Quesada

Universitat Autònoma de Barcelona

## Grush early proposal to bridge Husserlian phenomenology and cognitive neuroscience

In 2006 Rick Grush cogently criticized several ways of establishing a “bridge” between cognitive neuroscience and Husserlian phenomenology of time consciousness before advancing his own proposal about how to establish such a “bridge”. More recently (2016), he has substantially modified his position, although without altogether giving up his view of the relevance of a Husserlian-like theory.

His proposals, now and then, are based on his Trajectory Estimation Model (TEM), a computational model put forward to explain temporal features of perception, including well-known temporal illusions (like, e.g., those in the well known “cutaneous rabbit” illusion). Here, we are not going to put to critical scrutiny Grush’s claim that TEM, or something close to it, provides the best available explanation of those illusions, something for which Grush still argues forcefully in his more recent work, in which he also advances a “hybrid view” that includes TEM as a component. We will focus exclusively on the relation Grush claims there to be between TEM (or the TEM component of a model answering to such “hybrid view”), on the one hand, and a Husserlian philosophical analysis of temporal experience, on the other. Or, as it will turn out, between TEM and other available analysis of the temporal aspects of perceptual experiences.

TEM is one among a host of models that implement the *emulative hypothesis*, an hypothesis that holds that our nervous system systematically emulates a host of processes on the basis of previous experience. Moreover, under certain conditions, the sensory information registered by our perceptual system is filtered out on the basis that what the emulation system produces is deemed more probable, than such an information. In these cases, what we consciously seem to perceive is what the emulation system “says” has occurred, instead of what really took place. This happens under especially demanding temporal conditions; when those conditions are relaxed, the effect disappears, and we are able to perceive the phenomenon veridically. In the particular case of TEM, the model incorporates the idea that the filtering out at issue can produce representations of the states of a domain at times other than the present. That is, knowledge, at time  $t$ , of how the system behaves at that same time, can be used to yield estimates of the

state of the domain at time  $t + 1$  and at time  $t - 1$ , i.e., respectively, a prediction and a retrodiction. Since the process can be iterated, according to the model, at any time  $t$ , a representation of the states of the domain through a period of time, say,  $t - l$  through  $t + k$  can be produced. Moreover, the estimates themselves need not be limited to a particular time, but can be made through a period of time, e.g., the estimates of states of the domain from  $t - l$  through  $t + k$ , produced over the period from  $t - m$  through  $t + n$  (where  $l$ ,  $k$ ,  $m$  and  $n$  can all be different).

The proposal in Grush (2006) was that TEM is a good candidate for formulating a theory of the processing of temporal factors in experience underwriting a fundamentally Husserlian analysis of temporal experience, with its well known tripartite structure of retention, primal impression and protention.<sup>1</sup>

Now, Grush himself anticipated an objection one could address to such a claim. This is that the phenomena for which TEM provides an account—illusions involving temporal factors—happen over very short intervals (of up to approximately 100 ms, or 200 ms at most). For temporal intervals longer than those, simply there are no temporal illusions of the sort the trajectory theory sets out to explain. In contrast, “Husserlian protention and retention do not seem to be limited to intervals” of such a temporal order; e.g., “Husserl’s examples of melodies clearly indicate that he had ranges *at least* on the order of seconds in mind ... a few hundred milliseconds seems far short of the magnitude required” (Grush, 2006, p. 447).

Grush rejoinder was that the situation indicated “that there are at least two phenomena here, two different kinds of retentional-protentional structure in play, and Husserl focuses on the one active over longer durations, and the trajectory estimation model is an attempt to explain the one active over shorter durations” (*loc. cit.*, p. 447).

The problem is that no evidence was given that there is really a “retentional-protentional structure ... active over longer durations”. It certainly does not seem that we had even a hint of a neuroscientific theory at the time, or at present. In any case, Grush’s did not provide one (on his own admission).

---

<sup>1</sup>The aspects of perceptual experience that Grush’s took Husserl’s analysis to have accurately characterized—the alleged “tripartite structure” of temporal consciousness—are these (cf. Grush 2006, pp. 420–422):

- (i) One aspect concerns that which is a new element in the consciousness of an event. This aspect is called the *primal impression*, sometimes referred as “consciousness of the *now*”.
- (ii) Another aspect concerns what is just past, or has come just before. This is called *retention*.
- (iii) Finally, there is an aspect that concerns what is just about to come or to happen; a sort of implicit anticipation called *protention*.

## A new proposal: A-ish character, B-ish character and scale variance

In his recent work (Grush 2016), Grush has elevated part of the position expressed in the mentioned rejoinder to a general hypothesis about the temporal aspects of experience, while importantly modifying the rest of it. The general hypothesis is that time experience is not scale invariant. The important modification lies in that it is no longer part of his view that temporal experience *at very short intervals* has something like Husserl's tripartite structure.

In effect, Grush now distinguishes between experience which has an "A-ish" character and experience with a "B-ish" character. He takes it that the distinction is simply an assumption he thinks "most of the discussion" on these matters makes concerning the temporal character of perceptual experience, namely that "the relevant temporal content" of perceptual experience "is structured in terms of [either] (i) past, present, future, or (ii) earlier than, simultaneous with, later than" (cf. Grush 2016, p. 7). The respective terms Grush uses for alternatives (i) and (ii) are, of course, meant to echo McTaggart's celebrated metaphysical distinction between the "A-series" and "B-series".<sup>2</sup>

The specific hypothesis Grush now makes about experience at very short intervals is that such experience has a B-ish character. His argument for the latter conclusion seems to proceed along the following lines: TEM, in providing the right explanation for temporal illusions, proves that the processing machinery it postulates underwrite our temporal experience at such a scale, and therefore, reveals features of this experience. In particular, it would reveal that, at such a scale, we do not have an "experience of the now". This would indicate that experience at such a scale has not an A-ish character; hence, it has a B-ish character.

In any case, and on the basis of the specific hypothesis, Grush now makes clear that experience at such an scale *cannot* possibly be as described by a Husserlian theory, since, in postulating a primal impression phase, such theory would in effect be allowing that we have "experience of the now" (cf. *op. cit.*, p. 8). This is a very relevant correction of his former view: at such an scale, at least, there is no "bridging" of "cognitive neuroscience and Husserlian phenomenology of time consciousness". Only a vague kinship would remain, to the extent to which one can see in TEM the presence of a retentive element.

Some hope of establishing a "bridge" with at least something closer to the Husserlian theory remains in his other special hypothesis that "at a scale greater than 200msec, experience presents temporal relations A-ishly, and at that scale, a more standard retentive picture applies" (*op. cit.*, p. 11). But this possibility—the one that would retain a substantial part of his original stance—is not further explored in the recent paper.

The bulk of the paper is devoted to the defense of TEM against rival explana-

---

<sup>2</sup>The claim that experience—or a certain type of experience—has an A-ish character, or that it has a B-ish character can be articulated in a number of ways. Grush (2016) is content with a rough, intuitive rendering, and we are following this here. For good recent representatives of the two views see, respectively, Crowther and Soteriou (2017) and Hoerl (2009).

tions of the temporal illusions. These explanations come from diverse versions of extensionalism, the main rival to a retentionalist approach (cf. Dainton 2017a, Hoerl 2009, and Phillips 2014). Reasons are given for seeing extensionalism in its diverse forms as postulating an A-ish structuring of the contents of temporal experiences—at whatever scale—which are not completely clear to us (cf. *op. cit.*, p. 8). Anyway, at the specific level of defending TEM against the alternative explanations presented, Grush’s discussion is detailed and interesting. He might be right that TEM provides a better explanation of temporal illusions than those alternatives. But what this eventually implies for which putative neuroscientific theory best describes the processing which underwrites which philosophical theory of the temporal character of experience (if any) is far from clear.

## A misconceived project

Grush’s general reasoning concerning temporal experience at around 2006 was as follows (cf. Grush 2006, p. 441):

- (i) “[F]eatures of the neural information processing machinery in the central nervous system are relevant to those representational structures that underwrite the temporal aspects of our conscious experience.”
- (ii) “[T]o a first approximation at least, Husserl’s analysis does accurately characterize certain aspects of our subjective experience.”

What follows from these two premises is that there is something about the mechanisms of neural information processing which *underwrite* experience that has the “features of phenomenology revealed by Husserl’s analysis” (*loc. cit.*).

The aim of uncovering the information-processing structures that “underwrite” features of temporal aspects of perception described by a philosophical theory may be a perfectly sound project. At least, if we understand the word “underwrite” as summing up what Martin Davies described as “the need to descend”, that is, the need for philosophical theories of cognitive relevance, of finding support at some point in a cognitive or neuroscientific theory for their validation (that of the philosophical theory, that is; cf. Davies 2000, Fernández Prat 2008). And we think that the legitimacy of such aim can be extended to theories which make relevant claims about phenomenology. It can be checked that our objections and qualms in the previous section concern the particular project made explicit there, not the general aim of finding a scientific theory that can be seen as providing such sort of support.

Grush’s ambition, however, seems to be set for a higher aim. As we will presently see, this was explicit in his former paper, and there is no indication of being an ambition lost in the recent one. In this case, we do think that the aim itself is open to criticism, and that a project that pursues such an aim is misconceived. The reasons for this are connected with the general issue of how we should think of the relations between philosophical theories of (some aspects of) cognition or of consciousness and information-processing theories, neurophysiological or otherwise, and so it may be important to look closely at the issue.

Grush more ambitious aim concerns the uncovering of the information-structures that *explain* (not those which merely *underwrite*) the features of temporal experience allegedly described by the philosophical theory.<sup>3</sup> Indeed, we read in Grush (2006):

I am committed to there being something about the mechanisms [of] neural *information processing* that explains why our phenomenal experience [has] those features of phenomenology revealed by Husserl's analysis (p. 441; emphasis in the original).

Grush arrived to this conclusion by adding a new premise to the two that were quoted in the previous section, namely:

I also believe that our phenomenal experience is largely a function of the representational structures produced by neural information processing machinery, (*loc. cit.*)

After the previous section, we know that Grush no longer wants to be committed to “those features of phenomenology revealed by Husserl's analysis”. Indeed, for the case of very short intervals, he openly rejects them. His new general commitment is for the philosophical analysis of the temporal character of experience revealing either B-ish features or A-ish features (Grush 2016, p. 8). Hence, Grush's line of reasoning can be formulated thus:

(GR) Reflection on the temporal character of perceptual experience reveals that it is either A-ish or B-ish. But our phenomenal experience is a function of the representational structures produced by neural information processing machinery. It follows that there is something about the mechanisms of neural information processing that explains why our phenomenal experience has those features.

It will have been noticed that, in this formulation, we have omitted the first premise in the argument mentioned in the previous section, since it is no longer relevant here. Further, the qualification “largely”, which appears in Grush's original formulation of the conclusion, has also been removed. This is because its inclusion makes the claim vague, hence, weak for the wrong reason. Indeed, one would not very well know what would be claimed by saying that phenomenal experience is “largely a function” of certain representational structures produced by certain neural information processing machinery. It might suggest that other elements may also be intervening to go from the information-processing mechanisms to phenomenal experience. This would, indeed, open new, unknown perspectives. But it is not, we believe, relevant to the present discussion. What matters here is whether, maintaining fixed these other hypothetical elements, the activation of the same (neural) information-processing mechanisms would produce the same phenomenal experiences. One could then not be denying this and maintain at the same time that the relation is a functional one. One would have to settle for the weaker term “relation”. It seems that Grush wants to make a more significant—more strong—claim, and this is made clearer, without losing relevance, by the formulation above.

---

<sup>3</sup>This more ambitious aim seems to be a general one through Grush's work on information-processing models of consciousness. Grush does not explicitly differentiate between a less ambitious aim and a more ambitious one. Indeed, he might even reject the difference. In any case, the fact that it remains unmarked in his writings helps its being overlooked.

Although we have formulated GR following closely on Grush's steps, it seems to us that such line of reasoning informs at present much work in cognitive science and neuroscience. So our claim concerns the widely shared ambitious project which derives from that line of reasoning.

What we think is wrong about such a project is that it assumes that it is not possible that the same neuronal structures carrying fundamentally the same information—the information which is behaviourally relevant—have obtained such an information through essentially different channels, and that just this difference is crucial for the having of experiences with different phenomenal characters.<sup>4</sup> Hence, such phenomenal characters are not necessarily a *function* of the representational structures produced by neural information processing machinery. In consequence, pursuing the strategy of finding the information processing machinery of which key features of temporal experience are supposedly a function, is a misconceived project.

The general inescapable conclusion is that, if phenomenal consciousness is to be associated to neural activity, it must be done in a less conceptually simple way than the one Grush—and with him a considerable number of cognitive science and neuroscience theorists—has been devising.

## References

- Crowther, T. and Soteriou, M. (2017). "Time in the Dream." In Phillips, I. (ed.), *The Routledge Handbook of Philosophy of Temporal Experience*. London: Routledge, pp. 184–200.
- Dainton, B. (2017). "Temporal Consciousness." *Stanford Encyclopedia of Philosophy*.
- Davies, M. (2000). "Persons and their Underpinnings." *Philosophical Explorations*, 3:43–62.
- Fernández Prat, O. (2008). "Filosofía y neurociencia: Entre la interacción y el eliminativismo." In Pérez, D. and Fernández Moreno, L. (eds.), *Cuestiones filosóficas: ensayos en honor de Eduardo Rabossi*. Buenos Aires: Contextos, pp. 285–306.
- Fernández Prat, O. and Quesada, D. (manuscript). "Philosophy and Cognitive Science on Spatial and Temporal Experience."
- Grush, R. (2005). "Internal Models and the Construction of Time: Generalizing From State Estimation to Trajectory Estimation to Address Temporal Features of Perception, Including Temporal Illusions." *Journal of Neural Engineering*, 2:S209–S218.
- Grush, R. (2006). "How to, and How Not to, Bridge Computational Cognitive Neuroscience and Husserlian Phenomenology of Time Consciousness." *Synthese*, 153:417–450.

---

<sup>4</sup>This is exactly what happens in the spatial case, where the same elements in the posterior parietal cortex encode (behavioral) content which may correspond to quite different phenomenal experiences. We have analysed this case in Fernández Prat and Quesada (manuscript).

- Grush, R. (2016). “On the Temporal Character of Temporal Experience, Its Scale Non-Invariance, and Its Small Scale Structure.” Manuscript. doi:10.21224/P4WC73.
- Hoerl, C. (2009). “Time and Tense in Perceptual Experience.” *Philosophers’ Imprint*, 9(12).
- Phillips, I. (2014). “Experience of and in Time.” *Philosophy Compass*, 9(2):131–144.

# Qué es racional creer sobre las leyes Humeanas a la luz de la objeción de la coincidencia cósmica

Aldo Filomeno

## Resumen

Hay una antigua y fuerte objeción a la teoría de leyes Humeanas según la cual, si las leyes de la naturaleza fueran tal como cree el Humeano, sería una coincidencia cósmica inexplicable que el mosaico Humeano actual sea tan extremadamente regular como resulta ser. Sin embargo, muchos filósofos han desatendido esta objeción, principalmente debido a que recurre implícitamente al principio de indiferencia de forma injustificada. Debido a esto uno debería suspender el juicio sobre la solidez de esta objeción. Sin embargo, en esta charla argumento que suspender el juicio sobre la solidez de un argumento no debe siempre llevarnos a desatenderlo. Más concretamente, argumento que bajo ciertas suposiciones, que se cumplen en este caso, dada una creencia inicial positiva en una hipótesis  $H$ , el venir a conocer una objeción crucial a  $H$ , cuyo valor de verdad ignoramos, no es información irrelevante: uno debe rechazar su creencia inicial y suspender el juicio sobre  $H$ . Por tanto, aplicado a nuestro caso, la creencia epistémicamente racional acerca de la teoría de las leyes Humeanas es, en el mejor de los casos, la suspensión del juicio.

## Palabras clave

Leyes Humeanas; Best System Account of laws; Suspensión de juicio; Agnosticismo; Propagación de la suspensión de juicio

## 1. Introducción

¿Qué deberíamos creer sobre la teoría humeana de leyes de la naturaleza?<sup>1</sup> Por supuesto, una pregunta como esta se supone que depende de muchos argumentos a favor y en contra, y no parece haber un veredicto claro. En esta charla sin embargo argumentaré que, dadas ciertas suposiciones sobre el estado actual del debate sobre las leyes Humeanas, un agente con una creencia inicial positiva en la teoría de leyes Humeana debería modificar su creencia inicial positiva a suspensión del juicio sobre dicha teoría.<sup>2</sup> Argumentaremos esto porque hay cierta

<sup>1</sup> Nos referimos a la teoría elaborada principalmente por Mill, Ramsey y Lewis, también conocida como 'Best System Account' (Mill, 1884; Ramsey, 1978; Lewis, 1973, 1994); véase también Cohen and Callender, 2009; Loewer, 2007; Schrenk, 2014; Woodward, 2014. Para una versión extendida y en inglés de este documento ver Filomeno (2018d).

<sup>2</sup> Usaremos los términos 'agnosticismo' y 'suspensión de juicio' de manera intercambiable.



información que ha sido desatendida que deberíamos tener en cuenta, información que nos lleva a suspender el juicio. Esta información es el conocimiento de una objeción crucial a la teoría Humeana, objeción sobre la cual suspendemos el juicio. Por ‘crucial’ quiero decir que si la objeción fuera cierta, nos llevaría a tener desconfiar (e.d. tener una creencia negativa) sobre la teoría de leyes Humeana.

Esta objeción, argumento, ha sido incorrectamente desatendida. Aunque esto es comprensible e intuitivo porque es lo que se debe hacer en muchos casos, es incorrecto desatender *siempre* este tipo de objeción: argumentaré aquí que objeciones cruciales cuyo valor de verdad ignoramos no siempre deben desatenderse. Bajo ciertas condiciones, ignorar el valor de verdad de la objeción crucial implica, dado que es crucial, que el valor de verdad de la tesis que la objeción ataca – en nuestro caso la teoría de las leyes Humeana – es también algo sobre lo cual tenemos que ser agnósticos. En otras palabras, bajo ciertas condiciones, la suspensión del juicio sobre una objeción crucial “se propaga” a la tesis que la objeción ataca.<sup>3</sup>

## 2. La objeción de la coincidencia cósmica a las leyes Humeanas

No obstante sus virtudes y sofisticaciones recientes, desde sus orígenes la teoría de leyes Humeana se ha enfrentado a una gran amenaza. De acuerdo con esta teoría, las leyes de la naturaleza son meras *descripciones*: las mejores descripciones de toda la historia del universo, es decir, del mosaico Humeano, por lo cual las regularidades no accidentales son sólo aparentemente no accidentales. Mientras que esto significa que nuestra ontología se libera de la misteriosa noción de necesidad física, el problema es que también significa que el comportamiento regular aparentemente no accidental y extremadamente estable *ubicuo* en el universo es meramente un hecho bruto sin explicación, una *coincidencia cósmica*. El mosaico Humeano actual es un hecho bruto, por lo cual todas las regularidades dentro de él son un hecho bruto también. Sin embargo, resulta llamativo que estas regularidades son extraordinariamente ubicuas: nuestra experiencia y nuestra imagen científica describen un mundo donde el comportamiento regular es extremadamente ubicuo. Todos los planetas del sistema solar orbitan regularmente el Sol, todas las estrellas en todas las galaxias tienen los mismos ciclos de vida, todos los fermiones en el universo siguen las mismas interacciones fundamentales; en general, no hay evidencia empírica que las leyes de la física hayan cambiado *nunca* su forma en todo el espacio y el tiempo. Foster (1983, 89) ilustra la objeción de la coincidencia cósmica con el ejemplo de las interacciones gravitacionales:

The past consistency of gravitational behavior calls for some explanation.  
For given the infinite variety of ways in which bodies might have behaved  
non-gravitationally and, more importantly, the innumerable occasions on

---

<sup>3</sup> Este fenómeno de la propagación de la suspensión de juicio ocurre más generalmente, siempre que las condiciones se cumplan. Véase (Filomeno, 2018c) para un estudio general de las condiciones que deben cumplirse en el intrincado contexto de teorías científicas y experimentos cruciales.



FIGURE 30. In order to produce a universe resembling the one in which we live, the Creator would have to aim for an absurdly tiny volume of the phase space of possible universes —about  $1/10^{10^{22}}$  of the entire volume, for the situation under consideration. (The pin and the spot aimed for are not drawn to scale!)

Figura 1: El lienzo cúbico contiene todos los universos posibles, y cada punto representa un posible universo. Es un hecho matemático de la combinatoria que casi todos ellos están desordenados (e. d., son caóticos). El conjunto de universos ordenados se representa como la mancha negra sumamente pequeña a la que Dios está señalando. Imagen de Penrose (1989) (el cual la usó para otro debate).

which some form of non-gravitational behavior might have occurred and been detected, the consistency would be an astonishing coincidence if it were merely accidental – so astonishing as to make the accident-hypothesis quite literally incredible.

Así pues, si adoptamos la teoría de leyes Humeana, es una coincidencia cósmica que tales regularidades extremadamente estables hayan ocurrido, pues nuestro mosaico Humeano plagado de patrones parece una posibilidad extremadamente improbable entre todos los posibles mosaicos irregulares. En resumen, una explicación de este abrumadoramente alto nivel de regularidad debería proporcionarse; de lo contrario, la teoría de leyes Humeana ni siquiera puede considerarse una teoría candidata decente de las leyes de la naturaleza. La figura 1 ilustra esta situación (cf. Swartz, 2018, §7).

Este es el argumento, un argumento aparentemente plausible y potencialmente devastador para la teoría Humeana. Sin embargo, los humeanos a día de hoy se sienten tranquilos: el argumento oculta una premisa implícitamente falsa. Es válido, pero no sólido. En particular, el argumento se basa en la suposición implícita de que a cada posibilidad debe asignársele igual probabilidad; que todas las maneras en que el mundo podría haber sido – cada mosaico humeano – son igualmente probables. Esta premisa es una instancia del principio de indiferencia, que puede ser formulado por ejemplo así:

Supóngase que hay  $n$  posibilidades mutuamente exclusivas y colectivamente exhaustivas. Si no hay evidencia favoreciendo una posibilidad sobre otras, entonces a cada posibilidad debe asignársele la misma probabilidad. (Hájek, 2012, §3.1)

Por más intuitivamente correcto que parezca, el principio de indiferencia está hoy día seriamente desacreditado. Éste es la razón principal por la cual la objeción de la coincidencia cósmica se supone que es incorrecta.

### 3. Cuándo suspender el juicio

De hecho, en lugar de seguir el principio de indiferencia, modelos formales recientes representan nuestro estado doxástico de ignorancia total con el estado de suspensión del juicio. Dichos modelos han sido aplicados a argumentos en cosmología, donde se concluye que ciertas inferencias probabilísticas no están justificadas (Norton, 2010, Benétreau-Dupin, 2015).

Tal como Norton (2007, 2008, 2010, Forthcoming) ha instado recientemente, en vez de escoger *a priori* la lógica inductiva con la que representar nuestras creencias, deberían ser las condiciones empíricas o ‘materiales’ del problema bajo análisis las que tienen que justificar la lógica inductiva apropiada. Si bien es cierto que el típico marco probabilista – una asignación numérica precisa de probabilidades – está justificada en muchos casos, esto no siempre es así, y usar una lógica inductiva incorrecta puede llevarnos a errores en las predicciones.

Suspender el juicio sobre alguna proposición o conjunto de proposiciones significa no creer ni descreer en grado alguno ninguna opción u opciones, sino carecer de cualquier creencia o grado de creencia. Un número de representaciones formales se han elaborado en lógica inductiva para representar adecuadamente la suspensión de juicio. Estas diferentes representaciones incluyen por ejemplo el cálculo no-probabilista y no-numérico como el propuesto por Norton (2008) (aplicado a cuestiones cosmológicas en Norton 2010), y las probabilidades imprecisas propuestas por de Cooman and Miranda (2007) (aplicado a cuestiones cosmológicas en Benétreau-Dupin 2015). Para reseñas de otros acercamientos, ver Halpern (2003, Ch. 2) y Dubois (2007). Para discusiones recientes sobre el concepto de suspensión de juicio, ver Friedman (2013, 2015), Tang (2015) y (Filomeno, 2018c,d). Los orígenes de dicho concepto se remontan al escepticismo clásico de Pirrón (Empiricus, I c. A.C.).<sup>4</sup>

### 4. Suspender el juicio sobre la objeción de la coincidencia cósmica

Llamemos a la objeción de la coincidencia cósmica ‘OBJECCIÓN’. ¿Qué debería creer un agente racional ideal sobre OBJECCIÓN?

Una primera opción sería que OBJECCIÓN no fuera sólida porque una de sus premisas fuera falsa. Las principales críticas a OBJECCIÓN se puede decir que son:

- (i) OBJECCIÓN ignora, erróneamente, la evidencia que tenemos a día de hoy respecto al grado de orden del mundo real;
- (ii) OBJECCIÓN considera, erróneamente, que hay muchísimos más mosaicos humeanos irregulares que regulares; y
- (iii) OBJECCIÓN asume, erróneamente, que los eventos sobre los cuales un agente racional ideal tiene un grado de creencia sumamente bajo demandan explicación.

Si cualquiera de estas críticas fuera correcta, no se podría mantener la conclusión de OBJECCIÓN de que el alto grado de regularidad del mosaico Humean actual

---

<sup>4</sup>Para una discusión sobre el axioma de aditividad en situaciones de ignorancia, ver (Filomeno, 2018a).

exige explicación; ni mantener siquiera que uno debe suspender el juicio sobre OBJECCIÓN; más bien, uno estaría justificado a rechazar OBJECCIÓN, ya que alguna de sus premisas sería falsa.

Estas potenciales objeciones, sin embargo, considero que son erróneas y fácilmente refutables. Así lo argumento en (Filomeno, 2018d) y más en detalle en (Filomeno, 2018b). Lo que sin embargo es problemático en OBJECCIÓN, de forma casi oficial por la comunidad de filósofos de la ciencia, es un uso implícito del controvertido principio de indiferencia.

El principal problema del principio de indiferencia es la falta de justificación en la asignación de la probabilidad uniforme entre las posibilidades – un problema bien conocido que se ha destacado, por ejemplo, en los fundamentos de la mecánica estadística (por ejemplo, respecto a la distribución de probabilidad de Maxwell-Boltzmann (Uffink, 2006, §4), o respecto a la “Hipótesis del pasado” (Callender, 2004)).

Lo mismo ocurre con OBJECCIÓN, ya que no tenemos absolutamente ninguna idea respecto a las probabilidades que se deberían asignar a cada mosaico Humeano, y asignar la misma probabilidad para cada mosaico resulta completamente injustificado. Es decir, en nuestra situación de ignorancia total deberíamos suspender el juicio sobre las probabilidades de los mosaicos Humeanos. En conclusión, no deberíamos asignar grado de creencia alguno, ni positivo ni negativo: lo que deberíamos es suspender el juicio.

## 5. La propagación de la suspensión de juicio

Llegados pues a este punto, ¿qué le sucede a OBJECCIÓN si suspendemos el juicio sobre la probabilidad de cada mosaico humeano? ¿Es el argumento inválido o no? Su validez, que hemos asumido, por supuesto no está afectada por los valores de verdad de las premisas. Sin embargo, su solidez obviamente sí se ve afectada: OBJECCIÓN no es un argumento sólido, ya que no todas sus premisas y conclusión son verdaderas. Es por esto que OBJECCIÓN ha sido desatendida pero, como argumentamos en lo que sigue, desatenderla es erróneo.

Obviamente es correcto desatender argumentos que no son sólidos porque una premisa es falsa. Menos obviamente, también es correcto desatender *algunos* argumentos que no son sólidos porque el valor de verdad de una premisa no es ni verdadero ni falso. Sin embargo, hay otros de estos últimos argumentos que en cambio no deberían ser desatendidos: un agente racional debería suspender el juicio *también* sobre las hipótesis que dependan de tales argumentos. En otras palabras, la suspensión de juicio sobre el argumento se “propaga” a la hipótesis que el argumento ataca. A modo de ilustración, veamos como esto pasa con OBJECCIÓN.

En la medida en que consideramos que la suspensión de juicio es el estado doxástico correcto con respecto a las probabilidades de los mosaicos Humeanos, no podemos mantener la conclusión de OBJECCIÓN de que el alto grado de regularidad del mosaico Humeano actual exige una explicación, ya que esta conclusión se basa en la premisa de que la probabilidad de que un mosaico altamente re-

gular ocurra es abrumadoramente más baja que la probabilidad de que ocurra un mosaico irregular. Sin embargo, y aquí está la clave, esto no implica que tal conclusión es falsa, sino que ignoramos si esta conclusión es verdadera o falsa. Esto es lo que significa suspender el juicio. Así pues, suspender el juicio sobre la conclusión de OBJECCIÓN significa que *la teoría de leyes Humeana podría o podría no estar enfrentándose a una seria amenaza*. No se sabe.

Habiendo remarcado este primer punto, ahora debemos hacer una observación necesaria respecto a la fuerza de la conclusión de OBJECCIÓN. Tenemos que aceptar que si la conclusión de OBJECCIÓN fuera cierta, habría que admitir que dicha teoría de leyes no es realmente capaz de dar cuenta del aparente comportamiento de tipo ley que parece abundar en el universo. Es decir, tenemos que considerar a OBJECCIÓN como a una objeción crucial a la teoría Humeana. Esto a su vez significa que si estuviéramos seguros de que su conclusión es verdadera, lo racional sería desconfiar de la teoría Humeana de leyes. Esta condición la asumimos porque es necesaria para que se produzca la “propagación” de la suspensión del juicio. Más concretamente, es necesario que la fuerza de la conclusión de OBJECCIÓN supere cualquier otra suma de argumentos en favor de la teoría Humeana.

Hay argumentos para justificar esta suposición que acabamos de hacer. Podemos suponer (y considero tal suposición plausible) que los argumentos existentes actuales a favor y en contra de la teoría Humeana no son en absoluto decisivos ni especialmente fuertes. Así, la situación con respecto a los argumentos a favor y en contra de la teoría de leyes Humeana se parecería a la de la figura 2: la objeción O sería la conclusión de OBJECCIÓN, y la hipótesis H sería la teoría de leyes Humeana.

Por supuesto, un Humeano que deseara preservar la creencia en la teoría Humeana podría negar que la situación descrita realmente se asemeje al balance real de argumentos a favor y en contra de la teoría Humeana. Aun así, considero que, aunque sea de forma aproximada, es una representación relativamente fiel del equilibrio real de argumentos en este debate.

Así pues, dada la situación epistémica descrita, nuestra creencia en la teoría Humeana depende crucialmente de nuestra creencia en la conclusión de OBJECCIÓN (es decir, en si el orden del mosaico Humeano requiere una explicación). En el caso en que la conclusión de OBJECCIÓN fuera verdad, no deberíamos creer más en la teoría de leyes Humeana (inclinaria la balanza contra H en la figura 2); en el caso de que la conclusión de OBJECCIÓN fuera falsa, preservaríamos la creencia inicial positiva con la que hemos empezado (el equilibrio dibujado en la figura 2 permanecería como está). Entonces, la pregunta clave es: ahora que ignoramos el valor de verdad de la conclusión de OBJECCIÓN, ¿qué deberíamos creer sobre la teoría Humeana de leyes?

Hemos argumentado arriba que deberíamos suspender el juicio sobre la conclusión de OBJECCIÓN, y esto significa, como también hemos explicado arriba, que la teoría Humeana podría estar o podría no estar encarando una seria amenaza. Por tanto, el único estado doxástico que estamos garantizados a mantener es, como mucho, el que *ignoramos si deberíamos desconfiar de la teoría Humeana o preservar nuestra creencia positiva*. Y como hemos definido arriba (sección 3), suspender el juicio se define como el estado doxástico que corresponde a *ignorar*

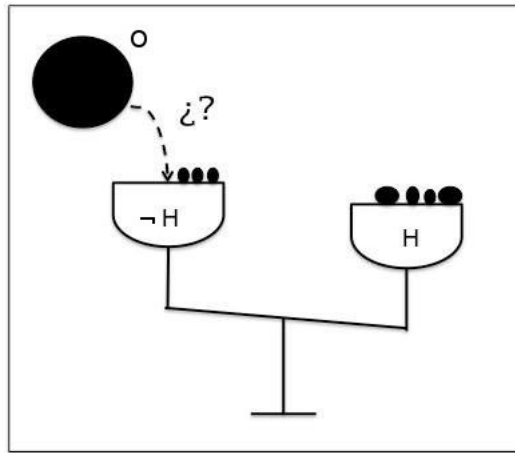


Figura 2: El escenario bajo estudio es tal que sabemos que hay una objeción crucial pero ignoramos si es verdadera o falsa, por lo que no sabemos si ponerla o no en la balanza.

*si p es el caso*. Por tanto, en esta situación, el estado doxástico epistémicamente racional es *la suspensión de juicio sobre la teoría Humeana de leyes*.

## Referencias

- Benétreau-Dupin, Yann. “Blurring Out Cosmic Puzzles.” *Philosophy of Science* 82, 5 (2015): 879–891.
- Callender, Craig. “Measures, Explanations and the Past: Should ‘Special’ Initial Conditions be Explained?” *British Journal for the Philosophy of Science* 55, 2 (2004): 195–217.
- Cohen, Jonathan, and Craig Callender. “A Better Best System Account of Lawhood.” *Philosophical Studies* 145, 1 (2009): 1–34.
- de Cooman, G., and E. Miranda. “Symmetry of models versus models of symmetry.” In *Probability and Inference: Essays in Honour of Henry E. Kyburg*, edited by W. Harper, and G. Wheeler. College Publications, 2007, 67–149.
- Dubois, Didier. “Uncertainty Theories: a Unified View.” CNRS Spring School “Belief Functions Theory and Applications”, 2007.
- Empiricus, Sextus. *Outlines of Pyrrhonism*. Harvard University Press, I c. A.C., 1994 edition.
- Filomeno, Aldo. “Agnosticism and the Additivity of Possibilities.”, 2018a. <https://philpeople.org/profiles/aldo-filomeno/publications>. Manuscript.
- . “Are non-accidental regularities a cosmic coincidence? Revisiting a central threat to Humean laws.”, 2018b. Manuscript.

- . “The propagation of suspension of judgment. Or, should we confer any weight to crucial objections whose truth-value we ignore?”, 2018c. Manuscript.
- . “Suspend judgment on the Humean account of laws.”, 2018d. Manuscript.
- Foster, John. “Induction, Explanation, and Natural Necessity.” *Proceedings of the Aristotelian Society* 83 (1983): 87–101.
- Friedman, Jane. “Rational Agnosticism and Degrees of Belief.” *Oxford Studies in Epistemology* 4 (2013): 57.
- . “Why Suspend Judging?” *Noûs* 50, 3.
- Halpern, Joseph Y. *Reasoning About Uncertainty*. MIT Press, 2003.
- Hájek, Alan. “Interpretations of Probability.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. 2012. Summer 2012 edition.
- Lewis, David K. *Counterfactuals*. Blackwell Publishers, 1973.
- . “Humean Supervenience Debugged.” *Mind*, 103 (1994): 473–490.
- Loewer, Barry. “Laws and Natural Properties.” *Philosophical Topics* 35, 1/2 (2007): 313–328.
- Mill, John Stuart. *A System of Logic*. Longman, 1884.
- Norton, John D. “Probability Disassembled.” *British Journal for the Philosophy of Science* 58, 2 (2007): 141–171.
- . “Ignorance and Indifference.” *Philosophy of Science* 75, 1 (2008): 45–68.
- . “Cosmic Confusions: Not Supporting Versus Supporting Not.” *Philosophy of Science* 77, 4 (2010): 501–523.
- . “Eternal Inflation: When Probabilities Fail.” *Synthese, special edition “Reasoning in Physics,”* .
- Penrose, Roger. *The Emperor’s New Mind: Concerning Computers, Minds, and the Laws of Physics*. New York, NY, USA: Oxford University Press, Inc., 1989.
- Ramsey, Frank P. *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*. International library of psychology, philosophy, and scientific method. Routledge & Kegan Paul, 1978.
- Schrenk, Markus. “Better Best Systems and the Issue of CP-Laws.” *Erkenntnis* .
- Swartz, Norman. “Laws of Nature.” In *Internet Encyclopedia of Philosophy*, edited by James Fieser, and Bradley Dowden. 2018.
- Tang, Weng Hong. “Reliabilism and the Suspension of Belief.” *Australasian Journal of Philosophy* 94, 2 (2015): 362–377.

Uffink, Jos. "Compendium of the foundations of classical statistical physics.", 2006. Chapter for "Handbook for Philosophy of Physics", J. Butterfield and J. Earman (eds) to appear.

Woodward, James. "Simplicity in the Best Systems Account of Laws of Nature." *British Journal for the Philosophy of Science* 65, 1 (2014): 91–123.



# Creación literaria y construcción de espacios de perspectivas

Natividad Garrido Rodríguez

Manuel Liz Gutiérrez<sup>1</sup>

Universidad de La Laguna

## Resumen

Un tema central del perspectivismo filosófico es la propuesta de modelos acerca de cómo la interacción de múltiples perspectivas, con sus respectivos contenidos, y posiblemente también en relación con una realidad que quede fuera de las perspectivas, es capaz de constituir una especie de “espacio” donde cada uno de estos elementos consiga tener una posición específica. Analizamos algunos aspectos importantes de tal noción de un espacio de perspectivas. Y mostramos como la creación literaria de una narración sirve como ejemplo sumamente ilustrador y sugerente del proceso de construcción de un espacio de perspectivas. Utilizaremos para ello tres autores sumamente representativos de las múltiples conexiones entre una realidad personal problemática y una creación literaria que explora de formas novedosas los recursos de la primera, la segunda y la tercera persona: Agustín Espinosa, Jean Genet y Hervé Guibert. Nuestra discusión permitirá, además, encontrar varios aspectos en los que el perspectivismo llega a distanciarse del relativismo.

## Comunicación

El perspectivismo es una posición sumamente recurrente dentro de la historia de la filosofía. Podemos encontrar dicha posición en autores como Protágoras, Leibniz, Kant, Nietzsche, Russell, James, Ortega, Carnap, Davidson, Putnam, etc.

A veces, el perspectivismo se sitúa muy cerca del relativismo. Pero hay diferencias cruciales cuando el perspectivismo se formula sólo en términos de “dependencia”. En este sentido, la tesis mínima del perspectivismo podría sintetizarse diciendo que lo que es real para un sujeto depende siempre de la perspectiva que adopte.

Dependencia no implica determinación. Desde una perspectiva accedemos a algo que, aunque dependa de la perspectiva, no está determinado por ella. Las perspectivas no construyen absolutamente la realidad sin ningún material previo.

---

<sup>1</sup>Este trabajo ha sido realizado en el marco del Proyecto de Investigación *Puntos de vista, disposiciones y tiempo. Perspectivas en un mundo de disposiciones* (FFI2014-57409-R), financiado por el Ministerio de Economía y Competitividad del Gobierno de España.

En palabras de Putnam (en su prefacio a *Razón, verdad e historia*, 1982), “la mente y el mundo construyen conjuntamente la mente y el mundo”.

A fin de precisar esa peculiar relación de dependencia, el perspectivismo ha solido recurrir a la idea de la construcción de un espacio de perspectivas (muchas veces sin esta denominación explícita).

Construir un espacio de perspectivas consistiría en ofrecer una manera, más o menos formal, tal vez sólo intuitiva o mostrativa, de poder asignar posiciones relativas: 1) a cada perspectiva relevante, 2) a cada uno de sus contenidos y 3) a los posibles hechos o factores que pudieran no estar dentro del alcance de ninguna de las anteriores perspectivas.

Algunos rasgos importantes de esta caracterización son los siguientes:

- Se asume la posibilidad de que una parte de la realidad no llegue a estar bajo el alcance de ninguna de las perspectivas que se consideren relevantes.
- En esa parte de la realidad pueden situarse los propios sujetos que adoptan perspectivas.
- Las perspectivas relevantes pueden ser de primera, segunda o tercera persona.
- En principio, todas esas perspectivas podrían compartir unos mismos contenidos y existir complejas dinámicas de cambios de perspectiva frente a ellos.
- La construcción del propio espacio de perspectivas podría convertirse en una perspectiva más dentro de otro espacio de perspectivas más amplio.

La generalidad de nuestra caracterización nos permite entender como propuestas de construcción de espacios de perspectivas planteamientos tan diversos como la monadología de Leibniz, la construcción lógica del mundo de Carnap, la construcción de Russell de un espacio objetivo, en base a los datos de los sentidos contenidos de nuestras perspectivas subjetivas, las interrelaciones entre lo subjetivo y lo objetivo a través de lo intersubjetivo en Davidson, incluso la construcción de una concepción absoluta de la realidad de Williams.

A veces se ha considerado que ciertas obras pictóricas sirven para ejemplificar la idea de un cierto espacio de perspectivas. Uno de los casos más conocidos es el análisis que Foucault hace de *Las Meninas* de Velázquez, en conexión con el enfoque básico de la epistemología moderna (*Las palabras y las cosas*, 1999). En este sentido, también la literatura se manifiesta como un terreno fructífero para ejemplificarlo. En nuestro trabajo, queremos mostrar cómo los esfuerzos de creación literaria pueden servir para analizar algunos rasgos cruciales de los propios procesos dinámicos de construcción de espacios de perspectivas.

La creación literaria de una narración sirve como ejemplo sumamente ilustrador y sugerente del proceso de construcción de un espacio de perspectivas. Lo que se pretende conseguir es que en la narración resultante cada perspectiva ocupe un lugar apropiado con relación al resto de las perspectivas. Esto se aplica de manera muy especial a las relaciones entre las perspectivas de primera, segunda y tercera persona. La articulación de una trama argumental es un muy buen ejemplo de cómo todas estas perspectivas pueden compartir contenidos, y

modificarlos, a través de complejas dinámicas narrativas. Los posibles hechos o factores fuera del alcance de las perspectivas también están muy presentes en la construcción de una estructura narrativa, incluyéndose aquí muchas veces, como parte de la narración, a los propios sujetos que adoptan las perspectivas.

Basaremos nuestros análisis en la actividad literaria de tres autores escasamente tratados: Agustín Espinosa, Jean Genet y Hervé Guibert. En los tres, tiene un gran peso tanto el entorno real, que incluye ciertos factores determinantes más allá de las perspectivas adoptadas, como un propio yo personal sumamente esquivo que funciona como referencia última de toda perspectiva. En Espinosa, se trata de una realidad insular atlántica, en Genet de un mundo sórdido de marginación social, prostitución y delincuencia, en Guibert el tratamiento médico y la marginación social de una enfermedad como el SIDA, que él mismo debe afrontar. A la vez, en la obra de estos tres autores hay una dinámica compleja de perspectivas que ejemplifican muy bien cómo la primera, segunda y tercera persona pueden llegar a compartir y cambiar unos contenidos interactuando intensamente entre sí y con las anteriores realidades situadas fuera de esas mismas perspectivas, pero siempre presentes y entremezcladas en sus narraciones. Espinosa hará esto a través de una creación simbólica cargada de modernismo y surrealismo (*Crimen*, 1990; *Diario espectral de un poeta recién casado y otros textos*, 2005; *Lancelot 28°-7°*, 1988), Genet utilizará los recursos de su autobiografía y los espacios marginales en los que ésta se desarrolla (*Journal du voleur*, 1991; *Notre Dame des Fleurs*, 1980), Guibert emprende una creación literaria basada en la autoficción, en la creación de un yo, a través del deterioro progresivo ocasionado por la enfermedad que padece (*A l'amî qui ne me sauve pas la vie*, 1988; *Cytomégalo-virus*, 2013; *Le protocole compasionel*, 1991). Algunas de estas obras ofrecen además, ejemplos importantes de usos imaginarios de la primera persona en los que no puede asumirse como realidad personal literal el contenido expresado, o por lo menos resultaría muy problemático hacerlo (esto ocurre especialmente en *Crimen*, 1990, de Espinosa).

Nuestro interés por esta cuestión permite, además, dar cuenta de como la manera en que se resuelve en una obra de creación literaria el problema de la construcción de un espacio de perspectivas ofrece también otro tipo de respuesta complementaria al problema de las conexiones entre el perspectivismo y el relativismo. El perspectivismo no sólo se distanciaría del relativismo en su forma de entender la noción crucial de dependencia de una perspectiva. También hay una gran diferencia en cuanto al valor concedido a lo que se obtiene desde una u otra perspectiva.

Tanto el perspectivismo como el relativismo asumen un pluralismo. Pero mientras que para el relativismo todas las diferentes alternativas dentro de este pluralismo tendrían el mismo valor, para el perspectivismo puede perfectamente haber perspectivas mejores y peores. Justo en el mismo sentido en el que, aunque las narraciones resultantes de un proceso de creación literaria puedan ser ciertamente muy diferentes, no todas ellas tienen el mismo valor. Ni para el público receptor de las obras literarias, ni para los propios autores de estas.

Tanto en su peculiar forma de tratar la noción de dependencia respecto de las perspectivas como en sus implicaciones evaluativas, el perspectivismo se distanciaría del simple relativismo.

# Cuando los conceptos son no-persistentes y no-representacionales: Un examen contextualista-consistente de la mente

José V. Hernández-Conde

## Contextualismo y conceptos instanciados

Dos son las principales posturas que cabe adoptar en filosofía de la mente con respecto a cuál es el grado de dependencia contextual de los conceptos, a saber, invariantista y contextualista. Mientras que el *invariantismo* identifica los conceptos con cuerpos de conocimiento estables entre individuos y tiempos, para el *contextualismo* son constructos creados al vuelo de manera específica para cada ocasión (Barsalou 1983; Sperber y Wilson 1995; Prinz 2002). La principal ventaja de esta segunda postura –contextualismo– es que explica la adaptación de nuestro comportamiento ante entornos cambiantes.

Un elemento característico de los enfoques contextualistas es el de considerar que los conceptos necesitan ser instanciados (Barsalou 1983; Casasanto y Lupyan 2015), esto es, contruidos específicamente para cada contexto particular. Con ello, la instanciación de un concepto no es más que su construcción al vuelo para un contexto determinado (el cual dependerá del sujeto que realice esa instanciación, y del momento de tiempo en que lo haga). Los conceptos instanciados son los que intervienen en tareas de categorización, comunicación, inferencia, etc., siendo posible diferenciarlos de la información almacenada de manera estable sobre sus correspondientes conceptos (Hernández-Conde 2017).

Mi tesis será que, para el caso de las aproximaciones contextualistas, los conceptos instanciados deben ser vistos como el resultado de procesos cognitivos, pero no en el sentido de *product* (o entidad psicológica persistente almacenada en una estructura mental), sino de *fenómeno* (como algo que ocurre, o se manifiesta, cuando la mente categoriza algo bajo ese concepto). Sobre esta base mostraré que los conceptos instanciados carecen de persistencia, y que por esa razón no pueden ser considerados representaciones de sus categorías asociadas.

## Los conceptos (instanciados) carecen de persistencia

La distinción producto-fenómeno resulta más clara cuando se consideran ejemplos de procesos no pertenecientes al ámbito de la cognición. Así, para el caso de un algoritmo que calculase el cuadrado de un número y lo almacenase en un registro de memoria, estaríamos ante un proceso cuyo resultado es un *producto* –por ejemplo, el estado físico de un conjunto de transistores–. En cambio, si consideramos el caso de un controlador que compara la temperatura recibida desde un sensor con un cierto umbral y emite un flash de luz en caso de que aquella lo supere, lo que tenemos delante es un proceso cuyo resultado es un *fenómeno* –a saber, emisión de fotones durante unas milésimas de segundo–. Y mientras que en el caso del *producto* el resultado del proceso es persistente, y puede «accederse» a él en múltiples momentos de tiempo futuros, para el caso del *fenómeno* ese resultado es no-persistente, y no puede contrastarse más allá del instante en que el fenómeno ocurre.

Esta distinción, aunque no expresada en términos de productos y fenómenos, no es ajena al funcionamiento interno de un computador, en donde ciertos estados son temporalmente persistentes (como, por ejemplo, la información almacenada en disco, memoria RAM o caché), mientras que otros carecen por completo de persistencia (como ocurre con la memoria de programa y, en menor medida, los registros del procesador). Así, aunque podríamos hablar de estados persistentes cuando nos referimos a los datos almacenados en registros del primer tipo, eso no es posible para el caso del estado de la CPU, pues la memoria de programa está en continuo cambio.

Finalmente, los conceptos instanciados son resultado de procesos cognitivos del segundo tipo –esto es, fenómenos–, que los construyen a partir de un contexto específico para cada ocasión (y cambiante para cada individuo y momento de tiempo). Por tanto, y al igual que sucedía para el caso de la señal luminosa producida por el controlador, los conceptos (instanciados) carecen de persistencia, existiendo solamente en el preciso momento en que su proceso de instanciación concluye, pues más allá de ese instante el contexto cambiaría y el concepto instanciado no sería ya ése sino otro. Bajo esta perspectiva podría decirse que los conceptos instanciados no son algo que exista, sino el punto final de sus procesos cognitivos asociados, y por eso que deban identificarse, no con entidades o estados psicológicos, sino con eventos mentales.

## Los conceptos (instanciados) no son representaciones

Por otro lado, la visión dominante en ciencia cognitiva es que los conceptos son representaciones mentales (Margolis y Laurence 2007), esto es, particulares con propiedades semánticas (Pitt 2017). Ésta es la perspectiva de la teoría computacional de la mente (TCM), la cual concibe los procesos mentales como operaciones gobernadas por secuencias de reglas constituidas por objetos

evaluables semánticamente. Ahora bien, ya fuera de la TCM clásica, algunos enfoques conexionistas sostienen que –dentro de la mente– no es posible aislar ningún particular susceptible de ser identificado con una representación, y que en vez de representaciones mentales debería hablarse de codificación *distribuida* (no *local*) de información en niveles, pesos y nodos. De ser así, ningún estado o parte de la red representaría ningún concepto en particular, pues éstos estarían almacenados holísticamente de modo distribuido sobre toda la red (Ramsey, Stich y Garon 1990). Por ello, en este tipo de aproximaciones los conceptos no serían representaciones mentales.

Aquí, y aunque mi conclusión coincida con la de los enfoques conexionistas descritos (a saber, que los conceptos instanciados no son representaciones), esa conclusión se alcanzará tomando como punto de partida una asunción, no con respecto a la arquitectura computacional / neuronal de la mente –como ocurre en el conexionismo–, sino sobre el grado de dependencia contextual de los conceptos. En este punto mi tesis es que el elemento determinante para rechazar que los conceptos (instanciados) sean representaciones no es tanto la arquitectura –conexionista o clásica– de la mente, como sí el requisito de *persistencia* exigido a las representaciones (el cual, como ya se ha indicado, no se cumple para una visión contextualista-consistente de la mente).

De hecho, uno de los requisitos exigidos en cualquier teoría cognitiva a la noción de representación es el de persistencia. Sobre la base de ese requisito puede establecerse la distinción siguiente entre: (a) *representaciones*, o aquellos objetos relativamente estables y persistentes que codifican información; y (b) *procesos*, o aquellas operaciones dinámicas que, operando sobre esos objetos, los pueden hacer cambiar (Danks 2014).

Considerando lo anterior, los conceptos instanciados no serían representaciones, dado que no codifican información de manera estable en el tiempo –al carecer de persistencia–. (Obsérvese que los conceptos instanciados no son «proxitypos» a la Prinz (2002) –almacenados y en espera de ser recuperados–, ni la parte activa de la información almacenada sobre esa categoría, sino una construcción –específica del contexto– producida por el proceso de instanciación.) No obstante, aún si olvidásemos por un momento el requisito de estabilidad y, con ello, llegásemos a aceptar que los conceptos instanciados tienen carácter representacional, lo que no se podría decir es que son representaciones de sus categorías, puesto que son ellos mismos los que crean tales categorías. O, dicho de otro modo, si los conceptos instanciados son eventos generadores de categorías, entonces no se puede decir que son representaciones suyas.

Llegados a este punto cabría argumentar que, aún cuando los conceptos instanciados pudieran no ser una representación mental, sin embargo sí que podría serlo la información almacenada de modo estable sobre sus categorías –conceptos almacenados–. Ahora bien, la noción habitual de concepto es la de aquello que interviene en procesos de categorización, comunicación, inferencia, etc., y éstos son los conceptos instanciados. De hecho, en ausencia de un proceso de instanciación (que sería específico para cada contexto –u ocasión de uso– concreto) podría replicarse –en línea con los enfoques conexionistas ya mencionados–, que los conceptos almacenados dependen holísticamente del resto de información almacenada por el sistema cognitivo, razón por la cual no pueden representar ningún concepto en particular.

## Referencias

- Barsalou, Laurence W. (1983). Ad hoc categories. *Memory & Cognition*, 11:211–227.
- Casasanto, Daniel y Gary Lupyan (2015). All concepts are ad-hoc concepts. En E. Margolis y S. Laurence eds., *The Conceptual Mind: New Directions in the Study of Concepts*. Cambridge, MA: MIT Press, 543–566.
- Danks, David (2014). *Unifying the Mind*. Cambridge, MA: MIT Press.
- Hernández-Conde, José V. (2017). Ciclo de vida de un concepto en el marco de la cognición ad hoc. *Theoria. An International Journal for Theory, History and Foundations of Science*, 32:271–292.
- Margolis, Eric y Stephen Laurence (2007). The ontology of concepts: abstract objects or mental representations? *Noûs*, 41:561–593.
- Pitt, David (2017). Mental representation. En E.N. Zalta ed., *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), <https://plato.stanford.edu/archives/spr2017/entries/mental-representation/>
- Prinz, Jesse J. (2002). *Furnishing the Mind*. Cambridge, MA: MIT Press.
- Ramsey, William, Stephen P. Stich y Joseph Garon (1990). Connectionism, eliminativism, and the future of folk psychology. *Philosophical Perspectives*, 4:499–533.
- Sperber, Dan y Deirdre Wilson (1995). *Relevance: Communication and Cognition* (2<sup>nd</sup> ed.). Oxford: Blackwell.

# Ser racional no equivale a responder correctamente a razones: el caso de las intenciones contradictorias

Julen Ibarrondo Murguialday  
Universidad Autónoma de Madrid

## ¿Es responder a razones suficiente para manejar racionalmente nuestras intenciones?

En las últimas décadas la tesis de que en última instancia todos los conceptos normativos han de poder reducirse a razones ha alcanzado gran popularidad. La racionalidad, entendida como concepto normativo no ha sido una excepción. Así muchos epistemólogos y filósofos morales han defendido la tesis de que ser racional es de alguna manera conceptualmente reductible o al menos equivalente a responder adecuadamente a razones.

**Equivalencia:** Necesariamente un sujeto es racional si y solo si responde adecuadamente a sus razones.

Esta tesis implica dos partes:

**Necesidad:** Necesariamente si un sujeto no responde adecuadamente a sus razones entonces no es racional.

**Suficiencia:** Necesariamente si un sujeto responde adecuadamente a sus razones entonces es racional.

Mi propósito aquí es, siguiendo la estrategia general presentada por Broome (2007), presentar brevemente un caso, a mi juicio paradigmático, en el que parece claro que nos encontramos ante una situación de irracionalidad pero que no podemos caracterizar como no respondiendo adecuadamente a razones: el mantener intenciones contradictorias. Si esto es así, dado que la tesis de la equivalencia implica la suficiencia, se sigue que la tesis de la equivalencia entre racionalidad y responder a razones es falsa. Hay casos en los que uno responde adecuadamente a todas sus razones y es irracional. De ello se sigue que todo intento de explicar en qué consiste la racionalidad reduciéndola o fijando relaciones de equivalencia con el responder a razones está condenado al fracaso y que debemos buscar un enfoque alternativo para explicar en qué consiste gestionar nuestros estados mentales de manera racional.

Por supuesto, existen serias discrepancias entre aquellos que defienden la equivalencia entre racionalidad y el responder adecuadamente a razones, tanto acerca de la naturaleza de las razones como en qué consiste responder adecuadamente



a estas. Sin embargo, creo que puede mostrarse que la tesis de la suficiencia es falsa para los casos de intenciones contradictorias apelando únicamente a una comprensión mínima de las razones y del responder adecuadamente a ellas compartidas por todos, de manera que mi crítica se aplica a todos estos modelos a pesar de sus significativas diferencias.

Mi argumento parte de la premisa de que mantener intenciones que uno cree racionalmente contradictorias es siempre irracional, de forma que toda teoría de la racionalidad satisfactoria debe ser capaz de dar cuenta de este caso. Por intención entiendo el compromiso con un determinado curso de acción, que típicamente se manifiesta en llevar a cabo una serie de acciones a lo largo del tiempo (Bratman, 1987). Hasta donde llega mi conocimiento ningún partidario de la equivalencia niega que la incoherencia entre intenciones no sea un caso de irracionalidad, lo cual no es de sorprender dado que los casos de este tipo intuitivamente se nos presentan como claramente irracionales. Consideremos el siguiente ejemplo:

**Camisa o camiseta:** Estoy preparándome para dar una conferencia y he decidido que llevaré camisa y comienzo a actuar en consecuencia; por ejemplo me pongo a plancharla. Sin embargo, sin haber reparado en que sé que si llevo camiseta entonces no podré llevar camisa, he decidido también llevar camiseta. Estoy comprometido con el curso de acción que me lleva a ir a la conferencia en camiseta. Podemos imaginar que esto se refleja también en mi conducta, por ejemplo echando a la lavadora la camiseta a fin de tenerla limpia para mañana.

Parece claro que en un caso como este estoy siendo irracional, que es condición necesaria de ser racional que si creo que dos intenciones son contradictorias entonces no puedo tener ambas. Que mantener intenciones que se creen contradictorias es irracional viene respaldado tanto por el hecho de que es un objeto típico de la crítica racional y el consejo como que, para las personas que consideramos racionales normalmente basta con que se percaten de que se encuentran en una situación así para que decidan no mantener las dos intenciones simultáneamente.

El problema está en que, tal y cómo es admitido por la inmensa mayoría de los teóricos de las razones, a menudo hay casos en los que las razones de las que dispongo no se inclinan decisivamente a favor de ninguna de las dos intenciones; que ambas opciones son permisibles en función de las razones de las que dispongo. En un caso como el del ejemplo tanto una opción como la otra son compatibles con responder adecuadamente a razones. Sin embargo, la tesis de la equivalencia exige que en los casos en los que ni una intención es contrario a mis razones ni la otra sea contrario a razones que mantenga ambas. Para el defensor de la equivalencia esto ha de ser así porque, de alguna manera, el hecho de que me las proponga conjuntamente tiene un impacto en mi conjunto de razones. ¿Pero qué clase de impacto es este? La mayor parte de mi exposición se dedicará a mostrar los que a mi entender son los tres principales candidatos más plausibles discutidos en la literatura. Mi propósito será mostrar que ninguno de ellos es satisfactorio.

## Consecuencias para la racionalidad instrumental

Antes de centrarme en la cuestión es reseñable remarcar que el problema puede extenderse a la racionalidad instrumental, a la exigencia racional de tomar los medios que consideramos necesarios para nuestros fines, si la caracterizamos (como parece que puede hacerse) en términos de intenciones contradictorias entre los fines y los medios.

**Irracionalidad instrumental como intenciones contradictorias:** Si S cree racionalmente que solo tendrá éxito al proponerse I1 si se forma la intención I2, entonces es irracional que S no adopte la intención I2.

En otras palabras, dadas ciertas expectativas racionales acerca de qué acciones propiciarán nuestros fines, no formarse la intención de adoptar los medios supondrá contradecir nuestros propósitos. Si no podemos probar la tesis de la suficiencia para las intenciones contradictorias el proyecto de dar cuenta de la racionalidad instrumental en términos de responder a razones también se ve comprometido. Si no hay razones por las que el hecho de tener un propósito nos compela a no adoptar otros propósitos que lo contravengan, la idea de que siempre hay razones en contra de no tomar los medios adecuados a nuestros fines que seriamente comprometida (Ibarrondo, próximamente).

Así, es de esperar que si la tesis de la suficiencia no se aplica satisfactoriamente a las intenciones contradictorias tampoco lo hará al menos a ciertos casos de irracionalidad instrumental. Los problemas para los defensores de la equivalencia serán dobles: habrá casos en los que uno responda adecuadamente a sus razones y sin embargo será instrumentalmente irracional.

## Posibles defensas de la suficiencia para las intenciones contradictorias

El defensor de la tesis de la suficiencia necesita una teoría que explique cómo el adoptar un propósito, que de por sí puede ser meramente permitido por razones, puede otorgar razones en contra de adoptar otros propósitos que, si bien por sí mismos serían también al menos permisibles, no lo son una vez se ha aceptado realizar la intención que previsiblemente los contradice. Estas son las explicaciones más verosímiles de este fenómeno:

### *Primer candidato: Razones relativas a actitudes*

Una primera explicación plausible de cómo la conjunción de dos intenciones contradictorias, cada una de ellas permitida por tus razones, pasan a ser contrarias a tus razones cuando las mantienes conjuntamente consiste en postular que el hecho precisamente de sostener creencias que consideras contradictorias es una buena razón (plausiblemente una razón decisiva) para no mantener esa conjunción de estados. Dado que a menudo uno abandona una intención al descubrir que se contradice con otra, es intuitivo pensar que la creencia de que I1 e I2 son

contradictorias combinadas con el tener la intención I1 le provee de una razón decisiva en contra de I2. Así, mientras mantenga la creencia e I1 será contrario a sus razones proponerse I2. Esto parece explica por qué casos así son irracionales. El hecho de tener ciertas actitudes es ya una razón en contra de adquirir una nueva que la contradiga.

Sin embargo la idea de que nuestras actitudes puedan por si mismas generar nuevas razones es del todo inaceptable, dado que nos lleva directamente a lo que se conoce como el problema del “bootstrapping” (Kolodny, 2007). Imaginemos que I1 e I2 son dos intenciones permisibles, pero contradictorias, y que sería contrario a mis razones no proponerme ninguna de las dos. Yo ya he resuelto hacer I1. Entonces, según el modelo de las razones relativas a actitudes yo tengo razones decisivas en contra de I2, lo que hace a su vez que mis razones me exijan I1. Por estipulación mis razones me permitían hacer I2 y no me exigían hacer I1. El hecho de que tenga la actitud de hacer I1 no puede añadir una razón a favor de I1 y contra I2. No puede ser que uno genere razones a favor de actitudes simplemente adoptando esas actitudes.

### ***Segundo candidato: las contradicciones entre intenciones implican contradicciones en las creencias***

Tener una intención implica que es posible llegar a llevarla a cabo. Sin embargo, si uno sabe que no va a poder realizar ambos objetivos porque se excluyen, entonces uno tiene una razón para no adoptar las dos intenciones, dado que implicarían creencias contradictorias (Velleman 1989, cap. 3 y Wallace, 2001) y tener creencias contradictorias es algo contra lo que siempre tenemos buenas razones, principalmente porque nos indica que al menos una de ellas es falsa. Tener intenciones contradictorias siempre conlleva una incoherencia en el ámbito de las creencias y tenemos razones independientes para tachar de irracional el mantener creencias contradictorias.

Esta línea de argumentación hace depender la irracionalidad en la intención de la irracionalidad en las creencias acerca de qué podemos hacer. Desgraciadamente no es cierto que proponerse I1 implique creer que es posible I1; puede que suspendas el juicio. Por ejemplo: tengo la intención de llevar la camisa al congreso mañana, pero tengo indicios inconclusos de si mañana hay congreso, así que suspendo el juicio acerca de si mañana podré llevar la camisa al congreso a pesar de que tengo la intención de llevarla. No descarto que sea imposible para mí llevar la camisa mañana al congreso (porque en realidad resulta que no hay tal congreso mañana) pero eso no hace que deje de tener la intención.

Si tener la intención de realizar algo no nos obliga a comprometernos con que sea posible hacerlo, entonces uno puede proponerse fines contradictorios sin que eso involucre creencias contradictorias: Si me propongo llevar la camisa al congreso y al mismo tiempo llevar la camiseta no necesito creer que podré llevar ninguna de las dos. Suspendo el juicio tanto acerca de si iré al congreso con una como con la otra, por lo que mis creencias acerca de lo que voy a llevar mañana al congreso son coherentes , pero mis intenciones no.

### *Tercer candidato: Evidencias a favor del derroche*

El hecho de que I1 y I2 y creas que no son compatibles te da evidencia suficiente de que estas malgastando tanto esfuerzos intelectuales como prácticos, de manera que tienes una razón para decidir si I1 o I2. En otras palabras, la previsión de los costes de mantener creencias contradictorias nos da una razón decisiva para que abandonemos una de las dos, de manera que al volver coherente el conjunto de nuestras intenciones estamos respondiendo a razones. El percatarse de una incoherencia provee de una razón pragmática para decidirse entre por una de las intenciones en liza (Kiesewetter, 2017, cap. 9).

El problema es que esto no puede fijar que uno siempre tenga razones decisivas en contra de tener creencias contradictorias. Esto es así por ejemplo si uno cree que haga lo que haga corre el riesgo de involucrarse en derrochar esfuerzos o que el derroche sea tan pequeño que en realidad no tenga relevancia normativa alguna. A fin de que el derroche tenga algún efecto normativo ha de alcanzar cierto mínimo y ha de poder evitarse abandonando una de las intenciones. Sin embargo, esto no siempre es así.

Me propongo a un mismo tiempo ver dos películas que echan simultáneamente en dos canales de televisión. Obviamente se trata de una pretensión irracional, ¿pero qué clase de derroche está en juego que me otorgue razones decisivas a favor de elegir ver una de las dos películas? Hay algo en mi manera de razonar que no es adecuado, pero lo realmente relevante que permite criticar mi conducta no parece basarse en unas pérdidas tan vagas e inocuas. Lo que me hace irracional no es no responder de inmediato para evitar este derroche aparentemente tan insignificante.

Además, esta explicación implica que dado que la razón para no tener creencias contradictorias es prevenir el derroche y que guiarse adecuadamente por razones requiere no sólo hacer aquello que coincide con las razones que tiene sino hacerlo por esas razones, uno es irracional si evita las intenciones contradictorias sencillamente porque reconocer que son contradictorias, pero esto parece ser precisamente lo que los agentes racionales hacen. Parece que la manera racional de abandonar creencias contradictorias no responde en lo esencial a evitar ningún derroche. Lo que nos motiva es el percatarnos de la contradicción misma, no el evitar ciertos efectos negativos que preveemos que se seguirá de esta.

## **Conclusiones**

La tesis de equivalencia tiene serios problemas para explicar por qué es irracional sostener intenciones contradictorias. El problema radica en que, al contrario de lo que parece ser el caso en las creencias, nuestras razones a veces nos permiten forjar ciertas intenciones sin llegar a exigirnoslas. Sin embargo, si la tesis de la suficiencia es correcta, es necesario que haya una evaluación posible en términos de ponderación de razones que impida siempre tener creencias contradictorias. Lo que necesitamos es reconocer algún tipo de razón decisiva que siempre decante la balanza en contra de sostener ambas actitudes.

Sin embargo no parece que exista una teoría que pueda explicar satisfactoriamente en qué consiste esta razón decisiva. Además, como he comentado brevemente, el problema no se circunscribe al ámbito limitado de la contradicción en las intenciones tal y como normalmente se concibe este ámbito, sino que el problema es también trasladable al campo de la racionalidad instrumental: no tenemos una explicación satisfactoria acerca de cómo el hecho de proponernos un fin nos otorga una razón decisiva a favor de proponernos también los que razonablemente consideramos que son los medios pertinentes.

Estos serios problemas a la hora de dar cuenta de la equivalencia en importantes ámbitos de la razón práctica ponen en entredicho el proyecto de explicar nuestro concepto de racionalidad a partir del de ser una razón. Si queremos dar cuenta de lo irracional en sostener intenciones contradictorias necesitamos una teoría diferente acerca de en qué consiste ser racional.

## Referencias

- Broome, J. (2007). "Does Rationality Consist in Responding Correctly to Reasons?", *Journal of Moral Philosophy*, 4(3):349–374.
- Bratman, M.E. (1987). *Intentions, Plans, and Practical Reason*. Chicago. The University of Chicago Press.
- Ibarrondo, J. (próxima publicación). "¿Estamos racionalmente obligados a ser eficientes?", *Análisis Filosófico*.
- Kiesewetter, B. (2017). *The Normativity of Rationality*. Oxford University Press.
- Kolodny, N. (2005). "Why be rational?", *Mind*, 114(455):509–563.
- Velleman, J. (1989). *Practical Reflection*. Stanford: CSLI (reimpreso 2007).
- Wallace, J. (2006). *Normativity and the Will*. Oxford University Press.

# Razonamientos que involucran perspectivas<sup>1</sup>

**Manuel Liz**

manuliz@ull.es

**Margarita Vázquez**

mvazquez@ull.es

Facultad de Humanidades (Sección de Filosofía)  
Universidad de La Laguna  
Islas Canarias. España

Nuestro punto de partida es una distinción entre, por una parte, razonamientos que dependen de una perspectiva y, por otra parte, razonamientos sobre perspectivas. Queremos dirigir la atención sobre estos últimos.

La distinción que acabamos de hacer es particularmente convincente si tenemos en cuenta dos consideraciones: 1) aunque todo nuestro pensamiento fuera independiente de las perspectivas, podríamos razonar sobre las perspectivas, y 2) aunque todo nuestro pensamiento fuera dependiente de una perspectiva, podríamos carecer de razonamientos acerca de las perspectivas.

De hecho, y muy a menudo, nuestro pensamiento sobre el mundo y sobre nosotros mismos involucra explícitamente referencias a perspectivas o puntos de vista (utilizaremos aquí de forma equivalente ambas nociones<sup>2</sup>). Se trata de un pensamiento acerca de cómo algunos contenidos son los contenidos de ciertas perspectivas y cómo se relacionan esos contenidos y las perspectivas que permiten acceder a ellos con otros contenidos, tal vez fuera del alcance de las perspectivas en juego. Hay una serie de patrones inferenciales que articulan estos pensamientos. Y apenas han sido objeto de análisis filosóficos detallados.

Existen algunos importantes planteamientos lógicos que introducen nuevos operadores en los sistemas habituales de lógica modal. Destacan los trabajos clásicos de Antti Hautamäki (1983 y 1986) y los más recientes de Steen Hales (1997 y 2006). Los nuevos operadores representan las nociones de “Bajo todo punto de

---

<sup>1</sup>Este trabajo ha sido realizado en el marco del Proyecto de Investigación *Puntos de vista, disposiciones y tiempo. Perspectivas en un mundo de disposiciones* (FFI2014-57409-R), financiado por el Ministerio de Economía y Competitividad del Gobierno de España. Los resultados que ofrecemos han sido discutidos en diversas reuniones vinculadas a dicho proyecto. Queremos agradecer a los miembros del mismo todos sus valiosos comentarios y sugerencias.

<sup>2</sup>Esta equivalencia no es en modo alguno inocente. Imaginemos que reservamos el rótulo “perspectiva” para designar un tipo general o abstracto de propiedades y relaciones, por ejemplo de acceso, que cuando son instanciadas por ciertos sujetos (además de por otras cosas) se convierten en “puntos de vista”. Así, al adoptar una perspectiva, un sujeto ejemplificaría un punto de vista. Con todas sus ventajas, este enfoque impediría entender directamente a estos sujetos tanto en base a los puntos de vista (sería circular o regresivo) como en base a las perspectivas (los sujetos se convertirían en algo abstracto). Sin duda, estos problemas son importantes. Pero queremos evitarlos.

vista, *p*” y “Bajo algún punto de vista, *p*”. Semánticamente, funcionan como nuevos índices de evaluación de las fórmulas. Hautamäki construye una gran variedad de posibles lógicas de los puntos de vista. Las mezclas de los operadores habituales de necesidad y posibilidad con los nuevos operadores abren un abanico de combinaciones difícilmente reducible. A su vez, Hales propone una interpretación modal del relativismo. En ella, la noción de “Relativamente, *p*” y “Absolutamente, *p*” se convierten en nuevos operadores modales intensionales interpretables, respectivamente, como “En cierto mundo posible en perspectiva, *p*” y “En todo mundo posible en perspectiva, *p*”. De esta forma, el relativismo y el absolutismo resultarían perfectamente compatibles.

Nuestro enfoque es compatible con estos tratamientos modales. Pero intenta precisar con más detalle los patrones inferenciales concretos que articulan nuestros razonamientos sobre las perspectivas. Tal vez no sea estrictamente posible una “lógica” sobre estos razonamientos. Tal vez únicamente sea posible un enfoque muy pragmático. En cualquier caso, nuestro objetivo es identificar y analizar un conjunto relevante de patrones inferenciales.

Al tratar con razonamientos explícitamente verbalizados, sorteamos el problema que plantean los contenidos no-conceptuales, contenidos cualitativos, que pueden tener muchas de las perspectivas que intervienen en ellos. Al razonar sobre perspectivas, lo único que puede contar son las descripciones proposicionales que hacemos de ciertos contenidos, incluyendo aquellos contenidos no-conceptuales, o cualitativos, típicamente característicos de las perspectivas en primera persona.

Ofreceremos un conjunto de patrones inferenciales en forma de diez reglas, presentadas de manera informal. Las variables “*p*”, “*q*”, etc., representarán indistintamente proposiciones, hechos y contenidos. Y podrán involucrar, como acabamos de decir, fenómenos no conceptuales. O referirse a ellos.

## R1: Identidad y diferencia

- Dos puntos de vista son el mismo *syss* 1) tienen el mismo portador, individual o colectivo, 2) tienen el mismo contenido, 3) las actitudes que se mantienen hacia ese contenido son las mismas, y 4) los puntos de vista satisfacen las mismas condiciones espacio-temporales de posesión.

Esta regla establece las condiciones de identidad y diferencia de los puntos de vista dentro de un discurso.

Las cláusulas 1-4 definirían los elementos estructurales de un punto de vista. Tomando como modelo las actitudes proposicionales, podemos decir que la naturaleza interna de un punto de vista PoV puede ser representada mediante una estructura<sup>3</sup>

$\langle B, A, n\text{-}CC, CC, Cp \rangle$ , donde

1. *B* es el portador del PoV, ocasionalmente un sujeto psicológico personal, pero no necesariamente,

---

<sup>3</sup>Hemos analizado en detalle este tipo de aproximación a los puntos de vista en Liz (2013), Liz (ed.)(2013), Vázquez y Liz (2011) y Vázquez y Liz (eds.)(2015).

2. n-CC y CC son, respectivamente, conjuntos de contenidos conceptuales y no- conceptuales,
3. A es un conjunto de relaciones entre B y esos contenidos, en nuestro caso particular se trataría de actitudes tales como percibir, creer, desear, recordar, etc.
4. Cp son las condiciones de posesión del PoV, quizás diferentes de las condiciones de atribución.<sup>4</sup>

En lo que sigue, supondremos que p puede representar tanto un contenido conceptual CC como un contenido no conceptual n-CC. Como ya hemos indicado, aun cuando se trate de un contenido no conceptual, el discurso que lo menciona puede identificarlo suficientemente, de manera que sea posible llevar a cabo razonamientos que lo involucran.

## R2: Particularización

- Si hay un acceso a p desde todo punto de vista, entonces hay un acceso a p desde cierto punto de vista particular.

Ejemplo, “Todos piensan p, luego también nosotros (vosotras, ellas, etc.)”.

R2 describe la particularización típica del razonamiento sobre perspectivas. A partir de “Desde todos los puntos de vista, p”, se deriva “Desde cierto punto de vista, p”.

Debe señalarse que es sumamente infrecuente encontrar en nuestras prácticas inferenciales derivaciones que a partir de “Desde todos los puntos de vista, p” lleguen a “Desde algún punto de vista, p”. A pesar de tratarse de una derivación lógica correcta, este tipo de inferencias particularizadoras, presentadas con el formato “Desde algún punto de vista”, no tienen ningún uso claro. Y sin embargo, sí lo tienen inferencias particularizadoras presentadas con el formato “Desde cierto punto de vista, p” (el de un sujeto o grupo sobre el que se quiere llamar la atención).<sup>5</sup>

## R3: Generalización

- Si un punto de vista particular ofrece acceso al contenido p, entonces hay puntos de vista ofreciendo acceso a p.

Ejemplo, “Hay gente que cree que p, sin ir más lejos yo mismo.”

---

<sup>4</sup>Si las perspectivas, o puntos de vista, no tuvieran condiciones de posesión, si únicamente tuvieran condiciones de atribución, su realidad sería tan poco sustantiva, tan arbitraria o caprichosa, como la realidad que parecen tener (a veces) los hechos relativos al gusto. De todas formas, también podemos evitar estos problemas cuando tratamos de analizar las reglas de nuestros razonamientos que involucran perspectivas.

<sup>5</sup>Debemos a José Zalabardo observaciones importantes sobre este punto, que sin duda necesita un desarrollo mucho mayor. Las diferentes formas de particularizar puede tener un efecto realmente muy diferente en nuestras prácticas argumentativas.



Típicamente, la generalización de perspectivas va desde la afirmación de que cierto punto de vista particular permite acceder a cierto contenido a la afirmación de que hay puntos de vista que permiten acceder a ese contenido. Por ejemplo, de que yo crea que  $p$ , de que exista un cierto punto de vista creencial mío que asume  $p$ , se infiere que hay, o existen, puntos de vista desde los cuales  $p$ .

R2 y R3 describen patrones inferenciales muy comunes. Respectivamente, particularizan y generalizan puntos de vista.

## **R4: Perspectivismo**

- Si  $p$  es un hecho, entonces ha de poder serlo desde cierto punto de vista.

Un sencillo ejemplo: “Ocurrió  $p$ , cualquiera que pasara por allí podría haberlo visto.” (Incluso, en ciertas situaciones, “tendría que haberlo visto”).

R4 junto con R5, que veremos a continuación, se conectarían estrechamente con el relativismo.

Sin embargo, como también veremos, R10 permitirá una importante posibilidad de distanciamiento respecto al relativismo, en un sentido dinámico.

## **R5: Pluralismo**

- Si bajo cierto punto de vista  $p$ , entonces es posible que bajo otro punto de vista  $\neg p$ .

Ejemplo: “Bueno, esa es mi opinión. Hay otras.”

En nuestras argumentaciones, R5 tiene un alcance sumamente amplio. Incluso llega a afectar a toda supuesta verdad lógicas. La variable  $p$  puede ser instanciada por verdades lógicas. En principio, cualquier verdad lógica podría instanciar  $p$ .<sup>6</sup>

R4 y R5 sugieren también cierto constructivismo respecto de los hechos. Pero quedaría contrarrestado con la siguiente regla R6.

## **R6: Acceso**

- Si bajo cierto punto de vista  $p$ , entonces por defecto  $p$ .

Ejemplo: “Yo estaba allí. Y lo que ví fue eso. Eso fue lo que ocurrió.”

La noción de “implicación por defecto” es aquí crucial.  $A$  implica por defecto  $B$  si lo implica sólo en la medida en que no ocurra nada que pueda cancelar de forma relevante tal implicación.

Nuestros puntos de vista permiten acceder a los hechos en este sentido. Implican por defecto transparencia respecto a los hechos.

---

<sup>6</sup>De nuevo, debemos esta observación a José Zalabardo.

R6 permite la entrada inferencial de contenidos que pueden ser accedidos desde múltiples perspectivas, o que incluso sean considerados independientes de la perspectiva.

### **R7: Enfoque**

- Si bajo cierto punto de vista  $p$ , entonces hay algún  $q$  tal que no es posible acceder a  $q$  desde tal punto de vista.

Ejemplo: “Si adoptas un punto de vista exclusivamente económico, no podrás entender la política.”

La noción de enfoque define un “campo” (de contenidos a los que se tiene acceso) y un “horizonte” (de contenidos accesibles). Ambas nociones son muy importantes.

Los puntos de vista siempre son selectivos respecto a sus contenidos en un sentido excluyente.

Como consecuencia de R7, no existiría un punto de vista completamente universal.

Una hipótesis interesante: Las diferencias fundamentales entre puntos de vista de primera, segunda y tercera persona pueden ser tratadas como diferencias de enfoque ligadas a distintas clases de actitudes.

### **R8: Adopción**

- El que desde un punto de vista se acceda al hecho de que desde otro punto de vista diferente se accede a  $p$ , no implica que desde el primer punto de vista se acceda a  $p$ .
- El que se acceda a  $p$  bajo un punto de vista no implica que exista otro punto de vista bajo el cual se acceda al hecho de que se accede a  $p$  desde tal punto de vista.
- Si desde cierto punto de vista se accede al hecho de que desde ese mismo punto de vista se accede a  $p$ , entonces ese punto de vista permite acceder a  $p$ .

La primera subregla establece una distinción sumamente importante entre “adoptar un punto de vista” y “tener un punto de vista sobre otro punto de vista”. Por ejemplo, un policía puede tener un punto de vista sobre el punto de vista de un delincuente sin adoptar el punto de vista del delincuente.

R8 perfila nuestra capacidad para adoptar “puntos de vista reflexivos”.

### **R9: Comprensión**

- Comprender un punto de vista implica comprender todo lo que se siga de su adopción

Ejemplo: “No puedes entender lo que pasó si no adoptas su punto de vista.”

Existiría una diferencia importante entre “comprender un punto de vista” y “explicar sus elementos estructurales y cómo se relacionan”.

Esta regla tiene conexiones sumamente interesantes con los múltiples principios de clausura del conocimiento y de la creencia.

Respecto a la comprensión de los puntos de vista, no parece bastar “comprender todo lo que se comprenda que se sigue de su adopción”. Se requeriría mucho más: “comprender todo lo que se siga de su adopción”.

### R10: Ajuste dinámico

- Si bajo un punto de vista se accede a  $p$  y bajo otro punto de vista se accede a  $\neg q$ , entonces 1) si hay un proceso que conduzca a la convergencia de  $p$  y  $q$ , habrá un proceso de diversificación de los anteriores puntos de vista, y 2) si hay un proceso de convergencia de dichos puntos de vista, habrá un proceso de diversificación de los contenidos  $p$  y  $q$ .

R6 permitía la entrada, por defecto, de contenidos en alguna medida independientes de la perspectiva. R10 describe la dinámica combinada de estos contenidos en relación a cambios de perspectiva.

R10 muestra la solución dinámica que nuestras prácticas inferenciales ofrecen a los casos de “faultless disagreements”. Como una cuestión de hecho ante estos casos o bien nuestras perspectivas acaban cambiando, o bien lo que cambian son las apreciaciones que hacemos, o bien acaban cambiando ambas cosas (y esto último es sumamente frecuente).

Ver los casos de “faultless disagreements” como planteando profundos problemas filosóficos (semánticos, epistemológicos, acaso también metafísicos) sólo es el resultado de la abstracción filosófica. Simplemente, hay contextos en los que no se llega a producir la dinámica descrita por R10 porque no hay nada importante que se siga de adoptarla.

Hay una conexión muy importante de todo esto con el pluralismo introducido por R5. Es una consecuencia de R5 que la diversificación de puntos de vista siempre ha de ser posible. Si bajo cierto punto de vista  $p$ , entonces han de ser posibles puntos de vista desde los cuales  $\neg p$ . Y aquí cabe una variedad muy amplia de alternativas a  $p$ .

Podemos concluir sugiriendo que las anteriores reglas determinan algunos de los aspectos más importantes de la idea de corrección que regula nuestras prácticas argumentativas que involucran perspectivas.

Acaso sea posible definir de modo más preciso una única relación, o un cierto operador especial, entre descripciones de perspectivas capaz de recoger todos los anteriores aspectos.<sup>7</sup> Sin duda, esto supondría un avance muy importante.

---

<sup>7</sup>Debemos esta sugerencia de formalización a Antti Hautamäki. Y actualmente estamos trabajando con él en este sentido. Hace años, Hautamäki hizo contribuciones importantes al análisis lógico del discurso que involucra perspectivas. Véase Hautamäki (1983 y 1986). Más

De cualquier modo, y sorprendentemente, el territorio que puede verse ante nosotros está en su mayor parte inexplorado. El análisis de las reglas que aceptamos y seguimos al razonar sobre las perspectivas, sobre los puntos de vista, es algo que está aún por hacer.

## Referencias

- Hales, S. (1997). “A Consistent Relativism”, *Mind*, 196(421):33–52.
- Hales, S. (2006). *Relativism and the Foundations of Philosophy*. Cambridge, MIT Press.
- Hautamäki, A. (1983). “The Logic of Viewpoints”, *Studia Logica*, 42(2/3):187–196.
- Hautamäki, A. (1986). *Points of View and their Logical Analysis*, *Acta Philosophica Fennica*. Helsinki, 41.
- Hautamäki, A. (1992). “A Conceptual Space Approach to Semantic Networks”. *Computers & Mathematics with Applications*, 23 [también en F. Lehmann (ed.), *Semantic Networks in Artificial Intelligence*, Oxford, Pergamon Press, 1992, 517–25].
- Hautamäki, A. (2015). “Points of view: A conceptual spaces approach”, *Foundations of Science*, 21(3):493–510.
- Liz, M. (2013). “Models and Points of View. The Analysis of the Notion of Point of View”, en L. Magnani (ed.) *Model-Based Reasoning in Science and Technology. Studies in Applied Philosophy, Epistemology, and Rational Ethics*, Vol. 8, Berlin, Springer, 2014, 109–28.
- Vázquez, M., and M. Liz (2011). “Models as Points of View. The Case of System Dynamics”, *Foundations of Science*, 16(4):383–391.
- Vázquez, M., and M. Liz (eds.)(2015). *Temporal Points of View. Objective and Subjective Aspects*. Berlín, Springer.

---

recientemente, ha propuesto el empleo de la herramienta topológica más amplia de los “espacios conceptuales”, cuyo principal desarrollador es Peter Gärdenfors. Véase Hautamäki (1992 y 2015). El análisis de las prácticas argumentativas que involucran la noción de perspectiva ofrecería otro tipo sugerente de aproximación.

# On the question of the indispensability of modal thought

Felipe Morales Carbonell

What is the role of experience in the acquisition of modal knowledge? What is the role of modal knowledge in experience? These questions loom large in contemporary modal epistemology (hence, for example, the emphasis on the debate between empiricist and rationalist approaches). Here, my goal is to frame these debates in terms of the question of whether and how modal thought is indispensable.

This way of posing the problem might sound reminiscent of the Kantian idea that modal categories are constitutive of experience. Recently, Brandom (2008) has developed a defense of what he calls the *Modal Kant/Sellars thesis* (KST), which is the conjunction of:

- 1) In using ordinary empirical vocabulary, one already knows how to do everything one needs to know how to do in order to introduce and deploy modal vocabulary.
- 2) The expressive role characteristic of alethic modal vocabulary is to *make explicit* semantic, conceptual connections and commitments that are already implicit in the use of ordinary empirical vocabulary.

It follows that the task of modal epistemology will be double: first, it should describe the conceptual capacities involved in modal thought, and second, it should give an account of how those capacities are established. Brandom's suggestion is that modal vocabulary is grounded in some of the same conceptual capacities empirical vocabulary makes use of. A similar point is made by Williamson (2007, p. 137):

Our overall capacity for somewhat reliable thought about counterfactual possibilities is hardly surprising, for we cannot know in advance exactly which possibilities are or will be actual. We need to make contingency plans. In practice, *the only way for us to be cognitively equipped to deal with the actual is by being cognitively equipped to deal with a wide variety of contingencies*, most of them counterfactual. (the emphasis is mine)

This leads him to commit to *anti-exceptionalism*, which is the claim that the means to acquire modal knowledge are instances of more general, ordinary, capacities, not *sui generis*.

As I indicated above, my suggestion here will be that the KST and Williamson's claim (and similar claims) should be spelled out in terms of *indispensability*.<sup>1</sup>

---

<sup>1</sup>As an antecedent, Hale (2013) refers to the indispensability of logical modality more explicitly.

What do I mean by this? A typical definition of indispensability is as follows:

**Indispensability:** X is *indispensable* for G iff there is no alternative to X for G that serves or would serve the goal of G better (alternatively: iff it would be impossible (in some relevant sense) to realize G unless with X)<sup>2</sup>

Indispensability claims are used to support *commitments*. The probably most famous example of an argument based on indispensability is Quine/Putnam's about mathematical entities:

- 1) We ought to have *ontological commitment* to entities which are indispensable for our best scientific theories.
- 2) Mathematical entities are indispensable for our best scientific theories.
- C) We ought to have ontological commitment to mathematical entities.

There is nothing, however, that restricts the type of commitment that any indispensability argument could support to *ontological* commitment. At its core, an indispensability argument transfers certain epistemic properties (justification, truth, etc.) from an admissible *datum* to some disputed claim (Panza & Sereni 2016). For example, we can try to support commitment to theories, principles, the applicability of methods of belief formation, and so on.

Enoch and Schechter (2008) present an argument for *vindicating* basic belief-formation mechanisms (like inference to the best explanation (IBE) or reasoning by *modus ponens*) which can serve as a template for an indispensability argument in the modal case. The central notion they use is that of a *rationaly required project* (RRP): a long term recurring task which can be assessed in epistemic terms, and which we would be rationally blameworthy for abandoning. They identify 4 RRP's: the explanatory project of making sense of and understanding the world, the deliberative project of making decisions, the project of planning for the future, and the project of self-evaluation. Their argument in defense of IBE and reasoning by *modus ponens* runs as follows:

- 1) We ought to commit to the belief-formation mechanisms which are indispensable for our *rationaly required projects* (RRP).
- 2) IBE and reasoning by *modus ponens* are indispensable for our RRP.
- C) We ought to commit to IBE and reasoning by *modus ponens*.

Indispensability claims like those in 2) take the form:

**Indispensability\*:** For some rationally required project *R*, it is or would be impossible\* to realize *R* unless one engaged in X.

Taking this as a template, we then have for the modal case the following claim:

**Indispensability-Modal:** For some rationally required project *R*, it is or would be impossible to realize *R* unless one engaged in modal thought (of type *t*).

There are reasons to think that modal thought is at least involved in all the projects that Enoch and Schechter identify as rationally required. This is the clearest, I think, in the case of decision-making, where reference and evaluation

---

<sup>2</sup>I follow Colyvan (1999) and Field (1989).

of alternatives (hence, possibilities for the one who makes decisions) seems incapable, along with consideration of what would happen if those alternatives were taken (cf. Stalnaker (1996)). The same goes for planning, where besides the consideration of alternatives in the form of different ways that aims could be realized, one often has to consider, on the one hand, contingencies that could prevent those aims to be realized, and, on the other, fixed points that we can rely on. Some notion of fixedness seem to be operative in that of an explanation, in the sense that whatever is explained is shown, through the explanation, to have been fixed in some way by something else that serves as its support or ground. Furthermore, looking for explanations requires the consideration of what could fill that role of fixing the *explanandum* in that way. This modal aspect of explanation is present no matter what kind of account of explanation we commit to. Finally, while self-assessment may appear to deal with the evaluation of actual *de se* properties, it seems to be equally important for it that we are capable of evaluating our dispositions and capacities. In an important sense, we engage in self-assessment *because* we need to consider what we *can* do.

The importance of modal thought for the engagement in rationally required projects is plausible. However, we might still need more assurance. How robust are the dependency relations we have adduced so far in counterfactual terms? That is, is it really impossible for us to engage in rationally required projects without engaging in modal thought? On the other hand, are we not overloading the notion of rationality if we make those projects required, and further, if we make it a requirement for those projects to engage in modal thought?

I think we can alleviate both worries. First, assuming that rationally required projects are constitutive of rationality (which would explain why we are criticizable if we don't engage in them), we cannot easily be rational and abandon them: so in relevantly similar possible worlds where we are rational, we are engaged in those projects either in the same way we are in the actual world, or in similar ways. The second question can perhaps be answered by appeal to the generality of the adduced rationally required projects (so that they are ongoing to some degree already as long as we have a minimal form of agency and guidance).

Beyond the *global* case for the indispensability of modal thought, it is worth thinking about the *local* indispensability of *parts* of it. If we can identify those, we can distinguish between different ways in which modal thought can be indispensable for epistemic and practical projects. Thinking on how these relations can be spelled out can, I will suggest, help us clarify certain debates concerning the epistemic support of certain disputed cases.

One of these is the case of *dispositions*. It is a platitude that dispositions have modal profiles: at the very least, the unrealized manifestations of an object's dispositions are possibilities for it. Recently, people have asked whether we can appeal to dispositions in order to provide explanations of interesting phenomena; the move has been particularly appealing for authors working in the philosophy of biology. A typical objection against this explanatory strategy is that it is possible to provide the same explanations, or some which are sufficiently similar, in terms of counterfactuals. However, from a conceptual point of view this objection faces the problem that dispositional notions cannot be captured fully by counterfactual analyses. In this sense, it might turn out to be the case that

dispositional talk is indispensable for these explanatory projects in biology.

Now, even the counterfactual case is undecided. In Williamson (2007), the development of counterfactuals requires to keep certain things fixed, making the minimum possible alterations to actuality. Among those, one finds *constitutive* (essential) facts. Against this, Roca-Royes (2012) argues, convincingly in my opinion, that essentialist vision is not a core part of counterfactual evaluation. If she is right, essentialist thought might be dispensable:

As far as we've been given reasons to believe, it *might* still be that essentialist and metaphysical modal thought can be 'removed from our conceptual scheme without collateral damage', contrary to what Williamson argues. (170–171)

We can capture the relevant dependencies using the following taxonomy of indispensability relations. If we carve out modal thought  $M$  and identify its set of roles/projects  $R$ , for any such carving out  $P$  we have:

**External Indispensability:**  $m$  is externally indispensable for  $M$  iff  $m \in P$  and  $m$  is indispensable for the realization of  $R$ .

In this case, it might be worth asking whether  $m$  refines  $M$ , that is, whether the claim that  $M$  is indispensable for  $R$  can be refined to a claim that  $m$  is indispensable for  $M$ .

**Internal Indispensability:**  $m$  is internally indispensable for  $M$  iff  $m \in P$  and  $m$  is indispensable for the realization of some of the roles of an  $m'$  such that  $m' \in P$  and  $m' \neq m$ .

We can now apply these notions to the cases we considered before:

- a) In the case of dispositions, the question is whether dispositional language/thought is externally indispensable for the explanatory project (as we suspected), *and* whether counterfactual language/thought is internally indispensable for whatever uses dispositional language/thought may have. If the second point is true, perhaps (if dispositional thought ungrounded in counterfactual evaluation is dispensable for the explanatory project) dispositional language could be dispensed with.
- b) In the case of counterfactual and essentialist thought, the question is whether essentialist thought is internally indispensable for counterfactual thought, and thus, whether essentialist thought is externally indispensable.

These observations could be further enriched in two ways. On the one hand, the fact that we can make several independent or orthogonal categorizations of modal thought suggests that we should introduce some notion of *cross-carving indispensability*. This suggests that questions of indispensability can turn out to be a matter of what classificatory schemes we should use. For example, in cases where different carvings of modal thought overlap we can ask whether some roles of  $M$  can be fulfilled in terms of one classificatory scheme instead of another, or whether several of those carving exhibit relations of mutual support.

On the other hand, we should take into account that the projects and function of modal thought may develop and change over time. In this sense, the questions



of indispensability can also be sensitive to historical factors.

To summarize: I have motivated the introduction of discussions concerning the functional role of modal thought through the use of the notion of indispensability. I have sketched a (defeasible) positive argument in favor of the global indispensability of modal thought, and addressed some objections. Finally, I have attempted to explain how finer-grained notions of indispensability can be useful to understand several current debates in modal epistemology. In doing so, I have moved from the issues concerning the implementation of modal epistemologies to architectonic problems that arise for any modal epistemology.

## References

- Brandom, Robert. 2008. *Between Saying & Doing: Towards an Analytic Pragmatism*. Oxford: Oxford University Press.
- Colyvan, Mark. 1999. "Confirmation Theory and Indispensability." *Philosophical Studies*, 96(1):1–19.
- Enoch, David, and Joshua Schechter. 2008. "How Are Basic Belief-Forming Methods Justified?" *Philosophy and Phenomenological Research*, LXXVI(3):547–79.
- Field, Hartry. 1989. *Realism, Mathematics and Modality*. Oxford: Basil Blackwell.
- Hale, Bob. 2013. *Necessary Beings: An Essay on Ontology, Modality, & the Relations Between Them*. Oxford: Oxford University Press.
- Panza, M., & Sereni, A. 2016. "The Varieties of Indispensability Arguments." *Synthese*, 193:496–516.
- Roca-Royes, Sonia. 2012. "Essentialist Blindness Would Not Preclude Counterfactual Knowledge." *Philosophia Scientiae*, 16(2):149–72.
- Stalnaker, Robert. 1996. "Knowledge, Belief, and Counterfactual Reasoning in Games." *Economics and Philosophy*, 12:133–63.
- Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell Publishing.

# Fake knowledge-how

Jesús Navarro

We are surrounded by impostors. Some of those are epistemic impostors: people who inappropriately hold—or deceitfully show—that they know, when they don’t. The kind of knowledge that some of those impostors pretend to have is knowledge-that—i.e. propositional knowledge, which aims at truth. But what some other epistemic impostors pretend to have is knowledge-how—i.e. a kind of knowledge that is more akin to abilities, capacities or skills than to the acquisition of truths. The focus of this paper is on the latter case of epistemic imposture. What does fake knowledge-how consist in? What makes it possible? What differentiates it from fake knowledge-that? In this exploratory paper, I aim to formulate those questions properly and to lay the foundations for their possible answers.

The paper has three parts, each one devoted to one of the aforementioned questions. In the first one—*What does fake knowledge-how consist in?*—I begin by introducing Edward Craig’s (1999) socio-genealogical approach to the concept of knowledge. The content of such concept, in Craig’s views, may be illuminated by understanding the function it accomplishes in our societies, which is to flag reliable sources of testimony for future informational exchanges. Our concept of knowledge would therefore have developed in order to track what Miranda Fricker (2007) calls “indicator properties”, namely, detectable properties of potential informants that correlate well with having a relevant true belief. I will hold that, even if Craig’s model is tailored to the concept of knowledge-that, it may also be applied to the concept of knowledge-how with some minor changes. I propose to focus on the role that the label “knowledge how” plays in a society where we systematically *entrust* others to perform tasks we care about. In order to navigate this network of trust, we are interested in searching for people who *know how* to do the thing—i.e., not only agents who luckily succeed in it, but those who are really reliable in attaining success and, in particular, those whose successful performances are particularly valuable. In this way, the first part of the paper establishes a parallel between knowledge-that and knowledge-how, which replicates in their respective *fake* varieties: epistemic impostors are those who either fake the property of being reliable sources of testimony, or the one of being reliable performers of tasks.

The second part of the paper—*What makes fake knowledge-how possible?*—inquires into the conditions of possibility for such imposture to occur. I do so by drawing extensively upon Katherine Hawley’s (2011) approach to knowledge-how. Craig and Fricker focus on what Hawley calls “upstream indicators” of true belief, which are those that show that agents were in the right conditions to obtain true belief (e.g. degrees, certificates, previous experience...). However, I follow Hawley in placing more emphasis on what she calls “downstream indicators” of true belief, which are typified by successful action of the sort that may

best be explained by true belief. When applied to knowledge-how, this distinction shows that we may attribute it either by considering the good position in which the agent was in, in order to obtain or develop certain capacities (e.g. his attendance to a good school, his previous participation in a prestigious team, etc.) or the excellence of her performances (e.g. how well she actually performs the task when casually observed, or upon request). My claim in this part of the paper is that fake knowledge-how may be the effect of two different mechanisms, depending on the sort of indicators that end up being deceitful, which allows me to distinguish upstream from downstream deceit—a distinction that I illuminate with the help of Wittgenstein’s (2009, §354) discrimination between symptoms and criteria (see Navarro 2010). Both are reliable indicators of some hidden property, but upstream indicators are based on mere symptomatic correlations (people in certain situations usually have the desirable property of true belief, or excellence in performance), whereas downstream indicators are criterial (i.e. being able to perform consistently well under the relevant conditions is what knowledge-how consist in). A putative confusion between both kinds of indicators may explain why some ‘fake experts’ sometimes signal their expertise in bizarre ways that are disconnected from actual performance—as, for example, happens with some political pundits that, when challenged on the grounds that they were not well informed, or that their presentation of topics was epistemically defective, would respond by signalling their own academic credentials.

The final part—*What differentiates fake knowledge-how from fake knowledge-that?*—shows that the differences between those two varieties of epistemic deceit run deeper than it seems at first sight because our position as knowledge-attributors is importantly quite unlike. A very salient feature of knowledge-that ascription is that we typically cannot directly detect true belief in a speaker just by the way she talks. This is the reason why examiners must have an independent access to truth in order to assess their pupil’s knowledge, and that is also what explains that testimony is valuable for inquirers, who typically lack such independent access. However, in the case of knowledge-how, as Hawley has convincingly shown, we typically have a way of assessing the excellence of the performance that does not rely on our having that piece of knowledge-how ourselves (or this is at least what sommeliers, art critics and literature editors like to think about themselves...). Hawley holds that that this possibility importantly affects our capacity to detect fake knowledge, with the effect that we are less prone to be deceived by knowledge-how epistemic impostors than by knowledge-that ones. Nevertheless, I will finally show that knowledge-how imposture may be even more resilient to detection than its propositional cognate, or at least resilient in distinctive ways. Motivation for this point is earned from the theoretical model we have advanced for the defeasibility of knowledge-how (Carter and Navarro, 2017).

## References

- Carter, J. Adam, and Jesús Navarro. 2017. «The Defeasibility of knowledge-how». *Philosophy and Phenomenological Research*, 95(3):662–685.

- Craig, Edward. 1999. *Knowledge and the State of Nature*. Oxford: Oxford University Press.
- Fricker, Miranda. 2007. *Epistemic injustice: power and the ethics of knowing*. Oxford: Oxford University Press.
- Hawley, Katherine. 2011. «Knowing How and Epistemic Injustice». In *Knowing How: Essays on Knowledge, Mind, and Action*, John Bengson y Marc A Moffett (eds.), 283–99. Oxford: Oxford University Press.
- Navarro, Jesús. 2010. «Speech Acts, Criteria and Intentions». *Lodz Papers in Pragmatics*, 6(1):145–170.
- Wittgenstein, Ludwig. 2009. *Philosophical Investigations*. Oxford: Wiley-Blackwell.

# Relevancia epistémica de las intuiciones

Manuel Pérez Otero

perez.otero@ub.edu

Universidad de Barcelona – LOGOS

[http://www.ub.edu/grc\\_logos/manuel-perez-otero](http://www.ub.edu/grc_logos/manuel-perez-otero)

Presento y defiendo determinadas hipótesis teóricas sobre las intuiciones, reivindicando su existencia y su relevancia epistemológica.

Una primera descripción, relativamente neutra, indica que cuando S tiene la intuición de que P, a S le parece verdadero dicho contenido (que P). En realidad, eso ocurre con cualquier proposición creída; si creemos P, nos parece verdad P. Todavía sin apartarnos mucho de la descripción poco controvertida, podemos elucidar algo ese “*parecer*” o impresión correspondiente a las intuiciones con una socorrida analogía: la intuición es una impresión, *intelectual*, que tiene similitudes (y diferencias) con las impresiones *perceptivas*. En condiciones normales, mirando una manzana roja nos parece estar viendo una manzana roja, mirando un palo rígido metido parcialmente en el agua nos parece estar viendo un palo quebrado. (Si sabemos que esa segunda impresión perceptiva es errónea, no creemos –sin embargo– lo que visualmente parece ser el caso.) Al tener una intuición también nos parece que algo es verdad, aunque esa apariencia no deriva de la percepción inmediata. (Me refiero a la intuición *intelectual* o *racional*; en ocasiones se usa “intuición” en un sentido muy amplio que engloba algunos casos de impresiones o apariencias derivados de la percepción inmediata.)

Mi propuesta invoca la noción de *justificación derivada de la comprensión de los conceptos constituyentes de una proposición*, tal y como sucede con otras teorías sobre las intuiciones, o sobre la justificación de verdades procedente de intuiciones. Las verdades así justificadas son las verdades conceptuales. Voy a asumir que entre tales verdades conceptuales hay una distinción (con toda probabilidad, una distinción gradual). Algunas son muy básicas, o inmediatas, o directas; la mera comprensión conceptual proporciona directamente o inmediatamente justificación para creerlas. Pero para otras verdades conceptuales, la justificación que tenemos para creerlas es –comparativamente– algo más indirecta.

Respecto a las proposiciones del primer tipo, mi enfoque está muy emparentado con tesis que –desde perspectivas diversas; a veces no apelando a la *comprensión de los conceptos* sino a la *comprensión del significado*– encontramos en textos de Wittgenstein (1953, 1958), Carnap (1952), Peacocke (1989, 1992, 1993), Sosa (1998, 2007), Goldman (2007) y Bealer (1998). Los aspectos específicos y originales de mi propuesta conciernen a cómo basándonos en la aceptación de ese primer tipo de proposiciones (aquellas directa o inmediatamente fundamentadas en la comprensión de los conceptos) las proposiciones del segundo tipo (también meramente conceptuales, pero no tan directas) nos *parecen* verdaderas y –si nada interfiere– las aceptamos. Es una propuesta sobre qué condiciones

acompañan a (o constituyen) esa impresión intelectual. (No obstante, mi análisis también permite describir la justificación del primer tipo de proposiciones como una justificación por intuición.) La idea directriz es ésta: en relación con tales proposiciones no tan directamente basadas en la comprensión conceptual, la apariencia de verdad del contenido intuitivo deriva de una estimación típicamente implícita sobre simplicidad; sobre la simplicidad que supone aplicar los conceptos (constituyentes de la proposición intuitiva como verdadera) a nuevos casos de conformidad con la proposición intuitiva.

Llamemos SC al sistema de creencias de un sujeto en un momento de tiempo  $t$ . Simplificando mucho, SC podría identificarse con el conjunto de proposiciones creídas por S en  $t$  (o con ese conjunto y una ordenación en él inducida por ciertas relaciones de dependencia entre tales proposiciones). Si P es una proposición, digamos que SC<sub>P</sub> es el P-*subsistema* de creencias (de S en  $t$ ), entendido del siguiente modo: SC<sub>P</sub> contiene todas las proposiciones de SC tales que la mera comprensión de algunos de los conceptos integrantes de P proporciona justificación inmediata *prima facie* para creerlas.

La afirmación sustantiva crucial que definiendo establece lo siguiente: la intuición de que P es identificable (o al menos se correlaciona) con una estimación (típicamente implícita) de simplicidad y/o de naturalidad. Más concretamente, propongo esta tesis (cuyas siglas están por *estimación de simplicidad*):

(ES) Si S tiene la intuición de que P es verdad, entonces

- (i) S tiene una inclinación primitiva a tomar como verdadero P directamente derivada de la mera comprensión de algunos de los conceptos integrantes de P, o bien
- (ii) S estima que el sistema de creencias resultante de añadir a su P-subsistema de creencias la proposición P es más simple y/o natural que el sistema resultante de añadir a su P-subsistema la proposición No-P.<sup>1</sup>

Dicha tesis establece sólo una condición necesaria para tener la intuición de que cierto contenido proposicional es verdadero. De forma más tentativa, sugiero también la siguiente definición:

(INT) S tiene la intuición de que P es verdad si y sólo si

- (1) S considera la cuestión de si P es o no verdad,
- (2) (2.i) S tiene una inclinación primitiva a tomar como verdadero P directamente derivada de la mera comprensión de algunos de los conceptos integrantes de P, o bien (2.ii) S estima que el sistema de creencias resultante de añadir a su P-subsistema de creencias la proposición P es más simple y/o natural que el sistema resultante de añadir a su P-subsistema la proposición No-P, y
- (3) la justificación que pueda tener S para creer que P es verdad (3.i) es inmediata y deriva de la mera comprensión de los conceptos que inte-

<sup>1</sup>Es típicamente implícita la estimación –por parte de S– de la simplicidad y/o naturalidad de extensiones de su P-subsistema de creencias. Obviamente, cuando de forma explícita y consciente S se pregunta si P es el caso, su estimación sobre el valor de verdad de P no es implícita.

gran P, o bien (3.ii) procede principalmente de la estimación indicada en (2.ii)

La cláusula (1) viene motivada por el hecho de que –según algunos filósofos– probablemente tener una intuición deba considerarse un estado “activado” [*occurent*], un estado en el cual no podemos estar de forma sólo latente o disposicional, en contraste con otros estados epistémicos o mentales (como creer, desear, saber, tener justificación, etc.). Quien discrepe puede considerar el resultado de retirar de (INT) dicha cláusula. Respecto a (3), se incluye para descartar que, por ejemplo, un caso sencillo de conocimiento perceptivo sea también un caso de intuición, pues probablemente satisface (2.ii).

Exploraré diversos aspectos de esa tesis, varios de ellos concernientes a la analogía con las impresiones perceptivas. Estas últimas tienen tres rasgos de interés:

- (a) son (al menos muchas de ellas) *persistentes*: al sujeto S puede seguir pareciéndole perceptivamente que P aunque crea que esa apariencia es una ilusión perceptiva;
- (b) proporcionan justificación para creer el contenido aparentemente percibido;
- (c) las impresiones perceptivas de dos sujetos ante una misma situación pueden diferir significativamente en la justificación que aportan (una de esas impresiones puede ser mucho más fiable que la otra), debido a factores muy variados (por ejemplo: los factores conducentes a requerir el uso de gafas, experiencia en la observación de radiografías adquirida en una facultad de medicina, etc.).

Veamos algunos rasgos de las intuiciones, conforme a la hipótesis que he presentado:

- (1) Que las intuiciones sean o no persistentes (de forma análoga a las impresiones perceptivas) es una cuestión controvertida. Es decir, no es obvio que tras haber intuido que P y rechazar (en virtud de alguna teoría) que P sea verdad podamos seguir teniendo la intuición de que P. Mi hipótesis es neutral respecto a eso.
- (2) Bajo ciertos supuestos plausibles, el hecho de que *S estime que su P-subsistema de creencias sería más simple y/o natural si incorporase P* le confiere a S (si ningún otro factor interfiere) justificación para creer P. Éstos son dos de tales supuestos:
  - (2.i) Los modos en que el uso de conceptos debe extenderse a nuevos casos (y por tanto la verdad o falsedad de tales aplicaciones) depende de la simplicidad y/o naturalidad de esas extrapolaciones. Eso se ejemplifica con casos sencillos de *seguimiento de reglas*: a la serie 1, 3, 5, 7, 9 le sigue el 11, aunque reglas menos simples establecerían que le sigue, por ejemplo, el 23. Igualmente: será incorrecto aplicar el concepto *árbol* para describir una piedra. Igualmente: será incorrecto aplicar el concepto *saber* para describir la situación epistémica de un protagonista de un caso Gettier.<sup>2</sup>

---

<sup>2</sup>La noción de simplicidad que invoco procede, en lo sustancial, de Lewis (1983). Creo que esa noción –o alguna otra que desempeñe un papel muy cercano a ella (naturalidad, semejanza

- (2.ii) Las estimaciones –por parte de los sujetos– de la simplicidad y/o naturalidad de extensiones de sus subsistema de creencias tienen, en general, un grado apropiado de fiabilidad.

Eso explica que las intuiciones confieran justificación dadas las hipótesis teóricas aquí defendidas. Pero el mero hecho de que las intuiciones sean (o se correspondan con) impresiones o apariencias también permite explicarlo, pues una apariencia (sea en el caso perceptivo o en el caso intuitivo) proporciona justificación *prima facie*.

- (3) Que las intuiciones (ante cierto contenido P) de dos sujetos, S1 y S2, pueden diferir significativamente en la justificación que aportan es una cuestión controvertida. Mi enfoque no es neutral respecto a eso. Las intuiciones de S1 y S2 pueden no ser epistémicamente equivalentes debido a dos tipos de factores diferenciales entre S1 y S2: (3.i) diferencias en sus respectivos subsistemas de creencias; (3.ii) que S1 y S2 no sean igualmente fiables al estimar la simplicidad y/o naturalidad de extensiones de tales subsistemas. Esa segunda diferenciación entre S1 y S2, (3.ii), puede derivar de factores muy variados, innatos o adquiridos, incluyendo diferencias significativas en la familiarización previa con tareas emparentadas (extraer conclusiones, explorar escenarios alternativos, detectar falacias, captar semejanzas, etc.). Por ejemplo, sujetos diferentes responden con capacidades diversas (con diferente fiabilidad, con diferente rapidez, etc.) ante los problemas planteados en un cuaderno con ejercicios sencillos de lógica. Cuando las intuiciones difieren debido a este factor, (3.ii), también difieren en el grado de justificación que aportan. En general –y como sucede con las impresiones perceptivas–, las impresiones intelectual-racionales podrían proporcionar mayor o menor justificación dependiendo del sujeto que las experimenta.<sup>3</sup>

## Bibliografía

- Bealer, George (1998). “Intuition and the Autonomy of Philosophy”, en M. DePaul y W. Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, Lanham, MD: Rowman and Littlefield, pp. 201–239.
- Carnap, Rudolf (1952). “Meaning postulates”, en R. Carnap, *Meaning and Necessity*, Chicago: The University of Chicago Press, 2ª edición, 1956, pp. 222–229. Publicado originalmente en 1952.

---

natural, etc.)– es crucial para establecer la diferencia entre aplicaciones correctas e incorrectas de un concepto; cf., por ejemplo, Pérez Otero (2008; 2016; 2018, cap. 5).

<sup>3</sup>Este punto es relevante para afrontar bastantes críticas (no todas) al papel epistémico de la intuiciones. Una variedad de objeciones, procedentes de la filosofía inspirada por encuestas –también denominada, de forma algo confundente, “filosofía experimental”–, presuponen que todas las intuiciones tendrían la misma validez epistémica. Pust (2017) responde a diversas críticas contra las intuiciones; pero no menciona este asunto (aunque varias de las objeciones que comenta asumen esa presuposición). Incluso si las encuestas realizadas desde tales enfoques estadísticos se llevan a cabo con rigurosidad y sin prejuicios (no al revés), su relevancia para objetar a unas u otras intuiciones depende de esta cuestión. (Cf. Ludwig 2007; Pérez Otero 2017.)



- Goldman, Alvin I. (2007). "Philosophical Intuitions: Their target, their source, and their epistemic status", *Grazer Philosophische Studien*, 74(1):1–26.
- Lewis, David (1983). "New Work For a Theory of Universals", *Australasian Journal of Philosophy*, 61:343–377.
- Ludwig, Kirk (2007). "The Epistemology of Thought Experiments: First Person versus Third Person Approaches", *Midwest Studies in Philosophy*, 31:128–159.
- Peacocke, Christopher (1989). "What are Concepts", *Midwest Studies in Philosophy*, 14:1–28.
- Peacocke, Christopher (1992). *A Study of Concepts*. Cambridge, Mass.: MIT Press.
- Pérez Otero, Manuel (2008). "The Humean problem of induction and Carroll's Paradox", *Philosophical Studies*, 141:357–376.
- Pérez Otero, Manuel (2016). "Respuestas disposicionalistas al problema wittgensteiniano-kripkeano sobre el significado y el seguimiento de reglas", *Crítica. Revista Hispanoamericana de Filosofía*, 48(143):21–45.
- Pérez Otero, Manuel (2017). "Teorías de la referencia, *filosofía experimental* y calibración de intuiciones", *Theoria*, 32(1):41–62. DOI: 10.1387/theoria.15463.
- Pérez Otero, Manuel (2018). *Vericuetos de la filosofía de Wittgenstein en torno al lenguaje y el seguimiento de reglas*. Zaragoza: Prensas de la Universidad de Zaragoza.
- Pust, Joel (2017). "Intuition", *The Stanford Encyclopedia of Philosophy*, Edición de Verano de 2017, E. N. Zalta (ed.), <https://plato.stanford.edu/archives/sum2017/entries/intuition/>
- Sosa, Ernest (1998). "Minimal Intuition", en M. DePaul y W. Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, Lanham, MD: Rowman and Littlefield, pp. 257–269.
- Sosa, Ernest (2007). "Intuitions: Their Nature and Epistemic Efficacy", *Grazer Philosophische Studien*, 74(1):51–67.
- Wittgenstein, Ludwig (1953). *Philosophische Untersuchungen*. Publicado originalmente en 1953. Traducción: *Investigaciones filosóficas*, Barcelona: Crítica, 1988, edición bilingüe.
- Wittgenstein, Ludwig (1958). *The Blue and Brown Books*. Oxford: Basil Blackwell. Traducción: *Los cuadernos azul y marrón*, Madrid: Tecnos, 1968.

# Social virtues upside down: Group epistemic agents as sources for individual epistemic virtues

Dani Pino  
Universidad de Sevilla

According to Robust Virtue Epistemology, a belief qualifies as knowledge if and only if the belief is true because of the subject's exercise of intellectual virtue or ability. Such achievement is implicitly understood as resulting from actions that are performed by an individual who manifests her competences as a knower. However, contrary to individualistic accounts of action, group agency approaches (Gilbert 2004, 2009; List & Pettit 2011; Pettit 2005; Tollefsen 2002, 2004, 2006, 2015) hold that collectives may count as entities whose actions are not reducible to those performed by group members, nor to the mere aggregation of their contributions. Given this, one may ask whether group agents fit in the picture of Virtue Epistemology.

In his "Group Virtue Epistemology", Kallestrup holds that, while a group may form apt beliefs none of its members share, the innermost competences of the group are «nothing over and above those of its members when suitable combined» (2016: 17). This leads to «a crucial disanalogy» between the group and its members in relation to the competences required in order for a true belief to count as knowledge. In a nutshell, if Kallestrup is right, a group epistemic agent *can* know in its own, but its competences (which are the condition for its virtues) depend on those of the individual members from whom these group competences supervene.

In order to motivate his position, Kallestrup's idea of group agency relies on Schmitt's notion of *chartered group* as «founded to perform a particular action or actions of a certain kind [...] with the understanding that the group will perform only such actions» (1994: 273). In his view, Kallestrup describes the agency of such a group as constrained by an organizational and intentional structure (i.e. an office) and the way this structure is «determined by the relevant charter (i.e. rules, norms, standards)» (2016: 9). If one of the defining targets of the group is knowledge, we can talk about a group epistemic agent. However, the rules, norms and standards that comprise the group's charter do not constitute by themselves genuine group competences in Kallestrup's account. On the contrary, these supervene on its member's innermost competences, which are to be understood as the structural seat for a skill to be performed (Sosa, 2015). The problem of such an understanding of competences is that the structural seat that matters seems to be necessarily identified with *what* each of the individuals has in her mind. In short, this account seems to impose that in the game of manifesting competences as a virtuous cognitive agent, if there is no brain, there is

no game. Hence, since there is no group brain, but group's member's brains, no innermost competences might be attributable to a group, but to its members.

In this paper, I challenge Kallestrup's view that only individuals may have epistemic innermost competences. In contrast to that, I hold that groups may also have innermost competences required to explain social epistemic virtues. In order to defend this, I will make use of arguments by Lahroodi (2007) and Goldberg (2009) regarding collective and social virtues.

On the one hand, Lahroodi introduces a taxonomy that distinguishes "social groups" (intimacy groups or task groups) from mere "populations" (social categories or loose associations), a distinction that is explained in terms of two properties: "entitativity" and "group member interaction" (2007: 284–285). The former is a property groups have resulting from the coherence of its members, whereas by the latter Lahroodi understands the degree of relations between the individuals that are gathered in the group. Furthermore, group member interaction is a requirement that, in my opinion, may be illuminated by applying Dynamical Systems Theory to group cognition, in particular the idea of a «*coupled system*» that is due to «mutual (i.e., non-linear) relations [...] between the contributing parts» (Pritchard & Palermos 2016). The conjunction of both requirements, entitativity and group member interaction, results in the emergence of collective traits that face a kind of disanalogy that is similar to the one pointed out by Kallestrup, though not totally identical. The difference between both disanalogies relies on what Lahroodi calls an «anti-correlative account of collective traits and virtues» (2007: 290), according to which «members do not have to jointly accept to exercise a trait for the group to have that trait» (2007: 292). This idea suggests that a social feature, such as a virtue that results from a group competence, does not have to result from the contribution of individual members, which is contrary to the view that Kallestrup holds.

On the other hand, I employ two accounts of social virtues by Sanford Goldberg in connection with testimonial knowledge. The first one concerns the roles that individual virtues play in the social routes to knowledge, whereas the second one is related to the individuation of social virtues (2009: 238). According to the former, one of those social routes relates testimonial knowledge in so far as «if my knowledge that *p*, acquired through my having accepted your testimony that *p*, amounts to a cognitive achievement, surely I cannot claim the whole of the achievement» (2009: 242). The relation between the individuals involved in testimony cases plays a role by itself in the manifestation of some kind of competences that are not properly reducible to the members of such relation. So to the extent that this *social* role may be highlighted as significant in cases of testimonial cases, so it is as a competence that is not exclusively seated in neither of the participants.

Regarding the latter, Goldberg claims the need to «regarding social factors as relevant to *the very individuation* of the social virtues themselves» (2009: 245). This account sets out a comparison of the cases of social virtues with cases of perceptual virtues according to «a parallelism between two divisions, inner vs. outer, and cognitive vs. brutally causal» (2009: 246). The comparison establishes a dichotomy of two terms, one of which, that the external factors in the belief-forming process are to be seen as merely background features and conditions, but not primarily relevant in attaining the cognitive achievement that counts

as virtuous performed, is rejected by Goldberg. The rejection rests on the claim that «the division of the process into the parts that take place *within vs. without* the subject's own mind/brain does *not* line up with the division between the *cognitive vs. brutally causal* part of the process» (ibid.), because «the virtues are *virtues* insofar as they conduce to successful performance, but successful performance in connection with perception is success in being attuned to perceptual features of our environment» (2009: 247). In other words, the inner processes involved in perception do not exhaust what it is required in order to individuate a perceptual knowledge achievement, addressing to an «externalism regarding the individuation of the perceptual virtues» (ibid.) , just as the individuation of social virtues requires «that processing bears the right sorts of relation to the perceived “external” environment» (ibid.).

I benefit from both accounts in order to argue for an anti-individualistic understanding of virtues in which the production of knowledge may represent a «“social” cognitive achievement» (2009: 242) resulting from social conditions, which might count as competences.

Combining both arguments, I conclude against Kallestrup's view by arguing that social competences would not be the mere result of the combination of their member's competences, but genuinely and irreducibly group virtues. My claim is that group innermost competences, in spite of their dependence upon the individual competences of its members, are to be identified with the kind the relations that matters for a group to count as such. And this view is a challenge to Kallestrup's one to the extend that it contributes to a really anti-individualistic approach to the very nature of group agency. In brief, if a group agent genuinely manifests group virtues while performing its innermost competences as an epistemic agent, then the group would take into account its own competences in order to attain its achievements and, thereby, opening a new understanding of the group responsibility. Moreover, I defend that the virtues of a group are in general not the result of the upstream-driven contributions of individual virtues, but the reverse: at least some of the individual epistemic virtues of the members of a group may only be acquired by their belonging to intimacy and task groups. Consequently, the more virtuous and competent the group is, the more virtuous its members may become.

## References

- Gilbert, M. (2004). Collective epistemology. *Episteme*, 1(2):95–107.
- Gilbert, M. (2007). Collective Intentions, Commitment, and Collective Action Problems. In Fabienne Peter & Hans Bernhard Schmid (Eds.), *Rationality and Commitment*. Oxford: Oxford University Press. pp. 258.
- Gilbert, M. (2009). Shared intention and personal intentions. *Philosophical Studies*, 144(1):167–187.
- Goldberg, S. (2009). The Social Virtues: Two Accounts. *Acta Analytica*, 24(4):237–248.
- Kallestrup, J. (2016). Group Virtue Epistemology. *Synthese*:1–19.

- Lahroodi, R. (2007). Collective Epistemic Virtues. *Social Epistemology*, 21(3):281–297.
- List, C. & Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press.
- Palermos, S. O., & Pritchard, D. (2016). The Distribution of Epistemic Agency. In P. Reider (Ed.), *Social Epistemology and Epistemic Agency: Decentralizing Epistemic Agency*, MD: Rowman & Littlefield:109–126.
- Pettit, P. (2005). Group Agency and Supervenience. *Southern Journal of Philosophy*, 44 (Supplement):85–105.
- Schmitt, F. (1994). The justification of group beliefs. In F. Schmitt (Ed.), *Socializing epistemology: The social dimensions of knowledge*. Lanham, MD: Rowman and Littlefield:257–287.
- Sosa, E. (2015). *Judgment & Agency*. Oxford: Oxford University Press UK.
- Tollefsen, D. (2002). Collective intentionality and the social sciences. *Philosophy of the Social Sciences*, 32(1):25–50.
- Tollefsen, D. (2004). Collective epistemic agency. *Southwest Philosophy Review*, 20(1):55–66.
- Tollefsen, D. (2006). From extended mind to collective mind. *Cognitive Systems Research*, 7(2):140–150.
- Tollefsen, D. (2015). *Groups as Agents*. Cambridge: Polity.

# An hybrid ontology for musical works: Between perceptualism and epistemicism on musical understanding

Pablo Seoane Rodríguez  
USC

Against the idea that musical understanding or, to be more precise, our knowledge of a musical work, is conditioned by the ontological commitments we assume regarding to the object we deal with, authors like Aaron Ridley (2004) or James O. Young (2014) have claimed that ontology is a scarcely valuable enterprise as a part of the philosophy of music and, accordingly, it must be abandoned in the interest of more useful disciplines, like aesthetics. In opposition to these assumptions, this paper defends an ontological status of musical works that, unlike other widespread theoretical stances (like Platonism, nominalism and fictionalism), allows a more rewarding epistemic approach to music: namely, the musical work conceived as a hybrid entity made up of a set of instructions plus a performance.

Firstly, we will face Ridley's claims, noting that his dismissal of any kind of musical ontology—as Kania (2008) has noted—is flawed from the start by a misconception of ontology's purpose. We think that ontology shouldn't determine the *content* of a musical work—as Ridley argues—but establish what kinds of musical entities there are and which relations hold between them. Furthermore, from an alleged position of ontological neutrality (which is not that neutral), Ridley articulates a rather vague or even obscure view of *content*, partially grounded on his refusal of the identity conditions of the musical work and a problematic equivalence between the faithfulness to the work and the aesthetic value. We will show how this notion of *content* doesn't warrant a suitable epistemic relation with musical works and why, at any rate, it would never affect, as Ridley purports, ontology status broadly speaking, but merely, as we'll see, the so-called hard ontology.

We will argue as well that Platonism about musical works (both the strong version and the weak one) is not an appropriate guiding principle to account for musical understanding either, because is weighed down, among other issues, by the problem of the epistemic access to the work understood as an abstract object independently of its performances: a problem closely linked to the *werktreue* paradox or the faithfulness to the work that a performance should keep. This is a question that Ridley analyses properly to a certain extent, but from which he draws, in our opinion, misguided consequences.

Secondly and intimately related to the issues brought about by the different attempts to articulate work and performance, we'll examine the notion of twofoldness or Wollheim's theory of experiential access to certain artworks (1980),

recently applied to musical works by authors like Andy Hamilton (2007, 2009) or Bence Nanay (2012), according to which our experience of the work would be a compound of an alleged perceptual appraisal (the *surface*) and an epistemic one (the *scene*) whose objects are, respectively, what in pictorial terms we could state as the stroke and the represented by the stroke. We'll review as well the debate between positions akin to perceptualism, like Levinson's concatenationalism (2006), epistemicist theses, like Kivy's architectonicism (2001), and conciliatory attempts between both extremes, like those carried out by Eggebrecht (2004) and Huovinen (2011).

We defend that this device of *twofoldness* shouldn't lead us to perpetuate the realist conception of work/performance as a hierarchical arrangement which favours abstraction at the expense of the concrete object, an implicit manner, for instance, in the text of Nanay, who seems to take the separation between work and performance for granted. We think that this is a clear and regrettable vestige of romanticism and shows a subordination to the so-called *textualist paradigm* (Cook 2013). To be more specific: As Lydia Goehr claims in her *thesis of 1800*, towards the end of the 18th century it sprouts a concept of musical work understood as an autonomous entity (Ridley 2004), independent, among other things, of the performance; that is to say, the idea that 'tonal, rhythmic, and instrumental properties of works are constitutive of structurally integrated wholes that are symbolically represented by composers in scores' (Goehr 1992: 2). And, precisely, the score, despite being born an ancillary object, a memory aid and an instrument of preservation, is going then to replace the musical event as the central object of considerations about music. A fascinating movement in which the symbolized object is devoured by the symbol, and whose main consequence is the study of music as writing, as the musicologist Nicholas Cook has shown (Cook 2014: 1). Admittedly, although we all use to link naturally music with the sense of hearing, it's curious to check out how, in virtue of the establishment of this textualist paradigm, musical studies during the last 200 years have consisted mainly in a textual scrutiny. To sum up, we have received an idealist conception of the musical work which postulates a disembodied and intangible entity of which we get a glimpse thanks to the textual codification, and that travels, according with the Romantic doctrine of genius, from the glorious mind of the composer to our ears by means of the always degraded and insufficient intervention of the performer. Paraphrasing Walter Benjamin, the concrete musical event becomes the death mask of the musical work, the material excrescency of an abstraction.

In the light of the above, we don't see any epistemic or aesthetic benefit in fixing a correspondence between both constituents of twofoldness and these two musical objects—work and performance, as Nanay does. We find difficult to admit, however figuratively, that the performance is merely the *stroke* of the pair, the *surface*, i. e., just a vehicle to reach the work. And if it is really more than that, that is, if performance is the titleholder of its own *surface* and *scene*, as I deem reasonable to believe, then a twofoldness whose horns are performance and work are pointless or redundant. And we should consider, therefore, more rewarding forms of applying twofoldness to music, like Andy Hamilton's twofold thesis (2007, 2009) a theoretical approach construed mainly in opposition to Roger Scruton's acousmaticism (1997).

According to acousmaticists like Roger Scruton, our experience of music consists in listening to the sounds severed from their sources or material origins. We have to engage, in other words, in the organisation of sound through a process of *getting inside* and, at the same time, beyond the sounds merely perceived. Scruton plainly identifies both elements of a possible musical twofoldness, the *surface* and the *scene*, but denies its possibility by rejecting the necessity of any attention to the first element of the pair in order to have a proper musical experience. Thus, acousmaticism accepts the existence of non-acousmatic features in every musical sound we hear, impure ingredients that point at gestures or images related to the sources of sound (musicians, instruments, and any other elements belonging to the sound environment), but claims that ‘only the acousmatic is a genuinely musical aspect of musical experience’ (Hamilton 2009: 160). On the opposite side, Andy Hamilton (2007, 2009) endorses a musical *twofold* thesis, according to which we need both *surface* and *scene*, the non-acousmatic and the acousmatic, that is, we engage in a process of ‘hearing-in’ to get an adequate musical experience. One of the reasons that grounds this thesis, Hamilton argues, is that certain central or inalienable properties of musical sound can be seen as non-acousmatic. For example, *timbre*, that is, the feature that permits the listener to identify the material source of sound, presents a problem for the acousmaticist, insofar as it involves a complex of auditory phenomena, which includes the noise elements that usually come along with tones, or the *attack*, the initial production of a musical sound by the friction, hitting, blowing, etc., of the performer on the material objects used as sources of sound. Commonly, some of these features seems to link our listening with sight or touch—with an extra-sonic world. Consequently, says Hamilton, experience of music is not in fact a purely auditory phenomenon, and an appeal to sensorial multimodality, to the fact that other senses affect what we hear, is here used to support the non-acousmatic value of music. We don’t even need to think about the easy visual experience of a live performance; the point is that a simple aural experience always involves the contribution of other senses as long as ‘the input of different perceptual channels cannot be absolutely separated’ (Hamilton, 2009: 178). Taking seriously the question of multimodality, according to another critic of Scruton’s acousmaticism, Rob Van Gerwen, involves considering sounds as ‘events as heard’ (Van Gerwen 2012: 226). Thus, it seems plausible, and more akin to our musical practices, to consider a musical experience as a phenomenon which involves an awareness of both *surface* and *scene*, or, in Hamilton’s terms, to a literal and sensuous dimension (non-acousmatic) and a metaphorical dimension (acousmatic).

After this review of two different kinds of musical *twofoldnesses*, we’ll outline briefly the dispute between Jerrold Levinson’s *concatenationism* and Peter Kivy’s *architectonicism*. Levinson has defended a perceptualist approach to musical understanding, maintaining that the listener engages in an accomplished musical experience by paying attention to perceivable effects on concrete points, and establishing connections and implications on a small scale, in virtue of what he calls *quasi-hearing*; as a consequence, the genuine listener shouldn’t be aware of large scale structures, as long as we cannot perceive them directly, but through reflection and analysis. As a matter of fact, the pleasure of detecting or recognizing huge patterns, says Levinson, is more intellectual than perceptual, and therefore a minor one. On the opposite side, Peter Kivy’s *architectonicism*



defends the so-called epistemicist manner, a proposal that emphasizes the importance of grasping large structures over the short-term perception. Of course, both theories have to make concessions to the other part, and a few authors have tried to reconcile the extremes by lowering the zeal of impractical commitments. It is not, however our purpose to delve into this dispute. In general terms, It may seem reasonable to accept that there is a perceptual grasping of music, perhaps the cause of what Hamilton calls the ‘sensuous pleasure’ of the listener (Hamilton 2009: 169), and surely we might admit also the existence of a certain amount of intellectual grasping of formal structures. In view of all that, what are we talking about then when we talk about an intellectual grasping of music? Formal training? Or just a moderate awareness of the cultural tradition to which a particular music belongs?

So, thirdly, and as a way to answer or at least to understand properly what are we really asking when we formulate questions like the previous ones, we will try to draw the elusive line between the epistemic and the perceptual (if such a line exists or if it is worthwhile to be drawn at all), and make as well an attempt to distinguish an epistemic experience of music from a proper musical understanding. We can distinguish two questions here—we’ll see whether or to what extent both are ultimately one and the same: what is a musical experience? And, what do we get to know, if anything, when we appreciate a musical event? Regarding the first one, I believe that there is not a non-epistemic musical experience and that musical experience is mostly an ability knowledge or know-how, rather than a propositional one—although the latter of course also matters, albeit in a lesser way. That is, we couldn’t reduce completely an instance of musical knowledge to a set of propositions about this or that piece or musical event. My rejection of the possibility of a non-epistemic musical experience means therefore that there is no *musical perception*, strictly speaking, although musical experience is to be founded necessarily on a multimodal perception (perhaps auditory to a large extent, nonetheless). In other words, musical experience is already in itself a *post-perceptual* phenomenon. I submit consequently that listening to music is not perceiving music. We do not perceive music. We perceive sounds. But when we experience sounds as music we have an epistemic and aesthetic experience. Therefore, perception is not aesthetic yet, as it is (musical) knowledge. Note that a member of another species or even a human being arrived from a distant cultural milieu could perceive the sounds that make up what we consider a common musical piece without having a musical experience.

What do we know, therefore, when we apply our musical twofoldness? My scarcely exhilarating answer is that we *know* being in front of an object or event endowed by our cultural environment with a set of properties or attributes that makes this object or event susceptible of aesthetic appreciation, leaving aside now the multiple social functions that could come along with aesthetic experience. ‘A major question in philosophy of perception, says Bence Nanay, is about which properties are perceived and which ones are inferred or non-perceptually represented’ (Nanay 2018: 55). I think that the relevant properties of the musical object are inferred in virtue of the exercise of an aesthetic knowledge, that is to say, a both epistemic and aesthetic experience prompted by twofoldness. But if we admit that a musical event only can be properly appraisable inside a determinate framework, musical understanding cannot be reduced to the sheer epistemic experience of a musical event. Rather, it must be properly linked to a

set of norms or regulated practices. Twofoldness provides us with the experience of attending to the phenomenic and formal musical features: recurrences, symmetries, contrasts, countless kinds of nuances and temporal patterns, as well as the extra-sonic traits involved in the production of sound. But this knowledge does not amount to a proper understanding yet, since some of these things could remain undetected or misvalued if the listener were not able to locate the piece or event at issue within the right tradition or normative space (forms, genres, styles). This ability, that doesn't need to be totally conscious or deliberate, permits us to be aware (and somehow operate from) a sort of hierarchical set of relations and properties, that is, it indicates to us which is the nature of the object we are getting to know and appraise, or, in other words, which are the main features we have to attend to in order to understand it. Whether these traits and the categories to which they are linked are aesthetic or ontological, or both, is an issue we cannot pursue here. After discussing the virtues and shortcomings of Wittgenstein's (musical) intransitive understanding (*understanding as placing things together*) we end up by launching an open and exploratory proposal which involves in one way or another a rebuttal of formalism and the abovementioned autonomy of musical works, and, a proposal consequently akin to non-propositional forms of representationalism.

In this vein and in fair agreement with our ontological proposal (to remind: the performance, plus a set of instructions, *is the work*), we'll finally defend a third type of *twofoldness* in which both elements of this device act onto the musical event in a way akin to Kendall Walton's make-believe theory of fiction (2015) (See also Kania 2015). In this manner, we conceive twofoldness as a sort of articulation of two different kinds of attention, an articulation which may come about in a joint or an alternate fashion. On the one hand, the attention the listener pays to the props (*prop-oriented make believe*), that is to say, the attention to the techniques, actions and instruments involved in the production of sound. And, on the other hand, the attention the listener pays to the game (*content-oriented make believe*), or, in other words, the heed she pays to the musical fiction. (It should be noted that this usage of certain aspects of Walton's theory doesn't commit us by any means to his fictionalism). A slight survey of Walton's musical representationalism will be decisive to determine the purported relevance of twofoldness for a substantial musical understanding. The advantages of Walton's theory regarding the epistemology of music, and also musical understanding, could remain notably clouded by the deep-rooted reluctance to consider music, any music or, at least, *most* music, and not just programme music, as a representational art. Or even, and this sounds still more problematic, as a sort of fiction. Besides the countless hassles that plague representation and fiction as concepts, I think that in the work of certain philosophers it seems to linger an unfair identification of representation with semantic or propositional representation, in order to stage an idle confrontation of the latter with a stale formalism (see Kivy 1991, 2009, for instance). But, as in Hamilton's twofold thesis, with Walton's theory we are able to distinguish between the literal (the props or non-acousmatic) and the metaphorical (the content or the acousmatic) sides of music. Accordingly, Walton's theory preserves the spirit of the twofold thesis, but also adds to Hamilton's view a depth—the fictional dimension—that deserves to be explored.

That we decide the *content* in the concrete sound event or, in other words, that

understanding, conceived as an adequate epistemological access, is accomplished on the performance and not on the work as an abstract entity, is for Ridley a triumph of aesthetics over ontology. But, actually, it only entails a rebuttal of Platonic ontology. Our ontological proposal of a work construed as a hybrid, artefactual entity, an entity undetachable from performance, is in this sense a privileged theoretical space from which we can aptly account for the understanding of musical works, as long as it avoids several lasting concerns that afflict the main contemporary theories of musical ontology, namely, the epistemic access to abstract entities or the so-called wrong note paradox. Moreover, this proposal hopes to make clear that there is no substantive incompatibility between perceptualism and epistemicism and that any sound scrutiny of musical understanding should incorporate elements of both theoretical standpoints. We won't conclude, however, without giving a response to possible objections in relation to ontological prodigality or the individuation problems that our proposal might involve.

# La discusión en torno al estatus epistémico de la doctrina filosófica del falibilismo

Edgar Serna  
USC

La pregunta general, sobre la que versa esta ponencia, es la siguiente: ¿cuál es el estatus epistémico de la dimensión normativa de la epistemología en la tradición de la filosofía analítica? En lo particular, la pregunta cuyas posibles respuestas analizaré aquí, se refiere a *cuál* es el estatus epistémico de la doctrina del *falibilismo*. La importancia que tienen estas preguntas radica en que la solución que se brinde a las mismas permitiría dilucidar *cuál* es el estatus epistémico de la teoría del conocimiento.

Al respecto, y al margen de la filosofía analítica, la respuesta tradicional se cristalizó en la defensa del programa metafísico en pos de un fundamento irrefutable del conocimiento, que pudiera alcanzarse mediante el cultivo de una filosofía autónoma, independiente y, además, distinta de cualquier ciencia particular. De acuerdo con lo anterior, la dimensión normativa inherente a la epistemología se concibe como *irreducible* a un ámbito científico concreto. Por ejemplo, este proyecto fue emprendido por Descartes, Kant y Husserl. Todavía Apel (1991 [1987]), después del giro lingüístico (cf. Quine 1981 [1975]), buscó dar cuenta del estatus epistémico del falibilismo con base en el programa metafísico en pos de una filosofía primera. Durante el siglo pasado, este ambicioso proyecto fue severamente criticado por Quine (1981 [1975]; 1995), Rorty (1979) y, muchos años antes que ellos, por Neurath (1983a [1931]). Cabe mencionar que este último, como promotor y representante de la incipiente naturalización de la epistemología, le reprochó a Popper seguir atado a esta añeja posición.

No obstante, defendiendo la tesis de que esta específica objeción que Neurath (1983b [1935]) hizo en su reseña de *Logik der Forschung* es, con todo, errónea. Nada hay en ese libro que sustente el proyecto de proporcionar una filosofía primera. Para entender cuál es el estatus epistémico que para Popper tiene la teoría del conocimiento y la reconstrucción racional del cambio científico que brinda al respecto, sostengo, a continuación, que es indispensable (a) estudiar el vínculo que dicho problema tiene con su temprana filosofía de las ciencias sociales (verbigracia, Popper 1961 [1944/45]; 1976 [1961]), examinando atentamente qué argumento le permite justificar la distinción que traza entre las humanidades y las ciencias de la naturaleza. A su vez, planteo que (b) es forzoso tomar en cuenta la forma como, dicho autor, explica cuál es el estatus epistémico de la doctrina filosófica del falibilismo para advertir hasta qué punto es deudor de una teoría no cognitivista de la ética. Considero que ambos aspectos de su filosofía permiten entender que, para Popper, la autonomía de su reconstrucción racional del

cambio epistémico con respecto a una investigación científica concreta, se basa, tanto en:

- Una naturalización *sui generis* de la epistemología (sustentada en lo que, en su filosofía de las ciencias sociales, caracteriza a las disciplinas humanísticas y las distingue de áreas como la física o la astronomía), como
- En la concesión que hace a una teoría *emotivista* de la ética al responder la pregunta:  
¿Por qué debería elegirse la actitud racional (el falibilismo) y, por consiguiente, la racionalidad dialógica o argumentativa?

Si estoy en lo correcto, este análisis nos permite aclarar por qué la posición de Popper al respecto se puede situar en el contexto de los intentos de explicar aquella autonomía, en lo general, y el estatus epistémico del falibilismo, en lo particular, desde el panorama de una formulación *débil* (o *parcial*) de la epistemología naturalizada (verbigracia, Habermas 1990 [1981]).

La estructura de la argumentación es la siguiente. En primer lugar, enumero cuáles han sido los principales motivos debido a los que una definición falibilista del conocimiento ha sido ampliamente aceptada en la epistemología contemporánea. Acto seguido, indago *qué* significa ‘falibilismo’ y distingo dos acepciones de dicho concepto: una *axiológica* y otra *epistemológica*. Expongo, en términos generales, algunas definiciones que se han propuesto de lo que significa ‘falibilismo’ (por ejemplo, las de Peirce, Dewey, Haack y Rescher) y analizo brevemente el debate acerca de cuál es el estatus de la doctrina del falibilismo en su acepción propiamente axiológica o valorativa. En este contexto discursivo, explico por qué es una definición falibilista del conocimiento lo que constituye el núcleo de la propuesta metodológica de Popper y, por ende, de la reconstrucción racional del cambio científico que propuso. Después, brindo una exégesis de los dos principales motivos por los que, al defender su criterio de demarcación, el cual identificó con una *propuesta normativa de convención* (cf. Popper 1959 [1935]), afirmó que, incluso cuando las opiniones acerca de si el mismo es, o no, adecuado, puedan discrepar, y admitiendo que una discusión razonable de tales asuntos

...únicamente es posible entre partes que tengan un propósito en común, [habrá de quedar claro, con todo, que] la elección de semejante propósito habrá de ser materia de una decisión que, por supuesto, va más allá de cualquier argumento racional (Popper 1959 [1935], 37).<sup>1</sup>

El motivo de lo anterior es que, debido a la índole *autorreferencial* de la pregunta sobre el estatus epistémico del falibilismo, Popper tuvo que responderla situándola en el contexto de la acepción *ética* (y no epistemológica) del mismo. Una vez que se acepta que *todas* las proposiciones de la ciencia son falibles y, a su vez, tan pronto como se admite que la teoría del conocimiento *forma parte* de la ciencia... ¿en qué situación se encuentra, entonces, una epistemología naturalizada cuyo núcleo lo constituye la doctrina filosófica del falibilismo? ¿Acaso la tesis de que *todas* las proposiciones de la ciencia son falibles (por lo

---

<sup>1</sup> «...is only possible between parties having some purpose in common [it would be clear that] the choice of that purpose must, of course, be ultimately a matter of decision, going beyond rational argument» (Popper 1959 [1935], 37 –la traducción es mía).

que, tarde o temprano, habrán de ser abandonadas y sustituidas por otras) es, ella misma, falible? Si se contestara afirmativamente, cometeríamos una petición de principio, pues estaríamos dando por sentado lo que intentamos probar. Asimismo, tampoco se podría suspender la duda con respecto a tal reserva falibilista sin caer en el dogmatismo (pues, ¿por qué interrumpir el examen sobre el problema en un punto determinado, en vez de proseguir con la búsqueda de tal justificación?). Y, en fin, no es posible persistir en la búsqueda de lo que sustenta la aceptación del falibilismo sin caer en una regresión al infinito de las justificaciones.

# Implicit bias, accessibilism and justification

Josefa Toribio  
ICREA-UB

## Introduction

Accessibilism is the view that only consciously accessible factors are relevant to epistemic justification. It has been recently argued (Puddifoot, 2016) that the justification of beliefs formed indirectly as a result of implicit biases, i.e., beliefs formed as the result of how things seem to us given implicit biases' influence on the available evidence, pose a problem for accessibilism—since implicit biases are consciously inaccessible, yet they seem to be relevant to epistemic justification. In this talk, I set out to do three things. First, I focus on how best to understand the way in which implicit biases are said to be inaccessible. I do this by reviewing some recent empirical evidence which suggests that, while we may typically lack conscious access to the source of implicit attitudes and their impact on our beliefs and behaviour, we do have access to their content (Gawronski et al. 2006; Hall & Payne 2010). Second, I discuss the notion of accessibility required for Puddifoot's argument to work in light of the reviewed empirical evidence and argue that accessibilism could meet the challenge posed by implicit biases in at least two ways. Finally, I show that these versions of accessibilism only get us out of the implicit bias challenge by positing an implausibly over-intellectualized and over-reflective subject. Although ultimately both strategies fail, the way in which they do, I conclude, reveals something general and important about our epistemic obligations and the role of implicit biases in (accessibilist) justification.

## Accessibilism

Accessibilism is a variety of internalism. According to accessibilism, whether or not a belief is justified depends solely on factors that are consciously accessible to the subject—typically, the mental states a subject can reflect about (see e.g. Chisholm 1988; Steup 1999). Accessibilism is a thesis about propositional justification, i.e., about which propositions a subject is justified to believe given the available evidence. It is not a thesis about doxastic justification, i.e., about a subject's justifiably believing what she does. One can have good reasons  $R$ , and thus be propositionally justified to believe  $p$ , even if one does not believe it or believe it for reasons other than  $R$ .

## Implicit bias and accessibilism

Implicit biases or implicit attitudes (henceforth, I take the two expressions to be synonymous) are representational mental states that reflect stereotypical properties of members of, and items in, all kinds of different categories: racial groups, professions, women, nationalities, members of the LGBTQ community, moral and political values, etc. They typically connect one or two concepts and a valence (either negative or positive) or two or more concepts, one of which has either a negative or a positive slant.

Especially when considering implicit biases such as racism, sexism or homophobia, the central idea seems to be that, despise sincerely and justifiably considering ourselves to be unprejudiced agents, consciously committed to egalitarianism in all its forms, we are often surprised to discover that we still harbour implicit attitudes that betray our unprejudiced, egalitarian explicit views. This kind of mismatch between our explicit and our implicit attitudes is often used to argue that implicit biases are unconscious. The assumption that implicit attitudes are unconscious mental states whose content can be diametrically different to the content of our explicit (self-reported) ones also makes it plausible to think that only indirect methods would give us information about them.

In order to highlight how accessibilism seems to deliver the wrong verdict when accounting for the justification of beliefs formed indirectly as a result of implicit biases, Katherine Puddifoot (2016) asks us to imagine two different scenarios. In the first one, Jones, a member of a jury in a rape case involving a black man, considers all available evidence provided by the prosecution and finds it convincing that the defendant is guilty. Jones thus has good reasons to believe that the defendant is guilty and believes that he is guilty for those reasons. In the second scenario, the evidence remains the same, but it seems convincing only because Jones holds an implicit bias against black men. In the second scenario, Jones associates black men with violence, so, were not for his implicit racist attitude, the available evidence would not seem compelling to him. In this second scenario, Puddifoot claims (2016, p. 422), Jones' belief that the defendant is guilty is not justified or, at a minimum, its justificatory status should strike us as much weaker. Puddifoot relies on this pre-theoretical intuition to argue against (both forms of) accessibilism. Here is her argument (Puddifoot, 2016, p. 422. ACCESS henceforth):

ACCESS:

1. According to accessibilism, only consciously accessible factors can be relevant to epistemic justification.
2. Implicit biases are consciously *inaccessible* factors.
3. Implicit biases are relevant to epistemic justification.

Therefore:

4. There are some consciously *inaccessible* factors that are relevant to epistemic justification.



5. Accessibilism is wrong.

This is how I will proceed from here. I show that recent research in social psychology warns us against the widespread conception of implicit biases as unconscious when thinking about their content. I rely on this research to reject premise 2 in ACCESS, but offer instead a refined version of the argument, ACCESS\_2 (Section 5). I then argue (Section 6) that the accessibilist can still meet the challenge posed by ACCESS\_2—or so it seems. Ultimately, I seek to debunk the intuition behind ACCESS\_2 (Section 7), but, hopefully, we would have learnt a lot about accessibilism, epistemic responsibility and implicit biases on our way to the final conclusion.

## References

- Chisholm, R. (1988). The indispensability of internal justification. *Synthese*, 74(3):285–296.
- Gawronski, B., W. Hofmann, & C. Wilbur (2006). Are “implicit” attitudes unconscious? *Consciousness and Cognition*, 15:485–499.
- Hall, D. L. & Payne, B. K. (2010). Unconscious influences of attitudes and challenges to self-control. In R. R. Hassin, K. N. Ochsner & Y. Trope (eds.) *Self Control in Society, Mind, and Brain* (pp. 221–242). Oxford: OUP.
- Puddifoot, K. (2016). Accessibilism and the challenge from implicit bias. *Pacific Philosophical Quarterly*, 97:421–434.
- Steup, M. (1999). A defense of internalism. In *The Theory of Knowledge: Classical and Contemporary Readings*, 2nd ed. Belmont, CA: Wadsworth, 373–84.

# Thinking and experiencing the embodied self

**Víctor M. Verdejo**

Logos Research Group/BIAP (Barcelona Institute for Analytic Philosophy)  
University of Barcelona

In a representational framework, there are two senses of embodiment: embodiment as body representation or representation carrying body information and embodiment as bodily awareness or awareness of our own body. One might concede that body representation is a form of self-representation and bodily awareness a form of self-awareness. However, are these reflexive cognitive capacities constitutive conditions on self-thought or first person thought? In this piece, I shall defend that whereas body representation is a plausible constitutive or necessary condition on self-thought, bodily awareness is not: one can have self-thought without bodily awareness but one cannot have self-thought without body representation. I therefore propose to detach the constitutive connection between body representation and self-thought from various forms of bodily awareness such as spatial awareness (Martin 1995), agentive awareness or the sense of being under direct control (Vignemont 2007, O'Brien 2007, Gallese and Sinigaglia 2010) or affective or emotional awareness (Damasio 1999).

In order to defend this standpoint, I will not merely appeal to extreme science fiction cases, such as deprivation tanks in which one's ordinary routes to knowledge of oneself are suspended (Anscombe 1975) or quasi-proprioception cases in which one is wired up to someone else's body (Campbell 1994, O'Brien 1995) but look into neuroscience data of patient studies. Bodily illusions, such as the Rubber Hand Illusion (Botvinick and Cohen 1998), better illustrate the sense in which bodily awareness may involve a deficient epistemic relation to one's body that is independent of self-thought and action capacities. In this experimental setting, one has the illusion that a rubber hand is one's own but one can still exhibit non-illusory self-thought and moving limb capacities (cf. Kammers et al 2009). Nonetheless, these independent action and thought capacities are plausibly anchored in unconscious body representation.

Deafferented subjects with no proprioception or touch in large parts of their bodies (Cole 1991, Cole and Paillard 1995) can also be used to illustrate the point: these subjects suffer severe alterations of their bodily awareness and hence their normal epistemic relations to their bodies are greatly impaired. However, while their capacity to feel one's body from the inside is severely affected, their self-thought capacities, such as the ones concerning thinking about their locations and actions, remain intact. In these cases, body representation is however present in a plastic and multimodal way (Wong 2015, 2017, forthcoming).

Similar analyses apply to other bodily illusions and disorders such as phantom limbs, autotopagnosia (inability to correctly identify body parts), out-of-body

experiences, or somatoparaphrenia (experience of one's body parts as someone else's), among others. However, even if we agree that body representation—as opposed to bodily awareness in its various manifestations—is a plausible constitutive condition on self-thought, there is a vast variety of body representations. It is clear that not all sorts of body representation will intervene in the constitution of self-thought since most of them will concern only parts or aspects of the body which may not involve a substantial form of self-representation. Body representations are present in a number of bodily receptors and channels of body information such as touch, vision, proprioception (which provides information about body position and movement), the vestibular system (responsible for balance), nociceptive system (which registers dangerous stimuli), and interoceptive system (which encodes optimal physiological conditions). Following the model of perception-action analysis in vision (Milner and Goodale 1995, 2008), body representation is typically divided into body representation for perception (the body image) and body representation for action (the body schema) (Paillard 1999, Gallagher 1995). I will take into account all these varieties of representation in order to suggest that body representation will constitutively intervene in self-thought only if it is at least potentially action-triggering, i.e. a type of self-representation that, however multimodal or unconscious, puts the organism in a position to take action on the basis of the target body information.

## Acknowledgments

This work has benefited from the support provided by the Secretary for Universities and Research of the Department of Economy and Knowledge (Government of Catalonia) and also by the Ministry of Economy and Competitiveness (Government of Spain) and the European Union through the research projects FFI2016-80588-R and FFI2015-63892-P (MINECO, AEI/FEDER, EU).

## References

- Anscombe, G.E.M. (1975). "The First-Person." In S. Guttenplan (ed.), *Mind and Language*. Oxford: Clarendon Press, pp. 45–65.
- Botvinick, M. and J. Cohen (1998). "Rubber Hands 'Feel' Touch that Eyes See", *Nature*, 391:756.
- Campbell, J. (1994). *Past, Space and Self*. Oxford: Clarendon Press.
- Cole, J. (1991). *Pride and the Daily Marathon*. London: Duckworth.
- Cole, J. and J. Paillard (1995). "Living without Touch and Peripheral Information about Body Position and Movement: Studies with Deafferented Subjects." In J.L. Bermudez, T. Marcel, N. Eilan (eds.), *The Body and the Self*. Cambridge, MA: MIT Press, pp. 245–266.
- Damasio, A. (1999). *The Feeling of What Happens*. London: William Heinemann.

- De Vignemont, F. (2007). "Habeas Corpus: the Sense of Ownership of One's own Body", *Mind and Language*, 22:447–449.
- Gallagher, S. (1995). "Body Schema and Intentionality." In J.L. Bermudez, T. Marcel, N. Eilan (eds.), *The Body and the Self*. Cambridge, MA: MIT Press, pp. 225–244.
- Gallese, V. and C. Sinigaglia (2010). "The Bodily Self as Power for Action", *Neuropsychologia*, 48:746–755.
- Kammers, M.P., F. de Vignemont, L. Verhagen and H.C. Dijkerman (2009). "The Rubber Hand Illusion in Action", *Neuropsychologia*, 47:204–211.
- Milner, D. and M.A. Goodale (1995). *The Visual Brain in Action*. New York: Oxford University Press.
- Milner, D. and M.A. Goodale (2008). "Two Visual Systems Reviewed", *Neuropsychologia*, 46:774–785.
- O'Brien, L. (1995). "Evans on Self-Identification", *Noûs*, 29:232–247.
- Paillard, J. (1999). "Body Schema and Body Image: A Double Dissociation in Deafferented Patients." In G.N. Gantchev, S. Mori and J. Massion (eds.), *Motor Control Today and Tomorrow*. Sofia, Bulgaria: Academic Publishing House, pp. 197–214.
- Wong, H.Y. (2015). "On the Significance of Bodily Awareness for Bodily Action", *The Philosophical Quarterly*, 65:790–812.
- Wong, H.Y. (2017). "On Proprioception in Action: Multimodality versus Deafferentation", *Mind and Language*, 32:259–282.
- Wong, H.Y. (forthcoming). *Embodied Agency*. Oxford University Press.

# FILOSOFÍA Y METODOLOGÍA DE LA CIENCIA

# What is really wrong with ontic structural realism? On the possibility of reading off ontology from current fundamental science

Haktan Akcin

I argue that the central conflict between epistemic (ESR) and ontic (OSR) versions of structural realism concerns whether it is possible to read off ontology from current fundamental science is.

Worrall provides a structuralist response to the tension between no miracle (NMA) and pessimistic induction (PIA) arguments. He agrees with anti-realists that introduction of “approximate truth” does not save realism from the destructive effects of ontological discontinuity implied in PIA. However, that does not mean that we should abandon realism altogether. If realist commits herself only to the mathematical or structural parts of theories, ontological discontinuity would not pose any threat to realism. Notice that ontological discontinuity is not denied here; rather, realism is reconstructed in a way that would not be vulnerable to the ontological discontinuity.

Reading off ontology from current fundamental science is not possible according to ESR. Although there is a structural continuity across radical theory changes, there is a discontinuity at the ontological level, and this discontinuity makes any reading off functionally otiose in time. Some structuralists, however, find such a restriction on reading off as problematic. Ladyman separates his ontic version of structural realism from Worrall’s epistemic version by arguing that metaphysics cannot be a restricted area for a realist: the correct form of structuralism should engage in metaphysical problems of modern science. The upshot of the argument is that we should replace our object-based ontology with a structuralist one. If objects are eliminated in favour of structure, then ontological discontinuity would disappear. For fundamental ontology consists of structure only, and so long as we can show continuity at the structural level, we can show the continuity of fundamental constituents in our ontology. Ontological discontinuity endorsed by Worrall is just a delusion since objects creating the discontinuity problem are eliminated in favour of structure. Therefore, according to OSR, reading off ontology from current fundamental science is possible, and the correct reading is that objects are dissolved into the structure at the fundamental ontological level.

After I set up the central conflict between ESR and OSR as the reading off ontology from current science, I look at two arguments in the philosophy of physics. The first argument is due to Ainsworth. He argues that there is no interpretation of modern physics that shows the ontological superiority of rela-

tions (or structures) over objects. I find Ainsworth's criticism quite convincing against both eliminativist and non-eliminativist versions of OSR. However, for the sake of argument, I assume that advocates of OSR could respond Ainsworth by insisting that relations are ontologically prior to relata according to modern physics. I introduce the second argument at this point to show that even if the ontological priority of structure over objects is established, that might not be sufficient for the claim that there is nothing except structure in the fundamental ontology. This second argument is related to a realist interpretation of entanglement in quantum information theory by Bub. The crucial claim is that entanglement should be taken as a new physical entity in the sense that it does not supervene on any other physical source like field or particle. I explain why I think the conclusion of Bub's argument is a challenge for OSR. However, again, for the sake of argument, I assume that proponents might find a way out and argue that Bub's argument does not damage OSR.

I then move on to what I find profoundly problematic about OSR. The locus in both Ainsworth's and Bub's challenges above is the philosophy of current fundamental physics. The discussion is about whether quantum particles are objects, or whether entanglement could be eliminated in favour of structure. Both opponents and proponents assume that it is possible to read off ontology from current fundamental physics. Although they disagree about the correct reading, in other words, they disagree about what would be the basic furniture of the world if modern science is correct, both sides assume that the reading off ontology is possible. Here, the discussion is *intra-theoretical*: within a theory, different parties discuss some metaphysical problems to find out the basic constituents of the fundamental ontology. However, ontological discontinuity argument that Worrall takes seriously is a problem concerning *inter-theoretical* relations: given the fact that these basic constituents of fundamental ontologies shift across theory changes, could the reading off ontology from current fundamental science be possible? After I write down the premises and conclusions of the arguments of both ESR and OSR, I argue that what is really wrong with OSR is that it includes the conclusion in one of the premises, hence begs the question. Since it begs the question, I conclude that the alleged superiority of OSR over ESR simply does not hold.

Despite fundamental problems I raise, I make it clear that OSR remains as a legitimate position in the philosophy of physics. Modern science is full of metaphysical problems, and discussions here might benefit a lot from OSR. The historical fact that OSR has been endorsed mainly by philosophers of physics shows that it touches upon some serious metaphysical problems of current fundamental science. However, in addition to the claim that OSR makes significant contributions to these metaphysical discussions, if the proponents also assert that their position provides a response to PIA, I am afraid that would not hold since OSR cannot solve the ontological discontinuity problem. Perhaps, instead of talking about ESR and OSR, it might be much more convenient to construct the argument as a discussion between Structuralism in Historical Perspective (SHP) and Structuralism in Modern Science (SMS). If the superiority claim of the latter is dropped, I believe they might coexist peacefully as two separate research programs. While SMS enriches metaphysical discussions in modern physics, SHP might provide a detailed analysis of shifting ontologies and structural continuities across theory changes; hence contributes to recent discussions

on global contexts and long-term developments in the history and philosophy of science.



# The explanatory power of topo-geometric diagrams in modern physics

Javier Anta  
LOGOS - UB

## Introduction

It comes as no surprise that almost all physics textbooks contain a vast manifold of images, ranging from photographs to sophisticated diagrams. Diagrams in physics are widely conceived to be epistemically insignificant representations, as mere visual support of symbolic or formulaic expressions of theoretical content. I will defend in this work that certain mathematically-structured diagrams, otherwise than mere unstructured or picture-like visual representation, might have a key explanatory (also predictive or understandability) import in certain modern physics contexts. The field of diagrammatic epistemology has been developed just in particular scientific areas, firstly within mathematics (Steiner; 1978) and just afterward in biology (Perini; 2005), but it still remains an underexplored topic in the domain of physics. This is precisely the purpose of this work. Section 2 will be devoted to evaluate, firstly, what constitute a diagrammatic inference of mathematical content, and secondly, how mathematical diagrams could be inferentially exploited to derive physical information. Afterward, I will assess two cases study of the epistemic role of diagrammatic representation in physics: Wheeler's geometric procedure in relativistic particle physics and phase portraits in chaotic systems theory. And finally, I will conclude defending in Section 3 the explanatory (or in general, the epistemic) import of topo-geometric diagrams within the physical realm.

## Inferential Character of Topo-Geometric Diagrams

While mathematical explanations of physical phenomena constitute now a hot topic, mathematical explanations of physical explananda properly based on diagrammatic representation has not been explored in the philosophical literature at all. It was precisely Mark Steiner (1978) who famously started the modern debate on mathematical explanations (of mathematical phenomena), the one pioneering the epistemic analysis of visual formal explanations of mathematical theorems, such as  $1 + 2 + \dots = n(n + 1)/2$  by means of a picture-proof. The widely accepted idea that “diagrams cannot be proofs” (Tennant; 1986) became

fiercely undermined during the 90s, mainly attacked by Nelsen's "Proofs without Words" (1993) and Brown's "Proofs and Pictures" (1998), wherein the last vindicated the high epistemic value and the formal validity of diagrammatical procedures as legitimate representational means for correctly testing mathematical hypotheses. Such a tendency will be extended into the empirical sciences and particularly in biology during the following decades, especially due to the work of Perini (2005). In this line, what epistemic role mathematical diagrams play in physics?

The epistemology of mathematical and non-mathematical diagrams become a largely fructuous area of research in the current century. Some authors were devoted to assess the justificatory role of diagrams in terms of "inferential support" (Norman; 2006), others, such as Molinini (2016), studied the formal correctness of the Euclidean diagrammatic-inference system. Here I will go further in exploring the idea that mathematical diagrams could be taken as representational vehicles physical information. Departing from Suarez's idea (2002) that representational mechanisms of theoretical content constraint the inferential power to derive theoretical information from that content, I will defend the thesis that one could correctly and efficiently explain a physical phenomenon by inferentially exploiting geometric and topological properties of physical diagrams.

Manders (2008) differentiated between mathematical-diagrammatical inferences grounding their epistemic power on co-exact (topological) resources and those relying on exact (metric) resources. In this line, Shimojima (1998), who pioneered experimental investigations on the inferential status of diagrammatic systems of representation), claimed that the inferential power of diagrammatic representations is determined on how efficiently are exploited graphical resources in order to encode relevant information. In his sense, a diagrammatic representation would be not just epistemically efficient, but moreover reliable if every inference carried on it is not deductively but wide-inferentially valid. Namely, that the conclusion diagrammatically derived depends, both syntactically and semantically, on the (topo-metrically codified, diagrammatically represented) information contained within the premises. This can be called "Criterion of Inferential Validity" (CIV). I would argue that not just mathematical theorems but also certain properly physical theoretical content is perfectly suited for being diagrammatically represented and graphically encoded<sup>1</sup> by using metric and topological structures. Let takes an illustrative example of CIV from modern physics: abducting that one positron moves on a space-time trajectory parallel to that of an electron by mean of a Feynman diagram (depicting, for instance, electron-positron annihilation) would constitute an invalid inference, since the linear path visually represented does not directly encodes information about this positron's trajectory but otherwise about the multivariate integral fixing its electrodynamical behavior. It does not mean that Feynman diagrams cannot epistemically reliable, but otherwise that their cognitive role is restricted to visually simplify (Kaiser; 2005) complex differential operations.

The fact that topo-geometric representations, or any other mathematical structure that exploit spatial resources, are well-suited to support valid inferences of physical phenomena is granted (thus satisfying CIV) by what Molinini (2016)

---

<sup>1</sup>Here, I differentiate between a metric or topological (formal) codification and diagrammatic (morphological) representation of physical information.

calls “inherent inferences”. This kind of inferential processes preserves physical meaning by means of correctly performed diagrammatic procedures following geometric rules, wherein the physical content of the conclusion semantic-syntactically depends on the physical content of the premises. For instance, diagrammatically abducting the relativistic phenomenon of “time dilation” in a high-acceleration context by means of geometrically manipulating the correspondent Minkowski diagram would constitute a valid inference due to the fact that every correct diagrammatic manipulation performed is based on a correct geometric rule and has a well-defined and metrically-encoded physical meaning.

## Case Study 1. Geometric Diagrams in Relativistic Particle Physics

If one would explore the vast production of diagrammatic representations in modern physics, one of the most philosophically interesting examples that it might be found in the particular procedure developed by John Wheeler in 1963 to relativistically calculate certain missing quantities within a nuclear reaction, such as mass of a tritium particle. His idiosyncratic method consisted on diagrammatically constructing an odd polygon, wherein each segment entails a three-dimensional projection on the dimensional page of a momentum-energy 4-vector embedded within three-dimensional energy-momentum coordinates. Each one of the three-dimensional component of the 4-vector encodes one relativistic quantity associated to a particle in the nuclear scattering; namely: rest mass (relativistic energy) on its y-projection as well as momentum<sub>x</sub> and momentum<sub>z</sub> on its x and z-projection respectively. On a higher level of description, vertexes represents relative energy-momentum (also velocity) between two particles as well as triangles represents the total relativistic energy-momentum information at a certain pre-collision or post-collision time. Beside such diagrammatic semantics, Wheeler’s procedure contained a set of syntactic rules that formally correspond to three-dimensional Euclidean geometry, like Pythagorean Theorem. This diagrammatic system allows deriving relativistic data, for instance, rest-mass of particle P, by manipulating the initial geometric elements just by means of correctly-performed Euclidean rules.

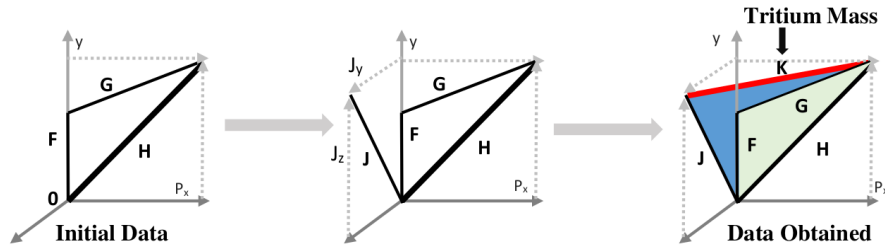


**Figure 1:** Diagrammatic inference of the mass of a Tritium particle via Wheeler’s procedure.

Let see a brief example of how Wheeler’s procedure works. The initial data of two incoming deuterons D and D’ and an outgoing proton p and a tritium T in a nuclear scattering is given on a triangle FGH plus projections J<sub>z</sub> and J<sub>y</sub>; wherein

F, G and J are momentum-energy 4- vector of D, D' and p respectively, and T is the total momentum-energy 4-vector. From projections  $J_x$  and  $J_y$  one could obtain segment J (energy-momentum of proton p) simply by applying the Pythagorean Theorem. And given J and FGH, we can derive K (energy-momentum of proton p) by tracing a segment between J and F. Thus, FGHJK contains all the energy-momentum data unfolded in this nuclear scattering. I would strongly argue that such derivation constitutes an inferentially valid operation since the physical meaning of geometric elements contained in the initial data will always be preserved under semantically-coherent and correct Euclidean transformation (thus, CIV is satisfied). Then, Wheeler literally demonstrated, in the strong sense of the word, that symbolic formula-based procedures were not the unique representational systems for inferring empirical data, explaining these data and also generating predictions; precisely because, as it should be noticed at his point, any valid geometric derivation of empirical data would constitute a testable prediction as well.

Interestingly, the epistemic power of Wheeler's diagrammatic system is not limited to data-derivation and prediction, but moreover it also has a significant explanatory import. For instance, we can geometric-diagrammatically explain the law of relativistic conservation of momentum- energy (preservation of the quantity described by stress-energy tensor  $T_{\mu}^{\nu}$ ) with a statement a polygon built of 4-vectors in space-time. The geometric fact that the segment H constitute an edge for both pre-scattering triangle FGH and post-scattering triangle HJK convey a geometric-diagrammatic explanandum for the amodal explananda (momentum-energy conservation law). Such geometric facts could be verbally formulated as "the quantity of momentum-energy between the two colliding deuteronx is equal to the quantity of momentum-energy between the two outgoing particles". This explanation of an empirical fact by means of a geometric diagram as explanandum do not just constitute a paradigmatic case of multi-modal explanation, but a type of explanation that is empirically superior (in terms of theoretical simplicity, understandability, predictability power, and so on) to any suitable verbal or formulaic alternative. Nevertheless, these epistemic virtues are domain-specific to cases wherein a small number of particles are involved. That is precisely the reason why Wheeler's procedure was not used anymore.

## Case Study 2. Topological Diagrams in Chaotic Dynamical Systems

In our second case study, we will focus on topological and diagrammatic explanations carried on phase portraits describing the dynamics of physical systems. A phase plane portrait is properly a bidimensional graphical representation of a plane section of phase space  $\Gamma$  configured by general coordinates  $Q$  and general momentum  $P$ . All possible initial conditions of a physical system is represented by a 2-form  $\omega$  in phase space  $\Gamma$  and every possible dynamical time-evolution is determined by curves on  $\omega$ . Valid and physically significant diagrammatic inferences on phase portraits could be easily carried given the appropriate conceptual tools. For instance, dynamic chaos is directly represented by graphical noise in phase portrait  $P$ , which exactly corresponds to the system having positi-

ve Kolmogorov-Sinai entropy (nearly-stochastic behavior) in that phase region. It would also be possible to locate “separatrix layers” on  $P$ , which are precisely those curves that separate different possible dynamical behavior of the same system; like the separatrix layer of a simple pendulum separating between its harmonic oscillation and its rotation.

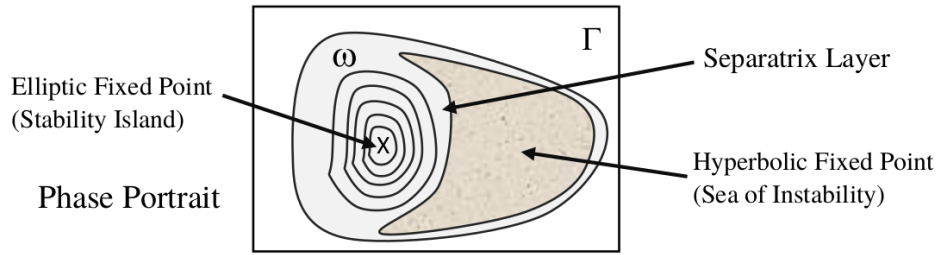


**Figure 2:** Diagrammatic inference of stable/chaotic dynamical behaviors of a system on its phase portrait.

From the Poincaré-Birkhof fixed-point theorem one could diagrammatically predict which phase regions will remain more stable (and which not) when the whole phase space becomes progressively chaotic, as well as it might be topologically explained why some phase regions become unstable and some remain stable when its energy increases. This theorem correlated elliptical curves with the so-called “Island of Stability” and hyperbolic curves with “Zones of Instabilities”; then, we might appeal to the disconnectedness of curves within pro-hyperbolic phase regions as the topological explananda that explain the exponential increase of positive Kolmogorov-Sinai entropy in such physical systems. Additionally, there exist possible diagrammetrical (using distances instead of topological tools) explanations of chaotic dynamics by relying on Brudno’s theorem and Poincaré maps as Lyon and Colyvan suggested (Lyon, A and Colyvan, M; 2014). If we visually track the phase trajectory of two nearby points and they exponentially diverge along time-evolution (diagram-metrical explanandum), then, the neighborhood of these two points will have positive Kolmogorov-Sinai entropy and therefore the system will behave chaotically at that point (diagram-metrical explanandum).

## Explanatory Power and other Epistemic Virtues of Topo-Metric Diagrams

I have tried to show with the above two cases is how it is possible, not merely to describe, but moreover to satisfactorily explain physical phenomena by relying on valid and correct diagrammatic inferences based on graphically encoded topo-metric properties. Their arguably epistemic superiority can be based on a two-fold enhanced indispensability argument (in the line of Saatsi; 2011), namely, that you cannot do epistemically better i.e. predicting, understanding, explaining, etc.; neither with a formula-based or a nominalist version. Concerning the enhanced indispensability of phase portraits as topological diagrams, Lyon and

Colyvan (2014) argued that physical theories based on phase spaces cannot have a nominalist reformulation precisely because its explanatory power depend on phase-space topological properties. Here, I will go beyond Lyon-Colyvan thesis and defend that a graphical codification of phase spaces as phase portraits would reinforce certain epistemic virtues, such understandability (cognitive accessibility) and computational-analytical tractability, that symbolic or formula-based versions would not possess. Of course, theoretical content diagrammatically represented on phase space will be much more understandable than its sentential reformulation, due to the nearly immediate visual accessibility of its theoretical content. The main argument is that computer-driven and sophisticated graphical methods used in chaotic systems theory would constitute a necessary tool for analyzing visually-distinguishable changes in the dynamical stability of physical systems, given numerical solutions. Thus, as long as our second case study is concerned, topology-based diagrammatic explanations (a.k.a. visual abductions exploiting topological resources) of why physical systems become unstable are not just possible, but necessary or enhanced-indispensable.

Regarding our first case, geometric explanations of relativistic energy-mass conservation constitute an elegant alternative, in a restricted sense epistemically superior, to the traditional symbolic account. It cannot be argued that the (domain-specific) Wheeler's procedure constitutes an epistemically indispensable method to derive relativistic quantities; otherwise, its epistemic merit rely on both (i) undermining the widely-assumed enhanced indispensability of formulaic procedures, and (ii) largely increasing the cognitive accessibility (Molinini; 2016) of such physical phenomena. What Molinini calls "inherited inferences" become instantiated in Wheeler's derivation of the relativistic mass of tritium, since certain invariance of geometric properties (e.g. length) in Euclidean operation "inherit" the physical meaning of such properties (e.g. energy-mass of a particle) in terms of conservation during physical processes. As it might be noticed, Molini's inferential inheritance follows directly by simultaneously assuming mathematical representations of physical phenomena and our CIV.

Up to this point, it can be sketched an inference-based account on why diagrammatic representations exploiting topo-metric resources are explanatory effective concerning these physical phenomena. The core idea is that diagrammatic representation as inferential devices (Suarez; 2002) trigger a belief-forming abductive (explanation-based) process, wherein its reliability or inferential validity is based on the reliable conformation between the perceptual content of such correct diagrams and their topo-geometric inferential rules. Or in other words, diagrammatic representations can be epistemologically well-suited in certain restricted contexts for inferentially exploiting topological and geometric structures in order to carry description, predictions and, moreover, explanations of physical phenomena.

## Conclusion

In this work, I have argued that diagrammatic explanations of physical phenomena are not just possible, but in some cases they are epistemically superior that their formula-based versions. Furthermore, I have defended that diagrammatic

inferences underlying these visual explanations depend essentially on exploiting mathematical structures, especially metrical and topological ones, to convey physical information. Then, one could talk about properly topo-geometric diagrammatic explanations of physical phenomena. As concrete examples of properly metrical and topological diagrammatical explanations in physics, I have rely on Wheeler's geometric procedure and phase portraits, respectively. In this sense, the inferential validity and efficacy of topo-geometric diagrammatic representations in modern physics would entail a highly explanatory import within the domain of modern physics.

## References

- Brown, J. (1999). *Philosophy of Mathematics: an introduction to the world of proofs and pictures*. London: Routledge.
- Barwise, J. & Etchemendy, J. (1996). "Visual information and valid reasoning", in *Logical Reasoning with Diagrams*, G. Allwein and J. Barwise (eds). Oxford: Oxford University Press.
- Bueno, O. & Colyvan, M. (2011). "An Inferential Conception of the Application of Mathematics", *Noûs*, 45:345–374.
- Bueno, O. and French, S. (2012). "Can Mathematics Explain Physical Phenomena?", *British Journal for the Philosophy of Science*, 63(1):85–113.
- Giaquinto, M. (2007). *Visual Thinking in Mathematics*. Oxford: Oxford University Press.
- Lyon, A. (2012). "Mathematical Explanations of Empirical Facts, and Mathematical Realism", *Australasian Journal of Philosophy*, 90(3):559–578.
- Lyon, A. & Colyvan, M. (2008). "The explanatory power of phase spaces", *Philosophia Mathematica*, 16(2):227–243.
- Mancosu, P. (2005). "Visualization in Logic and Mathematics", in P. Mancosu, K. Jørgensen and S. Pedersen (eds), *Visualization, Explanation and Reasoning Styles in Mathematics*. Dordrecht: Springer.
- Manders, K. (2008). "The Euclidean Diagram", in Paolo Mancosu (ed.), *The Philosophy of Mathematical Practice*. Oxford University Press. pp. 80–133.
- Molinini, D. (2016). "The Epistemological Import of Euclidean Diagrams", *Kairos. Journal of Philosophy and Science*, 16(1):124–141.
- Norman, J. (2004). "Can diagrams have epistemic value? The case of Euclid", in A. Blackwell, K. Marriott & A. Shimojima (eds.), *Diagrammatic Representation and Inference*. Springer. pp. 14–17.
- Perini, L. (2005). "Explanation in two dimensions: Diagrams and biological explanation", *Biology and Philosophy*, 20(2-3):257–269.

- Saatsi, J. (2011). "The Enhanced Indispensability Argument: Representational versus Explanatory Role of Mathematics in Science", *British Journal for the Philosophy of Science*, 62(1):143–154.
- Shimojima, A. (1996). "Operational constraints in diagrammatic reasoning", in Gerard Allwein & Jon Barwise (eds.), *Logical Reasoning with Diagrams*. Oxford University Press.
- Steiner, M. (1978). "Mathematical Explanation", *Philosophical Studies*, 34:135–151.
- Suárez, M. (2002). "An inferential conception of scientific representation", *Philosophy of Science*, 71(5):767–779.
- Tennant, N. (1986). "The Withering Away of Formal Semantics?", *Mind and Language*, 1(4):302–318.
- Wheeler, J. A., & Taylor, E. F. (1963). *Spacetime Physics*. A Series of Books in Physics. W. H. Freeman and Company.



# Scientific pluralism and scientific realism: Examining problems and reducing gaps

Paula Atencia  
UCM

The need to acknowledge the heterogeneous and pluralistic nature of science has been increasingly recognized in philosophy of science in recent decades and several philosophers have focused precisely on explaining such plurality, which is manifested in the methodological, theoretical and even ontological diversity shown by science. Those efforts have given rise to a philosophical stance called “scientific pluralism”. This label comprises different theses and approaches, but advocates of scientific pluralism usually share the firm belief that their stance is compatible with scientific realism. This view, however, is not unanimously accepted in philosophy of science, and some philosophers find troubles reconciling certain pluralist theses and scientific realism. Discussions regarding this particular issue take different forms. For instance, it is not uncommon that philosophers address the problem of the reconciliation of scientific pluralism and realism from the point of view of the debate between pluralism and monism. Monism is regarded as a philosophical view whose main core is the thesis that the aim of science is to provide a complete, comprehensive and all-encompassing theory of the natural world based on a set of fundamental principles. Pluralists point out that it might not be possible to develop such a theory, questioning the importance of committing to this particular view. It can be claimed, then, that the rejection of monism can ease the tensions between scientific pluralism and realism, since it allows us to recognize that the coexistence of multiple theories in one and the same context of research—or, in other words, the (theoretical) plurality of science—does not necessarily represent a temporary state of affairs. In connection to this, discussions regarding the compatibility between scientific pluralism and realism also fall under the framework of “representational pluralism”, an approach that revolves around those empirical cases where multiple theories or models of a single entity or phenomenon coexist. Some of these situations are not problematic, for such theories or models are compatible with each other; however, others do not. Those empirical cases where incompatible theories about the same entities coexist are widely regarded as the real challenge of pluralism to scientific realism.

This presentation addresses both issues, namely the debate between pluralism and monism and the incompatibility of theories. Within the context of the latter, two interrelated problems will be examined, ontological pluralism on the one hand and the unification of theories on the other. Additionally, some strategies will be suggested to reconcile scientific pluralism—or more particularly, ontological pluralism—and scientific realism.

# Kinds, joints and trees: The non-hierarchical structure of natural kinds

Javier Belastegui

A crucial feature of natural kinds is that they ground our classifications of the objects that can be found in the world. If theoretical classifications are to be objective, they must preserve these natural distinctions. So, what features natural kinds (necessarily) have determine what conditions objective classifications must satisfy. Now, there is an old thesis that goes back to Porphyry according to which natural kinds must be arranged as a tree. Two natural kinds cannot overlap unless one of them is a species of the other. This is the “Hierarchy Thesis” (H). According to another commonly held view, natural kinds have sharp boundaries. It is always determined whether a particular object belongs or does not belong to the kind. This is the “Categoricity Thesis” (C). Some authors have defended that (C) is a metaphysically necessary condition for something to be a natural kind (Ellis, 2001), and others that (H) is a logically necessary condition to be satisfied by a natural classification (Thomason, 1969). The aim of this talk is to argue against these two theses about the nature of natural kinds and classifications.

Before introducing the debate, it is convenient to explain what (H) amounts to and to distinguish two readings of (C) from (Ellis, 2001) and (Ellis, 2002):

**Hierarchy (H):** For every pair of different natural kinds  $K, K'$ , either  $K$  is a species of  $K'$  or  $K'$  is a species of  $K$ .

**Categoricity I (CI):** For every individual object  $x$ , for every natural kind  $K$ , it is objectively determined whether  $x$  is a  $K$  or whether  $x$  is not a  $K$ .

**Categoricity II (CII):** For any pair of individual objects  $x, y$ , which are members of natural kinds  $K, K'$ , respectively, there is no “continuous transition” from  $x$  to  $y$ .

Recently, several authors (Hendry, 2006), (Tobin, 2010), (Khalidi, 2013), (Hendry, 2015) have challenged these two assumptions with empirical counterexamples. According to them, (H) and (C) are contingently false, since they fail in different domains of science. Whereas the violations of (H) arise from intrataxonomic and intertaxonomic crossings (Tobin, 2010), the violations of (C) seem to arise from the fact that some classifications are based on continuous magnitudes (e.g. temperature) of the objects (Hendry, 2006).

The structure of the talk is as follows. First, after making some basic distinctions, I will argue that the aforementioned empirical counterexamples pose a real challenge to (H) and (C) and cannot be easily dismissed. Then, I will show that the conditions cannot be considered as logically necessary either, for there

are alternative models of classification that do not satisfy them. Against (H), I will argue that other theoretical virtues, such as simplicity or explanatory power, do not force us to choose between a classification that satisfies (H) and one that does not. Concerning (C), (Ellis, 2001) gives an argument to conclude that it is a metaphysically necessary condition. This argument requires that (CI) implies (CII). (Hendry, 2006) disagrees. I will consider a possible reply by Ellis, and I will argue against it that the violation of (CII) can be considered simply as a violation of (H). Finally, I will hint at some possible consequences of rejecting (H) and (C). We have seen that these conditions only hold contingently in some specific domains of science and that they are not particularly supported by theoretical virtues such as simplicity or explanatory power. As a consequence, there is no big difference from a methodological point of view between those classifications that satisfy them and those that do not. Furthermore, for those domains in which they do not hold, to impose them on the corresponding scientific classifications is tantamount to introducing conventional elements. Therefore, the absence of crossings or continuous transitions from one class to another is more likely to be an *artifact* of our representation (in the sense of (Swoyer, 1991)) than a real or objective feature of the world.

## References

- Ellis, B. (2001). *Scientific Essentialism*. Cambridge: Cambridge University Press.
- Ellis, B. (2002). *The Philosophy of Nature: A Guide to the New Essentialism*. Chesham: Acumen.
- Hendry, R. F. (2006). Elements, Compounds and Other Chemical Kinds. *Philosophy of Science*, 73:864–875.
- Hendry, R. F. (2015). Are Chemical Kinds Natural Kinds? En U. Mäki, I. Votsis, S. Ruphy, G. Schurz (eds.), *Recent Developments in the Philosophy of Science: EPSA13 Helsinki* (pp. 251–261). Switzerland: G. Springer International Publishing 1.
- Khalidi, M. A. (2013). *Natural Categories and Human Kinds. Classification in the Natural and Social Sciences*. Cambridge: Cambridge University Press.
- Ruphy, S. (2010). Are Stellar Kinds Natural Kinds? A Challenging Newcomer in the Monism/Pluralism and Realism/Antirealism Debates. *Philosophy of Science*, 77(5):1109–1120.
- Swoyer, C. (1991). Structural representations and surrogate reasoning. *Synthese*, 87:449–508.
- Thomason, R. (1969). Species, determinates and natural kinds. *Noûs*, 3(1):95–101.
- Tobin, E. (2010). Crosscutting Natural Kinds and the Hierarchy Thesis. En Helen Beebe & Nigel Sabbarton-Leary (eds.) *The Semantics and Metaphysics of Natural Kinds*. London y New York: Routledge Studies in Metaphysics.

# Operational definitions of life

**Leonardo Bich**

IAS-Research Center for Life, Mind, and Society. Department of Logic and  
Philosophy of Science

University of the Basque Country (UPV/EHU)  
Avenida de Tolosa 70, 20018, Donostia-San Sebastian, Spain

Despite numerous and increasing attempts to define what life is, there is no consensus on necessary and sufficient conditions for life. Accordingly, some scholars have questioned the value of definitions of life and encouraged scientists and philosophers alike to discard the project. As an alternative to this pessimistic conclusion, it has been argued that critically rethinking the nature and uses of definitions can provide new insights into the epistemic roles of definitions of life for different research practices. This talk discusses the possible contributions of definitions of life in scientific domains where such definitions are used most (e.g., Synthetic Biology, Origins of Life, Alife, and Astrobiology). Rather than as classificatory tools for demarcation of natural kinds, it highlights the pragmatic utility of what has been called ‘operational definitions’ that serve as both theoretical and epistemic tools in scientific practice. In particular, it examines contexts where definitions integrate criteria for life into theoretical models that involve or enable observable operations. It shows how these definitions of life play important roles in influencing research agendas and evaluating results, and it argues that to discard the project of defining life is neither sufficiently motivated, nor possible without dismissing important theoretical and practical research. Moreover, it shows how definitions built on the organisational perspective—such as those based on Ganti’s Chemoton (Ganti, 2003) and on autonomy and open-ended evolution (Ruiz-Mirazo et al. 2004; 2010)—perfectly fit the theoretical and epistemological requirements for operational definitions, and it proposes a possible expansion of the definitions of life in terms of autonomy in the domain of origins of life (Ruiz-Mirazo et al., 2004; 2010; Bich & Damiano, 2012), by explicitly including adaptive regulation in the set of mutually dependent (integrated) necessary conditions for life that characterise operational definitions.

## References

- Bich, L. & Damiano, L. (2012). Life, Autonomy and Cognition: An Organizational Approach to the Definition of the Universal Properties of Life. *Origins of Life and Evolution of Biospheres*, 42(5):389–397.
- Bich, L. & Green, S. (2017). Is defining life pointless? Operational definitions at the frontiers of biology. *Synthese*, doi: 10.1007/s11229-017-1397-9.

- Cleland, C. (2012). Life without definitions. *Synthese*, 185, 125–144.
- Cleland, C. E., & Chyba, C. F. (2002). Defining ‘life’. *Origins of Life and Evolution of the Biosphere*, 32, 387–393.
- Machery, E. (2012). Why I stopped worrying about the definition of life... and why you should as well. *Synthese*, 185(1), 145–164.
- Ruiz-Mirazo, K. Peretó, J., & Moreno, A. (2004). A universal definition of life: Autonomy and open-ended evolution. *Origins of Life and Evolution of the Biosphere*, 34(3), 323–346.
- Ruiz-Mirazo, K. Peretó, J., & Moreno, A. (2010). Defining life or bringing biology to life. *Origins of Life and Evolution of the Biosphere*, 40(2), 203–213.

# *Contra* static dispositions

Andrei A. Buckareff

Marc Andrews

Shane Brennan

Marist College

Poughkeepsie, NY (USA)

andrei.buckareff@marist.edu

**Short abstract:** Work on dispositions focuses chiefly on dispositions that are manifested in dynamic causal processes. Neil Williams (2005) has argued that the focus on dynamic dispositions has been at the expense of a richer ontology of dispositions. He contends that we ought to distinguish between dynamic and static dispositions. The manifestation of a dynamic disposition involves some change in the world. The manifestation of a static disposition does not involve any change in the world. In this paper, we concede that making a *conceptual* distinction between dynamic and static dispositions is useful and we allow that we can truthfully represent objects as manifesting static dispositions. However, we argue that the distinction is not *ontologically* deep. Rather, the truthmakers for our representations of static dispositions are actually dynamic dispositions to whose manifestations we may fail to be sensitive.

In the literature on dispositions, the focus has been on dispositions being manifested in *dynamic* causal processes. Neil Williams (2005) has argued that this has been at the expense of a richer ontology of dispositions that distinguishes between dynamic and static dispositions. We argue that any distinction between static and dynamic is not ontologically deep. The activity of all dispositions involves some change. Hence, any representations of static dispositions are made true by the work of dynamic dispositions.

We proceed in our paper as follows. First, we sketch Williams' case for making the distinction between static and dynamic dispositions. We then offer an argument against the existence of genuinely static dispositions. Next, we consider three cases of static dispositions Williams offers and argue that these are best understood as involving dynamic dispositions.

## Dispositions: dynamic and static

We agree with Williams' realist assumptions about the nature of dispositions. Dispositions are not to be confused with their manifestations. So if  $x$  is a disposition for a particular manifestation  $y$ , then  $x$  is not identical with its manifestation  $y$ .

An object has a particular disposition for a specific manifestation regardless of whether the disposition is ever manifested. The dispositions of objects are possessed in virtue of the properties an object possesses. They are manifested when they interact with other dispositions that serve as mutual manifestation triggers.

When we speak of dispositions and their subsequent manifestations, a focus is typically placed upon those dispositions that result in a change. Due to this conceptualization of dispositionality, as well as the linguistic connotations of the term ‘disposition’, it is presumed that all dispositions are dynamic in some way. Williams thinks the assumption that all dispositions are dynamic is a mistake. He urges acceptance of a distinction between *dynamic dispositions* and *static dispositions*.

Dynamic dispositions are dispositions whose manifestation involves some change in the world (Williams 2005, 303–304). In contrast, the manifestations of static dispositions do not involve any change (Williams 2005, 304, 309–312). These include dispositions for internal stability that are involved in the persistence and maintenance of an object’s intrinsic properties (Williams 2005, 304, 312–315). There are also dispositions for external stability that are “involved in maintaining an object’s extrinsic properties” (Williams 2005, 304). Finally, there are static dispositions involved in threshold conditions that serve as “some point or magnitude separating conditions that are adequate for bringing about some effect, from those that are not” (Williams 2005, 315).

## Our central argument

We agree with Williams that he provides a useful distinction between static and dynamic dispositions that accurately tracks how we think and talk about the world and the behavior of its constituent parts. So there is a real explanatory distinction that exists between static dispositions and dynamic dispositions. But this distinction is *merely* explanatory and conceptual.

Consider the following argument:

- (1) Every manifestation of a disposition involves some change at some point.
- (2) If (1), then there are no dispositions whose manifestation does not involve any change.
- (3) If there are no dispositions whose manifestation does not involve any change, there are no genuine static dispositions.
- (4) If there are no static dispositions, then the distinction between dynamic and static dispositions is not ontologically deep.
- (5) So the distinction between dynamic and static dispositions is not ontologically deep.

We present a case for premises (1) and (2). A synopsis of our cases for each premise follows.

Regarding premise (1). Dispositions do not come into the world already manifested. Suppose that some quantity of fundamental physical stuff has a disposition that is manifested and has been manifested continuously since the beginning of the universe. That disposition did not come into being without some initial change from being a mere potentiality to being fully manifested. If this is correct, then (1) is true.

In defending premise (2) we argue that if we assume that dynamic manifestations of dispositions involve change and that static manifestations do not involve change, then it seems that there are no static manifestations. Any possible lack of change is only at moments later than the initial onset of the manifestation. The initial onset of the manifestation is dynamic. Moreover, it is reasonable to expect subsequent changes to take place that may not appear substantial if we rely solely on our limited perceptual abilities and ordinary instruments to detect changes. So the total occurrence of an alleged static disposition's manifestation is not obviously static. Therefore, the total occurrence of a manifestation of a given disposition is not static. If this is the case, then there are no static dispositions. Importantly, the foregoing would suggest that (2) is also true.

Even if the foregoing suffices to show that any putative static dispositions are dynamic dispositions, it is instructive to consider the cases Williams discusses. He presents three possible scenarios in which static dispositions would be at play. In our paper we examine cases Williams offers involving dispositions for internal stability, dispositions for external stability, and threshold conditions for dispositions. We argue that in each of these cases dynamic dispositions are being manifested that provide us with the truthmakers for representing objects as possessing static dispositions. In the concluding section we consider two possible objections the defender of static dispositions may offer in defense of there being an ontologically deep distinction between static and dynamic dispositions.

## Works cited

Williams, N. (2005). Static and dynamic dispositions. *Synthese*, 146:303–324.



# La autoorganización como propiedad de nivel de sus entidades componentes<sup>1</sup>

Emilio Cáceres Vázquez

Departamento de Lógica, Historia y Filosofía de la Ciencia  
UNED, Spain.

## Introducción

Una de los conceptos clave para describir el comportamiento de los sistemas complejos es el de *auto-organización*, entendida como un proceso en el cual las interacciones locales entre los elementos de un sistema producen patrones emergentes de comportamiento sin que para ello sea necesario ningún tipo de coerción o control externo (Anderson 2002, 248–49). Esta idea de auto-organización se entiende como un tipo de “cierre organizacional” que apela a una emergencia en términos de relación entre los componentes de un sistema y sus capacidades o propiedades globales, lo que conduce por tanto a la postulación de propiedades ontológicamente irreducibles con poderes causales distintivos. La constricción ejercida por esta emergencia da lugar a un tipo de auto-organización y automantenimiento que diferencia los seres vivos del resto de sistemas (Mossio, Bich, y Moreno 2013, 153).

La intención de esta ponencia es tratar de explicar la idea de auto-organización, entendida como cierre organizacional, como una *propiedad de nivel* desde la perspectiva de los intervalos de cuasi-descomponibilidad (ICD) (Cáceres y Saborido 2017, 2018). El enfoque organizacional, que se fundamenta en el reconocimiento y análisis de esta autoorganización distintiva, puede considerarse como una forma explicativa de los sistemas vivos, y de otros sistemas complejos, pero desde un enfoque meramente heurístico sin la carga metafísica que muchas veces deja entrever.

En primer lugar, haré un breve esbozo de las ideas centrales del cierre organizacional basado en las restricciones intentando extraer las bases en las que se sustenta. Continuaré la idea de nivel como intervalo de cuasi-descomponibilidad y como esto nos permite determinar entidades sistema de forma arbitraria pero no caprichosa, describiendo la idea de propiedad de nivel como una propiedad modelizada de las entidades-sistema de un ICD, deducible de las propiedades de los elementos de un todo. De esta forma, aclararé como las propiedades

---

<sup>1</sup>Esta ponencia acerca de la autoorganización, forma parte de mi tesis doctoral acerca de las propiedades emergentes. A pesar de estar previsto tratar este tema desde mi TFM, debo agradecer el reconocer su pertinencia a la observación de un revisor anónimo de mi primera publicación en *Theoria*. En la nota 1 de dicho artículo hice referencia a un posterior desarrollo de la autoorganización que pretendo que se convierta en un nuevo artículo en breve. Este trabajo es un primer acercamiento al tema y así debe entenderse.

emergentes pueden tratarse como propiedades de nivel, así como la causación internivel también puede explicarse desde la perspectiva cuasi-descomponible y como afecta esto al enfoque organizacional.

## La perspectiva organizacional y el cierre de constricciones

Resulta evidente (creo) que en el intento de conocer la vida, tanto las visiones más crudamente reduccionistas como las visiones más ampliamente holistas proceden de puntos de vista previos a su argumentación. Cada uno de ellos busca su antecedente filosófico indagando en los análisis del mecanicismo cartesiano o del organicismo kantiano, fundamentando la cuestión en una tradición pasada. Sin embargo, al margen de enfoques no materialistas, creo que la diferencia entre todas las propuestas es principalmente de nivel de observación. Hasta la visualización de primeras células vivas por Leuwenhoek y sobre todo de la teoría celular de Schleiden y Schwann, la vida solo se percibía desde un nivel superior, pero a partir de este momento, la observación de lo vivo desciende un escalón para situar su mínimo en la célula. Sin embargo, independientemente de lo anterior, en ambos casos se observa un todo formado por partes que se integran dando lugar a un comportamiento conjunto. Es esta organización enfocada a un fin descrita ya por Kant la que impulsa la visión más holista de la vida, una vez finalizado el vitalismo con la declaración de Haldane (Mayr 2005, 30). La salida de la biología del *elán vital* no supuso el abrazo del reduccionismo, sino una reorientación hacia un nuevo concepto, el organicismo, que aceptaba la idea de que los componentes materiales y las reacciones fisicoquímicas eran en última instancia las responsables de lo vivo, pero que esta influencia disminuía con la complejidad de forma que *“los todos están tan relacionados con sus partes que no solo la existencia del todo depende de la cooperación ordenada y la independencia de sus partes, sino que el todo ejerce además un cierto grado de control determinista de sus partes”* (Mayr 2005, 31). Esta idea se basaba en la emergencia, término acuñado por Lewes (Lewes 1877) para referirse a las leyes heteropáticas descritas por Stuart Mill unos años antes (Mill 1843). El desarrollo de la emergencia basada en la impredecibilidad, la novedad cualitativa, la realizabilidad múltiple y la causación descendente, alcanzó su máximo en los años 20 del siglo pasado de la mano los denominados emergentistas británicos<sup>2</sup>. La organización y el organismo fue el núcleo del denominado *Club de Biología Teórica* de Cambridge, formado por Woodger, Needham, Waddington, Wrinch y Bernal, como nombres más destacables (Etxeberría y Umerez 2006, 7). Influidos por Bertalanffy, propusieron al organismo y a la organización como claves para entender la vida, desarrollando una teoría de jerarquías y de relación entre el todo y las partes. Esta biología basada en el organismo/organización decayó paralelamente al surgimiento de la biología molecular derivada de los trabajos de (Griffith 1928), (Avery, MacLeod, y McCarty 1944), (Hershey y Chase 1952), etc, que culminaron con el descubrimiento de la doble hélice en 1953 (Watson y Crick 1953). Sin embargo, el fallo del reduccionismo (Bunge 2004, 191) permitió un florecimiento de las ideas organicistas (Moreno y Mossio 2015, xxv) basado

---

<sup>2</sup>Principalmente (Alexander 1920), (Morgan 1923) y (Broad 1925).

en el surgimiento de las ideas de homeostasis como resistencia a la perturbación externa y la auto-organización en condiciones alejadas del equilibrio que culmina en la idea de autonomía biológica (Moreno y Mossio 2015, xxvi). Sobre la base de la vida como autonomía, se ha desarrollado el concepto de cierre, en el sentido de autocontención.

Desde que Varela habla de sistemas operacionalmente cerrados (Varela 1979, 58), la idea de cierre ligada a la autonomía ha trazado una línea de pensamiento en la filosofía de la biología que ha sido usado por “Howard Pattee, Robert Rosen y Stuart Kauffman, en un sentido similar o complementario” (Moreno y Mossio 2015, 1). Pero como bien afirman Moreno y Mossio en su “*Biological autonomy. A Philosophical and Theoretical Enquiry*”, el concepto de cierre no está convenientemente definido y, continúan, no se ha establecido ninguna relación entre el cierre como régimen causal en biología y otros regímenes causales en física o química, que en caso de existir sin perder información relevante, relegarían al cierre a una posición heurística sin potencia explicativa (2015, 2). Sin embargo, si como afirman los autores el cierre fundamenta muchas propiedades propias de la vida, como la individuación, normatividad y funcionalidad (2015, 4) incluso aunque solo fuera un heurístico, sería una buena aproximación a la biología.

De los distintos tipos de cierre, una de las variantes más desarrolladas en los últimos años, sobre todo por parte de los investigadores del IAS<sup>3</sup> es el denominado cierre de constricciones. Pattee define constricción como “*causas locales y contingentes, ejercidas por estructuras o procesos específicos, que reducen los grados de libertad del sistema sobre el que actúan*” (Pattee 1972). En este tipo de cierre, las constricciones son entidades que actúan sobre los procesos de un flujo termodinámico alejado del equilibrio de forma que, en su conjunto, se auto-mantienen (Moreno y Mossio 2015, 11). Estructuras como los sistemas disipativos de Prigogine (Prigogine 1978) son auto-organizativas y contribuyen a su automantenimiento, pero este aspecto es insuficiente para la biología, que requiere un sistema más complejo de constricciones (en la misma o distinta escala espacial y temporal) que dependen entre sí de forma mutua y que se cierra sobre sí mismo (Moreno y Mossio 2015, 15–20) que además, requiere apelar al *cierre organizacional* como un régimen emergente de causalidad (Mossio, Bich, y Moreno 2013, 155).

Otro aspecto importante de las constricciones es desde que nivel y sobre que nivel se ejercen. En la propuesta analizada, se afirma que no está claro si existe una relación de causalidad internivel, aunque a priori parece evidente que lo inferior determina lo superior y viceversa. Pero también dicen que la causación internivel puede interpretarse de dos formas: por una parte, el caso en el que una entidad de nivel superior condiciona *a otra* de nivel inferior; por otra, la situación en el que un todo condiciona *a sus propios* componentes. La primera la consideran ubicua y no problemática, como desde otra perspectiva hacen (Craver y Bechtel 2006). La segunda, que parte del enfoque de *parte-todo* de (Spaulding 1912, 158), (1918, 450–51) y que desarrollaron posteriormente (Campbell 1974) o (Sperry 1969) con el clásico ejemplo de la rueda. Sobre este tema, los defensores del cierre organizacional defienden que es suficiente el primero tipo de causación descendente para justificar las constricciones, mientras que la reflexiva es innecesaria (Mossio, Bich, y Moreno 2013, 169).

---

<sup>3</sup>Centro de Investigación del IAS para la Vida, la Mente y la Sociedad UPV/EHU.

Por tanto, la clave para que se pueda considerar el cierre de constricciones como algo más que una herramienta heurística queda sujeto a la existencia de las propiedades emergentes y de la causalidad descendente no reflexiva.

## Nivel como intervalo de cuasi-descomponibilidad

Este concepto de nivel parte de la idea de *nearly-decomposable systems* introducida por Herbert Simon (Simon 1962) para referirse a cierta forma de descomponer un sistema basada en un valor arbitrario  $\varepsilon$  rindiendo subsistemas que se influyen solo de forma agregativa. Pero a diferencia de Simon, mi propuesta considera la inexistencia de sistemas descomponibles, pues por pequeña que sea la influencia entre dos particiones de un sistema, nunca será nula y por tanto no cumple su premisa principal. Así, es posible cuasi-descomponer el continuo material a partir de varios criterios de cuasi-descomponibilidad  $\varepsilon_n$  de manera que un sistema quede dividido en varios subsistemas cuasi-descomponibles anidados definiendo un intervalo de cuasi-descomponibilidad.

Según nuestra elección de los diferentes  $\varepsilon_n$  habrá un *nivel fundamental aporoblemático* o microscópico  $F^4$ , un nivel macroscópico  $M$  y múltiples niveles mesoscópicos anidados  $m_i$ . Comenzando en los átomos<sup>5</sup>, podemos esgrimir razones energéticas para establecer valores de  $\varepsilon_n$  que discriminen moléculas, células, organismos, poblaciones, etc. De esta partición longitudinal de la materia se sigue que la relación entre los subsistemas superiores y los inferiores es simplemente composicional por lo que la relación entre dos niveles diferentes es de *identidad*.

Una vez definidos los ICD, una ciencia concreta se encargará de su estudio y lo describirá estructural y funcionalmente según su metodología, lo que incluye un nivel de observación. En función de sus fines, tomará solo aquellos *inputs* y *outputs* que considere necesarios estableciendo un sistema modelo que supondrá una simplificación heurística de la realidad. Esta pixelación se amplifica conforme ascendemos de nivel de manera que las generalizaciones observadas serán igualmente inexactas al margen de resultar explicativas y predictivas.

Desde este enfoque, solo las entidades del nivel  $F$  no están formadas por partes y tienen propiedades no cuasi-descomponibles debidas a su propia naturaleza, mientras que las entidades resultantes de la elección de un  $\varepsilon_n$ , son entidades-sistema (Cáceres y Saborido 2018, 139) y tendrán un conjunto de propiedades obtenidas de la simplificación realizada. Estas propiedades las denomino *propiedades de nivel*, pues se manifiestan solo a el nivel de observación debido a la metodología de la ciencia encargada del estudio de dicho nivel. Se trata de las propiedades tradicionalmente consideradas *propiedades emergentes*, pues se manifiestan impredecibles y genuinamente novedosas, entre otras características clásicas de dichas propiedades<sup>6</sup>. Sin embargo, según mi propuesta, la propiedades de nivel no emergen en el sentido habitual de la palabra, sino que son *resultan-*

<sup>4</sup>Para una correcta descripción de nivel aporoblemático, ver nota 2 en (Cáceres y Saborido 2017, 96)

<sup>5</sup>Por sencillez explicativa comienzo en los átomos, aunque igualmente podríamos partir de leptones y quarks, pasando por hadrones.

<sup>6</sup>Para una descripción detallada, ver (Klee 1984, 48) y (Cáceres y Saborido 2017, 93)

tes de las propiedades subyacentes en un sentido físico<sup>7</sup>. Hay varios tipos de razones para que dichas propiedades resulten *inesperadas* y *novedosas*: una de ellas es la ignorancia, como ya apuntaron (Mill 1843, 269) y (Spaulding 1912, 241) tiempo atrás; otra se debe a cuestiones heurísticas, pues una vez que se consigue predecir con cierta exactitud se vuelve (más o menos voluntariamente) innecesario investigar más; una tercera se debe a la rígida diferenciación entre entidad y entorno, que elimina *a priori* posibles influencias; otra es la forma clásica de investigar mediante análisis *top-down*, que delimita los fenómenos desde un nivel de observación preelegido. No obstante también puede ser por entender la emergencia como lo hace Mario Bunge, esto es, como propiedad que ninguno de sus componentes posee (Bunge 2004, 32).

De esta concepción se deriva que, conforme ascendemos niveles (como ICD) vamos encontrando *entidades-sistema* con *propiedades de nivel* arbitrariamente estables contrariamente a la idea de entidad clásica con ciertas propiedades resultantes y ciertas propiedades emergentes. Estas entidades-sistema interactúan con otras de su mismo nivel conforme a sus propiedades configurando un *mecanosistema*<sup>8</sup> que visto desde la perspectiva superior supone una entidad-sistema del siguiente nivel.

## Conclusiones

Considerando lo anterior, en un mundo cuasi-descomponible, las propiedades emergentes son propiedades de nivel derivadas de una elección arbitraria. De igual manera, la causación internivel no reflexiva, que obviamente depende de la consideración de descomponibilidad que permite distinguir un todo de otro, se hace indistinguible de la reflexiva<sup>9</sup> al ser la distinción entre entidades una elección a priori. Por tanto, el concepto de intervalo de cuasi-descomponibilidad es una herramienta que nos permite analizar y caracterizar la idea cierre como un método explicativo heurístico de forma que podemos hacer una visión reduccionista no eliminativista de la perspectiva organizacional.

## Bibliografía

- Alexander, Samuel. 1920. *Space, Time, and Deity*. Editado por Mcmillan. Vol. 2 vols. London.
- Anderson, Carl. 2002. «Self-organization in relation to several similar concepts: Are the boundaries to self-organization indistinct?». *The Biological Bulletin*, 202(3):247–255.
- Avery, Oswald T, Colin M MacLeod, y Maclyn McCarty. 1944. «Studies on the chemical nature of the substance inducing transformation of pneumococ-

---

<sup>7</sup>Es decir, exclusivamente la suma de las partes.

<sup>8</sup>Un mecanosistema (MS) es un conjunto de entidades-sistema (E-S) pertenecientes a un intervalo de cuasi-descomponibilidad (ICD) cuya organización y propiedades (P) son debidas exclusivamente a las propiedades de sus componentes. (Cáceres y Saborido 2018, 143)

<sup>9</sup>Ver un tratamiento extenso de la causación internivel en (Cáceres y Saborido 2017, 102–4)

- cal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III». *Journal of Experimental Medicine*, 79(2):137–158.
- Broad, Charlie D. 1925. *The Mind and Its Place in Nature*. London: Routledge.
- Bunge, Mario. 2004. *Emergencia y convergencia*. Barcelona: Gedisa.
- Cáceres, Emilio, y Cristian Saborido. 2017. «Intervalos de cuasi-descomponibilidad y propiedades emergentes». *Theoría*, 32(1):89–108.
- Cáceres, Emilio, y Cristian Saborido. 2018. «¿Realmente mató la bacteria al coronel?: Perspectiva sistémica, causación internivélica e intervalos de cuasi-descomponibilidad en las explicaciones mecanísticas». *Theoría*, 33(1):129–148.
- Campbell, Donald T. 1974. «Downward Causation in Hierarchically Organized Biological Systems». in F. Ayala, and T. Dobzhansky (eds.), *Studies in the Philosophy of Biology*, Berkeley, University of California Press, 179–86.
- Craver, Carl F., y William Bechtel. 2006. «Top-Down causation without top-down causes». *Biology and Philosophy*, 22:547–563.
- Etxeberria, Arantza, y Jon Umerez. 2006. «Organismo y organización en la Biología Teórica: ¿Vuelta al organicismo?». *Ludus vitalis*, XIV(14):3–38.
- Griffith, Fred. 1928. «The significance of pneumococcal types». *Epidemiology & Infection*, 27(2):113–159.
- Hershey, Alfred D, y Martha Chase. 1952. «Independent functions of viral protein and nucleic acid in growth of bacteriophage». *The Journal of General Physiology*, 36(1):39–56.
- Klee, Robert. 1984. «Micro-Determinism and Concepts of Emergence». *Philosophy of Science*, 51:44–63.
- Lewes, George Henry. 1877. *Problems of Life and Mind, second series*. Boston: J. R. Osgood.
- Mayr, Ernst. 2005. *Así es la biología*. Barcelona: Debate.
- Mill, John Stuart. 1843. *A system of logic: ratiocinative and inductive*. [8th, 1872]. London: Longmans, Green, Reader, and Dyer.
- Moreno, Alvaro, y Matteo Mossio. 2015. *Biological autonomy. A Philosophical and Theoretical Enquiry*. Springer.
- Morgan, C. Lloyd. 1923. *Emergent evolution*. Editado por Williams y Norgate. 2ª edición (1927). London.
- Mossio, Matteo, Leonardo Bich, y Álvaro Moreno. 2013. «Emergence, closure and inter-level causation in biological systems». *Erkenntnis*, 78:153–178.
- Pattee, Howard H. 1972. «Laws and constraints, symbols and languages». En *Towards a Theoretical Biology*, editado por C. H. Waddington, 4:248–58. Edinburgh: Edinburgh University Press.

- Prigogine, Ilya. 1978. «Time, structure, and fluctuations». *Science*, 201(4358):777–785.
- Simon, Herbert A. 1962. «The Architecture of Complexity». *Proceedings of the American Philosophical Society*, 106(6):467–482.
- Spaulding, Edward G. 1912. *The New Realism: Cooperative Studies in Philosophy*. New York: The Macmillan Company.
- Spaulding, Edward G. 1918. *The New Rationalism: The Development of a Constructive Realism Upon the Basis of Modern Logic and Science Through the Criticism of Opposed Philosophical Systems*. New York.
- Sperry, Roger W. 1969. «A modified concept of consciousness». *Psychological review*, 76(6):532.
- Varela, Francisco. 1979. *Principles of biological autonomy*. New York: North Holland.
- Watson, James D, Francis HC Crick, y others. 1953. «Molecular structure of nucleic acids». *Nature*, 171(4356):737–738.

# Varieties of pluralism and their ontological implications: The case of earth system sciences

Lino Camprubí

The plurality of sciences and their mutual irreducibility has not discouraged scientists and philosophers to pursue the dream of a unified science. When confronted with the successive failures of this project, they have often attributed it to an epistemological distortion without ontological consequences. Taking for granted a continuous and unified chain of being, some of them argue that it is only a matter of time before human knowledge finally matches the unity of reality while some others seek to establish a more definitive divide between the fragmented version of reality to which humans have access to and the holistic connections that would lay deep in the realm of *noumena*.

In recent decades, however, scholars have called into question this faith on ontological monism. In particular, philosophers and historians of science have argued from realism that epistemological divides often match ontological fractures (Cartwright, 1989; Duprée, 1993). I propose to compare different versions of this argument drawing classifying criteria from historical ontology (Hacking, 2002). If ontological entities are evolutionary and often result from animal and human operations, the diversity of ways for coming into being might favor scientific approaches that work at different scales and with different kinds of assumptions and techniques. Moreover, if the different sciences define their objects historically (an argument that does not necessarily entail a relativist approach but always an operationalist one attentive to material contexts), the irreducibility of different kinds of objects might be simultaneously ontological and epistemological.

I intend to use Earth System Science(s) as a guiding example to ground the discussion on pluralism and historical ontology. From the second half of the 19<sup>th</sup> century onwards, the earth and environmental sciences achieved an unprecedented scale defined by global circulation and interconnection. While all came to define spheres of planetary reach—the biosphere, the hydrosphere, the lithosphere, and the atmosphere—their respective efforts to encompass all spheres under a single science were limited by one another. In the late Cold War context, *earth systems science* appeared as an attempt to unify the variety of earth and environmental sciences. I will assess the successes and failures of this approach keeping in mind the critiques from pluralism.

Despite crucial connections, there is not a mathematical model able to interconnect or reduce sciences as diverse as climatology, geology and oceanography. The juxtaposition of these disciplines works at the level of local environments, but its application to the historically recent notion of the *global environment* results



from a jump of both epistemological and ontological nature. I will provide a historical and operational definition of the different earth spheres, defining them as the result of the different scales used by the disciplines involved in their study (Cartwright, Hacking, Bueno). At the same time, I will argue through a virtuous circle that the fractured ontology of nature is what explains the availability of these various and irreducible scales.

# Ockham’s razor in a minimalist guise

M. Inés Corbalán  
Giulia Terzian

Scientists and philosophers usually agree that simpler theories are preferable to more complicated ones; or in other words, that simplicity is a positive value, and a desirable theoretical goal. At the same time, they often disagree on the interdependent questions of how simplicity is to be (i) interpreted, and (ii) justified. As pertains to (i), for instance, it is well known that the term ‘simplicity’ is used fairly loosely in much of the philosophical and scientific literature, often being equated with other values such as beauty, elegance, coherence, unification and parsimony [Bar16]. Furthermore, requirements of simplicity seem to enjoy a certain special authority *qua* instances of Ockham’s razor (OR). And yet, OR is itself fundamentally ambiguous. Like simplicity, for instance, OR is open to different interpretations (methodological, epistemic, pragmatic). Indeed, Sober [Sob15] has convincingly argued that there is not one but several razors: on the one hand, OR is invoked under multiple formulations (e.g.: “Don’t multiply postulations beyond necessity” [Bak16], “entities should not be multiplied beyond necessity”, “it is futile to do with more what can be done with fewer”, “plurality should not be posited without necessity” [Sob15]). On the other, it is far from clear whether ‘postulation’, ‘more’, ‘fewer’ refer to the entities, and/or the hypotheses being postulated; if ‘multiply’ and ‘plurality’ should be interpreted qualitatively or quantitatively; and how ‘beyond/without necessity’ should be understood.

Considerations of simplicity, economy and parsimony have also acted as guiding concerns throughout the development of generative linguistics. In particular, under the most recent proposal—the Minimalist Program [Cho95]—they are attributed to the discipline’s main object of study: the human language faculty. Indeed, MP is guided by the following question: To what extent is the language faculty an “*optimal* solution” to the interface conditions imposed by the external performance systems?

At the same time, concerns of theory-simplicity are still very much present in minimalist inquiry. What’s more, under MP simplicity &co take centre stage as research goals of the highest priority. This is unlike other areas of inquiry, where simplicity &co are often treated as low-priority values.

In sum, MP is guided by two kinds or families of simplicity considerations, which come under the heading of *economy conditions*: methodological economy and substantive economy [Cho04]; [HNG05].<sup>1</sup>

[...] methodological and substantive minimalism are not to be confused [...] However, the methodological and substantive facets may

---

<sup>1</sup>Also called methodological minimalism and substantive/ontological minimalism, respectively [Cho04]. Chomsky himself recognizes that this distinction is subtle, but real.

not be independent: quite plausibly, it is by struggling for an elegant and parsimonious theory of an object that one can truly understand to what extent such an object is an optimal solution to the external conditions it must face. [AG09]: 8

These two kinds of economy notions [...] promote a specific research strategy: look for the simplest theory whose operations have a least effort flavor. [HNG05]: 8

Paraphrasing from these and similar quotes from all over the literature, we can give an initial, informal characterization of these notions. First, *methodological economy* is a methodologically-driven effort to improve scientific theories. Principles of m-economy can be understood as general guidelines in the search of the best theory, by and large recognizable as counterparts of the familiar theory-values of parsimony and simplicity (OR) in philosophy and other sciences [HNG05].

*Substantive economy* in turn seeks to discover whether the object of inquiry itself—the human language faculty—is optimal. Principles of s-economy thus take specific linguistic entities—typically, representations and derivations—as their target.<sup>2</sup> Note that the emphasis on subject matter-simplicity is consistent with the main theoretical hypotheses of minimalism: that Universal Grammar is an *optimal* way of satisfying the requirements imposed on the language faculty by the external (articulatory-perceptual, conceptual-intensional) systems with which it interfaces; and that the language faculty itself is characterized by *optimal*, computationally *efficient design*.

In sum, m-economy and s-economy operate on different levels (theoretical, objectual); to this extent, there is good reason to study them separately. And yet there is a clear sense—as the above quotes suggest—that they are not entirely independent; although it is far from clear just what their relation is. Against this backdrop, our aim in this paper is threefold: (I) to arrive at a better understanding of the nature and role of m-economy and s-economy; (II) to shed light on the largely overlooked question as to the relationship between the two (for instance: are principles of s-economy particular instances of m-economy?) (III) to determine whether the familiar philosophical readings of OR can also be found to underlie principles of s-economy.

As a first step in this direction, we focus on two specific substantive conditions: the Inclusiveness Condition on derivations (IC) and the Full Interpretation Condition on representations (FI).

Inclusiveness is one of the *best design* conditions predicated of the language faculty, in the sense that it is assumed to be met by ‘a perfect language’. The condition states, roughly, that the computational system—the ‘engine’ of the language faculty—is not ‘creative’: its outputs do not contain anything ‘new’ relative to its inputs. Arriving at a precise understanding of its role turns out to be tricky, however. For one thing, IC is presented under several, non-equivalent formulations. For instance:

---

<sup>2</sup>There are numerous examples of s-economy principles targeting derivations: Procrastinate, Greed, Last Resort, Inclusiveness, to name just a few. By contrast there is only one principle of representational economy: Full Interpretation. [Cho95].

[IC] states that no new features can be added in the course of derivation of an expression. [Bis13]: 1

Insofar as the condition of inclusiveness holds, the syntactic objects are rearrangements of properties of the lexical items of which they are ultimately constituted. [Cho15]: 207

Despite their differences, the above quotes clearly suggest that IC operates as a *parsimony* principle: it curbs potential expansions of the ontology by preventing new entities—in this case, features—from entering the derivation. Even granting that this is so, there are a number of worries in store for MP. One such worry is the following. We’ve just seen that IC blocks overgeneration of linguistic entities during the derivation; as a result, it also acts as a safeguard at the output end—the representational level. However, IC imposes no such constraints at the input level—the numeration,  $N$ . More precisely, it seems that under all of its formulations, IC allows for the numeration (a multiset of lexical items) to be augmented freely: either by new features, or by copies of features already in  $N$ , or any other theoretical entities.

At first brush, one might think—as we did!—that this possibility is ruled out by the above-mentioned Full Interpretation principle. The latter states, roughly, that no expressions occur idly in grammatical representations. More precisely, FI dictates that the only elements allowed to appear in a representation are those that are properly ‘licensed’; that is, those that are legible at the corresponding interfaces<sup>3</sup> between the language module and the external A-P and C-I systems. To this end, all the (uninterpretable) features of the lexical items of the numeration must be checked or valued in the course of the derivation.<sup>4</sup> If they are, the derivation is said to converge: it successfully makes it through to the interfaces. Otherwise, the derivation crashes, and won’t be interpreted at the interfaces.

Suppose  $F$  is an uninterpretable feature that fails to be checked in the course of the derivation. Then one might suppose that—given FI, and given the optimal design hypothesis— $F$  also cannot occur in the input numeration. If this were so, one might then reason as follows: while IC prevents the computational system from multiplying entities—but places no restriction on the size of the ontology outside the derivation, and specifically in the numeration.—FI moderately achieves the latter anti-superfluity goal as a by-product, insofar as it filters out those entities that cannot be legitimately licensed. Thus one might entertain the idea that these two s-economy principles operate *jointly* as a parsimony principle.

In this talk we will begin by reviewing the various extant formulations of IC, highlighting differences in their content and potential ramifications. In particular, we will raise the question of what IC is actually intended to target: the derivation alone, or the derivation together with the input numeration. We will then discuss whether IC is more appropriately understood as a qualitative or as a quantitative parsimony principle. Ultimately, we will argue that under each of these guises IC falls short of delivering its intended results. We will then discuss

---

<sup>3</sup>PF and LF: phonological and logical level of representation, respectively.

<sup>4</sup>Roughly speaking, a feature is uninterpretable if it is not legible at either interface. Note that all the lexical items of the numeration must be incorporated in the course of the derivation.

and ultimately reject the above conjecture, according to which FI might ‘make up’ for IC’s shortcomings. Finally we will offer a different reading of the latter principle. Specifically, we will argue that IC is most charitably understood, not as a substantive economy principle but rather as a methodological maxim: roughly, a ‘minimalist’ version of Ockham’s razor.

## References

- [AG09] Jordi Fortuny Andreu and ngel J. Gallego. Introduction. The Minimalist Program and the concept of Universal Grammar. *Catalan journal of linguistics*, 8(1):7–15, 2009.
- [Bak16] Alan Baker. Simplicity. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab. Stanford University, winter edition, 2016.
- [Bar16] Edison Barrios. Simple is not easy. *Synthese*, 193(7):2261–2305, 2016.
- [Bis13] Petr Biskup. On the NTC and labeling. Manuscript, 2013.
- [Cho95] Noam Chomsky. *The Minimalist Program*. Current Studies in Linguistics 28. MIT Press, 1st edition, 1995.
- [Cho04] Noam Chomsky. *The Generative Enterprise Revisited: Discussions with Riny Huybregts, Henk van Riemsdijk, Naoki Fukui and Mihoko Zushi*. Mouton de Gruyter, 2004.
- [Cho15] Noam Chomsky. *The Minimalist Program*. Current Studies in Linguistics. MIT Press, 20th anniversary edition, 2015.
- [HNG05] Norbert Hornstein, Jairo Nunes, and Kleanthes K. Grohmann. *Understanding minimalism*. CUP, 2005.
- [Sob15] Elliott Sober. *Ockham’s razors: a user’s manual*. CUP, 2015.

# Analysis of the infinite gods paradox in classical particle mechanics. Representation of a causal problem

Amaia Corral-Villate

## Introduction

Benardete (1964) introduces the Infinite Gods paradox in the context of his problems of the infinite in the following form:

A man decides to walk one mile from A to B. A god waits in readiness to throw up a wall blocking the man's further advance when the man has travelled  $1/2$  mile. A second god (unknown to the first) waits in readiness to throw up a wall of his own blocking the man's further advance when the man has travelled  $1/4$  mile. A third god... ad infinitum. It is clear that this infinite sequence of mere intentions (assuming the contrary-to-fact conditional that each god would success if given the opportunity) logically entails the consequence that the man will be arrested at point A; he will not be able to pass beyond it, even though not a single wall will in fact be thrown down in his path. The before-effect here will be described by the man as a strange field of force blocking his passage forward.

Priest (1999) starts the discussion on the problem with an article in which he introduces a logical analysis of the paradox and Yablo (2000) continues the debate reasoning that it contains a logical impossibility. This last conclusion achieves a generalised acceptance in the scientific community.

## Objectives

My communication is based on the introduction of a proposal for a representation of the Infinite Gods paradox in the strict context of Classical Particle Mechanics. The methodology consisting in the analysis of a metaphysical paradox in the frame of a specific theory is in line with what Grünbaum (1967) defended concerning the study of the supertasks and has later been followed by other philosophers of science who introduce representations of different paradoxes of the infinite. The only published representation of the Infinite Gods paradox is introduced by Pérez Laraudogoitia (2016) but it includes the participation of external agents. Thus, no strictly mechanical representation of this paradox has been published yet.

I follow this methodology in order to instrumentalise the analytical power of Classical Mechanics theory and achieve the objective of deepening in the comprehension of the paradox. This profound understanding implies clarifying the type of problem that underlies it. Concerning this last point, the results of my dynamic analysis contrast the conclusion argued by Yablo (2000), namely that the paradox contains a logical impossibility. My next objective is to point out what I consider is erroneous in his diagnosis.

## Formal and conceptual structure

I first present and contextualise the Infinite Gods paradox and then introduce my dynamic representation of it explicitly stating other particle systems I base on to construct it. These configurations are:

- i) The system proposed by Alper and Bridger (1998) which itself is a modification of a system previously introduced by Pérez Laraudogoitia (1996). These authors propose their particle configuration with some other purposes and don't mention the Infinite Gods paradox, but the similarities with it are clear. Still, the evolution of this configuration does not reproduce all the conditions in the paradox and so it requires additional ideas.
- ii) The self-excitable systems proposed by Pérez Laraudogoitia (1996). These are systems composed by infinite particles interacting one-dimensionally by means of binary collisions in which energy and momentum are entirely transmitted in each of the interactions, but in which total energy and momentum of the system disappear as a result of the infinite sequence of collisions. Because of the time reversal symmetry property of Classical Mechanics, the reverse process characterizes a case of indeterminism and this is a useful property to complement the configuration described above.

After introducing my proposal of dynamic representation I use it to analyse the evolution of the situation that characterises the paradox and conclude that the outcome of it is the one corresponding to that described by Benardete (1964) in his original formulation.

I then highlight the achievement of an additional result consisting in the identification of the type of interaction that characterises the paradox, namely the impenetrability interaction at a distance. This concept is first introduced by Pérez Laraudogoitia (2005) and independently suggested by Smullyan (2008). It is, indeed, already pointed out as the underlying interaction in the non-mechanical representation of the Infinite Gods paradox proposed by Pérez Laraudogoitia (2016). Since it has classically been supposed that impenetrability interactions occur exclusively in contact conditions, the identification of a situation in which this interaction occurs at a distance is conceptually interesting and relevant.

These results demonstrate that the type of problem that underlies the paradox is causal and clearly contrast the conclusion defended by Yablo (2000). Consequently, last part of my communication is dedicated to make a diagnosis of the point in which I consider his reasoning is erroneous.

## Conclusions

In clear contrast to the generally accepted conclusion introduced by Yablo (2000) my mechanical representation of the Infinite Gods paradox results in concluding that the type of problem that underlies it is not logical but causal and that the outcome of it is the one described by Benardete (1964) in his original formulation.

In addition to the achievement of the main objective consisting in deepening in the understanding of the paradox and thus clarifying the type of problem that underlies it, the analysis of the dynamic problem of evolution via my mechanic representation possibilitates clarification on the type of interaction in it, namely the impenetrability interaction at a distance.

## References

- Alper, J. S. & Bridger, M. (1998). Newtonian supertasks: A critical analysis. *Synthese*, 114(2):355–369.
- Benardete, J. (1964). *Infinity: An essay in metaphysics*. Oxford: Clarendon Press.
- Grünbaum, A. (1967). *Modern science and Zeno paradoxes*. Middleton: Wesleyan University Press.
- Pérez Laraudogoitia, J. (1996). A beautiful supertask. *Mind*, 105(417):81–83.
- Pérez Laraudogoitia, J. (2005). Achilles' javelin. *Erkenntnis*, 62(3):427–438.
- Pérez Laraudogoitia, J. (2016). Tasks, subtasks and the modern Eleatics. In F. Pataut (ed.), *Truth, objects, infinity*. Cham, Switzerland: Springer.
- Priest, G. (1999). On a version of one of Zeno's paradoxes. *Analysis*, 59(1):1–2.
- Smullyan, R. (2008). Memories and inconsistencies. In E. D. Demaine, M. L. Demaine & T. Rodgers (eds.), *A lifetime of puzzles*. Massachusetts: Taylor & Francis Group.
- Yablo, S. (2000). A reply to new Zeno. *Analysis*, 60(2):148–151.



# Justifying Lewis's kinematics of chance

**Patryk Dziurosz-Serafinowicz**  
University of Gdańsk

If there are chances (physical probabilities) attributable to the world, how can we account for the way they change over time? David Lewis (1986, p. 101) famously claimed that “a later chance distribution comes from an earlier one by conditionalizing on the complete history of the interval in between”. Interestingly, Lewis’s claim (hereafter, Lewis’s kinematics of chance, or LKC for short) parallels the well-known diachronic tenet of Bayesian epistemology, which says that a rational agent’s later credence function comes from an earlier one by conditionalizing on the total evidence gained in the interval in between. Moreover, much like Bayesian epistemologists who employed various arguments (most notably, Teller 1973; Lewis 1999; Brown 1976; Skyrms 1987; van Fraassen 1989; Maher 1992; van Fraassen 1999; Greaves and Wallace 2006; Leitgeb and Pettigrew 2010; Easwaran 2013) to show that Bayesian conditioning can be justified from more fundamental rationality requirements for credence (subjective probability), Lewis argued that his kinematical model for chance follows from the Principal Principle—the fundamental requirement for chance, according to Lewis.

Although recently there has been some interest in studying Lewis’s kinematics of chance (Lange 2006; Fisher 2006), remarkably little attention has been devoted to its justification. The main goal of this paper is to alter this situation by first giving more insight into Lewis’s own justification of his kinematics of chance, and then by motivating and providing a novel justification of it.

It is argued that, though initially attractive, Lewis’s own justification of his kinematics of chance is in tension with certain claims about chance kinematics that can be drawn by appealing to two other plausible chance-credences norms—the New Principle, advocated by Hall (1994) Thau (1994) New Principle, and the General Recipe proposed by Ismael (2008). I argue that this tension cannot be easily resolved in favour of the Principal Principle-based justification of LKC.

It is then shown that, under a fairly plausible assumption, Lewis’s kinematics of chance is equivalent to a particular principle connecting prior and possible posterior chances, which I call the *Generalized Chance-Chance Principle*. It requires chances at a particular time to be equal to a convex combination of possible chances at some later time. In the literature, this principle is regarded as a well-supported requirement for chance: either it is claimed to be a platitude about chance (Bigelow et al. 1993), or it is argued that the principle is an essential feature of resilient chance functions (Dziurosz-Serafinowicz 2017). So our new justification of Lewis’s kinematics of chance shows that this kinematical rule cannot be violated on pain of violating a well-established requirement of chance.

This new justification in turn gives us a useful perspective from which we can explain why so-called self-undermining chances may fail to satisfy Lewis's kinematics of chance. It is well known that those chances lead to probabilistic incoherence when constrained by the Principal Principle (see, for example, Thau 1994; Lewis 1994; Pettigrew 2015). We point out yet another serious problem in the conception of self-undermining chances, i.e., we show that since they may violate the Generalized Chance-Chance Principle, they may consequently violate Lewis's kinematics of chance.

I also show that, without additional assumptions, the Generalized Chance-Chance Principle entails any kinematical rule for chances which is *rigid*, i.e., it preserves the values of prior chances conditional on the intervening history in the posterior chances conditional on that history. This result in turn provides an interesting possibility of applying kinematical rules that are analogous to so-called Jeffrey conditionalization to situations where chances evolve in the interval from time  $t$  and  $t'$ ,  $t' > t$ , independently of the history that transpires in that interval.

## References

- Bigelow, J., Collins, J., and Pargetter, R. (1993). The Big Bad Bug: What are the Humean's Chances? *British Journal for the Philosophy of Science*, 44(3):443–462.
- Brown, P. M. (1976). Conditionalization and Expected Utility. *Philosophy of Science*, 43(3):415–419.
- Dziurosz-Serafinowicz, P. (2017). Chance, Resiliency, and Humean Supervenience. *Erkenntnis*.
- Easwaran, K. (2013). Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection. *Philosophy of Science*, 80(1):119–142.
- Fisher, J. C. (2006). On Higher-Order and Free-Floating Chances. *British Journal for the Philosophy of Science*, 57(4):691–707.
- Greaves, H. and Wallace, D. (2006). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, 115(459):607–632.
- Hall, N. (1994). Correcting the Guide to Objective Chance. *Mind*, 103(412):505–518.
- Ismael, J. (2008). Raid! Dissolving the Big, Bad Bug. *Noûs*, 42(2):292–8211.
- Lange, M. (2006). Do Chances Receive Equal Treatment Under the Laws? Or: Must Chances Be Probabilities? *British Journal for the Philosophy of Science*, 57(2):383–403.
- Leitgeb, H. and Pettigrew, R. (2010). An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy. *Philosophy of Science*, 77(2):236–272.

- Lewis, D. (1986). A Subjectivist's Guide to Objective Chance. In *Philosophical Papers*, volume 2, pages 83–132. Oxford: Oxford University Press.
- Lewis, D. (1994). Humean Supervenience Debugged. *Mind*, 103(412):473–490.
- Lewis, D. (1999). Why Conditionalize? In *Papers in Metaphysics and Epistemology*, volume 2, pages 403–407. Cambridge: Cambridge University Press.
- Maher, P. (1992). Diachronic Rationality. *Philosophy of Science*, 59(1):120–141.
- Pettigrew, R. (2015). What Chance–Credence Norms Should Not Be. *Noûs*, 49(1):177–196.
- Skyrms, B. (1987). Dynamic Coherence and Probability Kinematics. *Philosophy of Science*, 54(1):1–20.
- Teller, P. (1973). Conditionalization and Observation. *Synthese*, 26(2):218–258.
- Thau, M. (1994). Undermining and Admissibility. *Mind*, 103(412):491–504.
- van Fraassen, B. (1989). *Laws and Symmetry*. Oxford: Oxford University Press.
- van Fraassen, B. (1999). Conditionalization, a New Argument For. *Topoi*, 18(2):93–96.

# Knowledge vs understanding: Epistemic scenarios and scientific practice

Johannes Findl

Two goals occupy center stage in my talk. First, I argue that a difference between knowledge and understanding can be established by appealing to hypothetical epistemic scenarios where one but not the other can be had. For this purpose, I scrutinize two views in favor of this difference, one by Duncan Pritchard and the other by Jonathan Kvanvig. Kvanvig's view is that understanding why something is the case requires the grasping of relevant explanatory connections within a large body of information. Understanding is a cognitive ability that a subject has to continuously exercise regarding a certain problem which "requires the grasping of explanatory and other coherence-making relationships in a large and comprehensive body of information. One can know many unrelated pieces of information, but understanding is achieved only when informational items are pieced together by the subject in question" (Kvanvig 2013).

Pritchard has defended the view that what distinguishes understanding from knowledge is its compatibility with a specific form of epistemic luck and its distinctive relation to the exercise of one's cognitive skills.

I then shortly present two noteworthy epistemic scenarios—the burnt house and the Comanche scenario—which supposedly establish the difference between understanding and knowing.

The Comanche scenario is structurally analogous to so-called fake barn cases: a subject forms true and justified beliefs about the Comanche dominance in the southern plains in North America by reading a truthful history book about it. The scenario supposedly shows that while falling short of knowledge (if, say, her beliefs are only true as a matter of luck), the subject may achieve understanding of the Comanche dominance.

The burnt house scenario describes a case where a young child acquires knowledge of why his house burnt down yet does not understand why since she supposedly lacks a conception of how this come about.

While admittedly ingenious, I claim that these scenarios do for independent reasons not succeed in driving a wedge between knowledge and understanding. Neither of them provides necessary and sufficient conditions for the knowledge-understanding distinction.

I then present an argument for the claim that although non-reductionism about understanding, i.e. the view that understanding cannot be reduced to knowledge, appears as an initially attractive position, it turns out to be untenable. The initial attractiveness derives its support from two considerations. On the one hand, in ordinary language use, some constructions of sentences of the form 'I

know X but do not understand X' seem permissible and even felicitous. Such linguistic data is then often interpreted as to suggest that there might be a difference between knowledge and understanding to be investigated further. On the other hand, epistemologists have almost exclusively analyzed knowledge of single propositional sentences and have paid less attention to knowledge of more complex sets of interrelated propositional sentences. Most of them believe that one knows a complex set consisting of interrelated propositional sentences iff one knows every sentence in that set and knows how the sentences relate to each other. According to this strict requirement, if understanding is nothing over and above knowledge, an experienced physician who does not know every sentence that a theory X contains does not understand X. This is an undesirable consequence, since it seems at odds with our practice of charitably ascribing understanding of a theory to someone who may not know all its sentences and might even have some false beliefs about it. Theorists have in response developed various conceptions of understanding with the goal of alleviating this worry, either going the non-factual route which allows for understanding in the absence of having only true beliefs or, notwithstanding this matter, arguing that understanding is an ability that concerns a different level of cognition than knowledge.

Against the initial attractiveness, I firstly maintain that the analysis of ordinary uses of the terms *understanding* and *knowledge* reveals that even in the most felicitous cases, understanding locutions can be substituted by knowledge locutions if they are understood in a broad sense as to entail knowledge how (or more formally: explanatory knowledge) without any loss. Secondly, I argue that knowledge of complex sets of interrelated propositional sentences can be had even when one does not know all of these sentences and even when one has some non-central false beliefs attaining to these sets. The strict requirement from above then fails to hold, which arguably lessens the motivation for the non-reductionist left to embrace a different notion than knowledge on grounds of allowing for non-factivity. Thirdly, I raise some doubts about the feasibility of conceiving of understanding as a cognitive ability which is substantially different from the notions of coming to know and knowledge-how.

These conclusions are then applied to an alternative epistemic scenario where a scientist intends to account for a puzzling phenomenon running observational experiments, and finally succeeds. I intend to show that this scenario exhibits all three features outlined in Kvanvig's and Pritchard's scenarios, which supposedly support the distinctiveness of understanding. Second, I show that the three features my scenario exhibits are also constitutive of ordinary scientific practice: I claim that a special sort of epistemic luck, the exercise of distinctive cognitive skills and the grasping of connections within a large body of information are all constitutive of the context in which scientific research is carried out. Yet I think that the best way to accommodate them is to say that they describe the process of coming to know or coming to understand, which ultimately both yield knowledge. I try to show this via the historical example of the water channel which the German physicist Ludwig Prandtl once developed to visualize basic flow experiments. Prandtl was investigating the effects of friction (rapidly changing velocities) and concerned with the question of why these effects only occur *near* an object moving through a fluid? For the purposes of this talk, I shall divide his inquiry into different phases. In the first phase, Prandtl settled the questions to addressed with respect to the phenomenon under investigation, his

hypothesis being that there is zero relative velocity between fluid and wall but sharp changes in velocity in a so-called boundary layer and therefore very large shear stress. Careful and repeated observation of this phenomena in the water channel finally paved the way for Prandtl's correct and still valid explanation of the original question. Namely, that there are two regions of aerodynamic flow: *boundary layer* with friction and inviscid external flow without friction. External inviscid flow creates the boundary conditions at the edge of the boundary layer and therefore the velocity profile within the layer. On the other hand, the boundary layer is so thin that it has virtually no effect on the outer inviscid flow. In a nutshell, Prandtl was later able to find missing equations for Fluid Mechanics and Aerodynamics. I argue that this example shows that Prandtl's could only succeed in his research since, besides exercising his skills appropriately, he was lucky in Pritchard's sense. Moreover, following Kvanvig, I argue that Prandtl's underlying thought processes are best accounted for in terms of beliefs which are embedded in a large body of information and some of which are beliefs with less than full credence, hence not qualifying as knowledge. Having said that, I apply a distinction between *coming to know* and *knowing* in order to give the most adequate description of the Prandtl scenario. This distinction does not only neatly accommodate the various levels of Prandtl's thought processes vis-à-vis his research, it also establishes that the fact that Prandtl's final solution, the missing equations, can only be accounted for in terms knowledge that he acquired in his research. Pace Pritchard and Kvanvig, I claim that what they think is the sui generis notion of understanding boils down to what I call *coming to know* (or: coming to understand), but that the end product of Prandtl's research is knowledge simpliciter. Hence a genuine difference between knowledge and understanding cannot be inferred from this or similar examples, which is the main conclusion I draw.

# Truthlikeness for quantitative deterministic laws

Alfonso García Lapeña

## Abstract

Truthlikeness is a property of a theory or a proposition that represents its *closeness to the truth* of some matter. According to the *similarity approach* to say that a theory  $t_1$  is more truthlike than a theory  $t_2$  means that (i) given a space  $S$  where  $t_1$ ,  $t_2$  and the truth in question  $t^*$  are represented and (ii) given an *appropriate* metric  $d$  for  $S$ , then (iii)  $d(t_1, t^*) < d(t_2, t^*)$ . Scientific *quantitative deterministic laws* (QDL) typically have a real function representation  $f(x_1, \dots, x_n)$  in some  $n$ -dimensional space. According to the Niiniluoto (1982, 1987, 1998) and Kieseppä (1996), among others, we can define truthlikeness for QDL by the Minkowski metric. We will present two counterexamples to this definition and argue that the problem lies in the fact that it considers truthlikeness for QDL to be a function of just *accuracy*, but an accurate law can be completely wrong about the “causal structure” of the world. Finally, we will present a proposal where the truthlikeness of a scientific law is defined according to two parameters: *accuracy* and *nomicity*.

Truthlikeness is a property of a theory or a proposition that represents its *closeness to the truth* of some matter. Consider:

- a) The number of planets in our Solar System is 10.
- b) The number of planets in our Solar System is 10 billion.

As the number of planets is actually 8, both propositions are false. However, intuitively, they don't seem to be *equally false*: a) seems, in some sense, *less false* or *closer* to the truth in question, to how things actually are, than b).

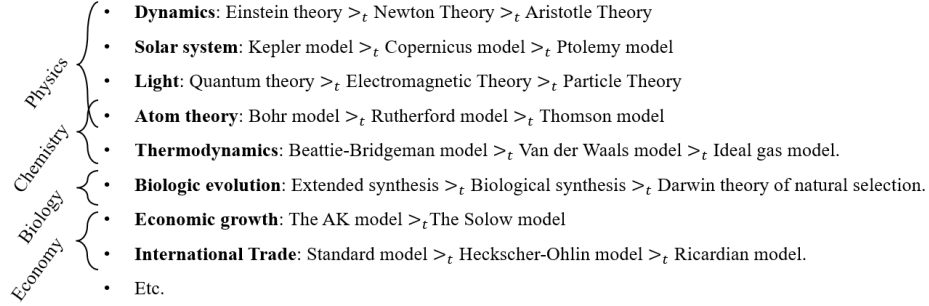
Consider:

- c) The shape of the Earth is a sphere.
- d) The shape of the Earth is a cube.

Both propositions are false, as the actual shape is something like an oblate spheroid. But, again, they don't seem to be *equally false*: c) seems *less false* than d), c) seems *closer* to the real shape of the Earth than d).

This relation also seems to stand regarding a large number of scientific theories (take ' $>_t$ ' to mean 'closer to the truth than'):

What these examples try to show is that, intuitively, *we don't consider all falsehoods to be on a par*. We consider some false theories or propositions to be



closer to the truth, nearer the truth, more approximate to the truth, more similar to the truth in question, than others. These considerations is what the notion of *truthlikeness* stands for.

Popper (1972) was the first one to take the concept seriously and to give a formal definition of it. However, two years later Miller (1974) and Tichý (1974) proved independently that Popper's definition was logically inconsistent. Since then, different strategies have been developed to try to define the notion. The so called "similarity approach" has been the most fruitful and the one that has attracted most attention.

On this approach (firstly developed by Hilpinen (1976) and Tichý (1974) and hugely expanded by Oddie (1986) and Niiniluoto (1987) among others) the truthlikeness of a theory or a proposition depends on the *similarities* between the state of affairs it allows and the actual state of affairs of the world.

The core idea of the similarity approach can be summarized in five points:

- (1) Given some phenomenon  $P$  and some language  $L$  to talk about  $P$  we can construct a space of possibilities  $S_p^L$  which contains all the ways the world ( $P$ ) can be regarding  $L$ . We can think of each point or cell of  $S_p^L(i, j, k, \dots)$  as a possible world.
- (2) Any theory or claim  $h$  of  $L$  will be expressible as an element or set of elements of  $S_p^L$ .
- (3) We introduce a *metric*  $d(i, j)$  which defines the distance between elements of  $S_p^L$  and an *extension* of  $d$  into another metric  $d'(h, s)$  which defines the distance between theories or claims.
- (4) Some element  $t^*$  will be the truth in question, the actual world.
- (5) The truthlikeness  $Tr$  of a claim  $h$ , when  $d'$  is normalized, can be defined as:

$$Tr(h) = 1 - d'(h, t^*)$$

Where:

- $Tr(t^*) = 1 - d'(t^*, t^*) = 1$
- $Tr(h) > Tr(s) \leftrightarrow d'(h, t^*) < d'(s, t^*)$



Therefore, what does it mean to say that a theory  $t_1$  is more truthlike than a theory  $t_2$ ? That (i) given a space  $S$  where  $t_1$ ,  $t_2$  and the truth in question  $t^*$  are represented and (ii) given an appropriate metric  $d$  for  $S$ , then (iii)  $d(t_1, t^*) < d(t_2, t^*)$ .

This is a general definition of truthlikeness, where the concept is *quantified* (according to a measure of similarity) and can be used to *order* different rival answers to some specific cognitive problem (Niiniluoto, 1987). The crucial point is (3), the introduction of an *appropriate* metric. In many spaces we have already a natural way of measuring the distances (similarity) between elements. For example, if our space is the set of real numbers the distance between two elements  $x, y$  is naturally defined as  $d(x, y) = |x - y|$ . As we will see in a moment, the kind of spaces where quantitative deterministic laws are represented already present natural metrics. However, some spaces do not present a straightforward natural metric. One of the main difficulties of the similarity approach has been to develop an *appropriate* metric for qualitative statements, represented in spaces produced by first and higher order languages.

What science usually does when studying some phenomenon  $P$  composed by  $N$  individuals interacting in a certain way is to postulate some magnitudes  $m$  that those individuals instantiate—in different values—and to explain their interactions according to some mathematical relations among the magnitudes. In the simplest case where we have just one individual and  $n$ —scalar—magnitudes, all the possible states  $\langle m_1, \dots, m_n \rangle$  of the system can be naturally represented in a mathematical  $n$ -dimensional space  $S^n$ . For example, in classical mechanics a particle has at each instant some position  $r^3 = (r_x, r_y, r_z)$  and some velocity or momentum  $p^3 = (p_x, p_y, p_z)$ . Therefore,  $S^n$  can be taken to be  $\mathbb{R}^6$ , being each point or state a 6-tuple of real numbers  $(r_x, r_y, r_z, p_x, p_y, p_z)$ .

In that kind of infinite, continuous and multidimensional  $S^n$  spaces a *quantitative deterministic law*  $A$  will typically have a real function representation  $f^A(x_1, \dots, x_n)$  and the truth in question  $T$  (the true connexion between magnitudes  $x_1, \dots, x_n$ ) will correspond to some  $f^T(x_1, \dots, x_n)$ . As we have seen, according to the similarity approach what we need to define truthlikeness given some representational space is an appropriate *metric*, in our case an appropriate metric for functions. This distance is given by the Minkowski metric for functions<sup>1</sup>:

$$d(A, T) = \int \left( |f^A(x) - f^T(x)|^k dx \right)^{\frac{1}{k}}$$

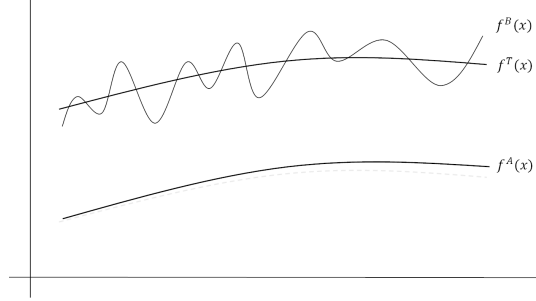
The three most used in mathematics are the Manhattan, the Euclidian and the Chebyshev:

- *Manhattan distance* ( $k = 1$ ):  $d^{ma}(A, T) = \int |f^A(x) - f^T(x)|$
- *Euclidean distance* ( $k = 2$ ):  $d^{eu}(A, T) = (\int |f^A(x) - f^T(x)|^2)^{\frac{1}{2}}$
- *Chebyshev distance* ( $k \rightarrow \infty$ ):  $d^{ch}(A, T) = \sup |f^A(x) - f^T(x)|$

Niiniluoto (1982, 1987, 1998) and Kiesepä (1996), among others, consider that  $d^{ma}$  and  $d^{eu}$  properly define truthlikeness for quantitative deterministic laws.

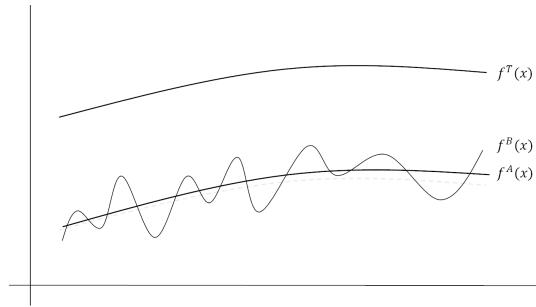
<sup>1</sup> Always for some interval  $C$ .

However, consider the following case (presented by Thom (1975), Weston (1992), Liu (1999) among others):



Which law,  $A$  or  $B$ , is intuitively more truthlike? According to  $d^{ma}$  and  $d^{eu}$ ,  $B$  is closer to  $T$  than  $A$  because  $d(A, T) > d(B, T)$  and therefore  $B$  is more truthlike,  $Tr(B, T) > Tr(A, T)$ . However, in some intuitive “sense”  $A$  seems to be more truthlike than  $B$  as it is closer to “the shape” of  $f^T(x)$ , to the *actual true relation between the magnitudes*.  $B$  seems to totally fail in representing the real way in which the magnitudes are related in the world. This is why Thom (1975) argues that in such situation scientists will typically prefer law  $A$  to law  $B$ , Liu (1999) also claims that  $A$  should be considered more truthlike than  $B$  contrary to what available approaches of truthlikeness conclude, and Weston (1992) argues that because of our incapability to measure the *shape* of the functions we can’t *a priori* determine the truthlikeness of  $A$  or  $B$ , and therefore that truthlikeness for quantitative laws becomes context-dependent: for some aims we will consider law  $A$  to be more truthlike—for instance, for qualitative predictions—and for others we will prefer law  $B$ —for instance for accurate predictions.

Although  $A$  is closer to the actual true relation between the magnitudes, so in that sense seems to be more truthlike than  $B$ ,  $B$  is a much more accurate law, so in that other sense seems to be more truthlike than  $A$ . In order to overcome this discrepancy consider the following modified case:



And suppose two situations:

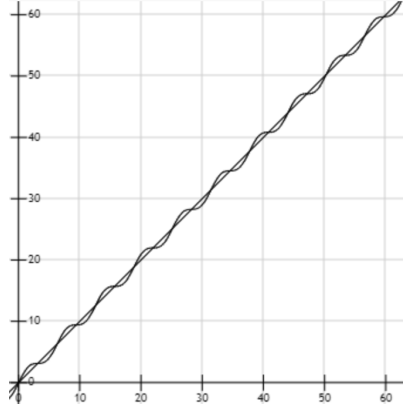
- (1)  $d(A, T) = d(B, T)$
- (2)  $d(A, T) = x$  and  $d(B, T) = x - dx$ , meaning that  $d(B, T)$  is just infinitesimally smaller than  $d(A, T)$ .

According to available theories of truthlikeness, in (1)  $Tr(B, T) = Tr(A, T)$  and

in (2)  $Tr(B, T) > Tr(A, T)$ . I take it that either in (1) or (2) we have strong intuitions that  $A$  should be considered more truthlike than  $B$ , contrary to what available approaches predict. Therefore, it seems that we have a clear counterexample to available theories of truthlikeness regarding quantitative laws.

The problem relies in the fact that available approaches take  $Tr$  to be just a function of *accuracy*, but an accurate law can be completely wrong about the “causal structure” of the world. Suppose a very simple case where a particle is moving on a straight line at a constant speed of  $1\text{ m/s}$ . The true function that describes the movement of the particle is  $y^T = x$  (being  $y$  position and  $x$  time). Suppose we postulate law  $A$  as a description of the particle’s movement:

- $y^T = x$
- $y^A = x + \sin x + \frac{e}{(x+5)^{10}} - \pi \cdot 10^{-3}$



Law  $A$  is highly accurate regarding the position of the particle (with a maximum error of  $0.9968\text{ m}$  and an average error of  $0.6334\text{ m}$  in predicting it). However, it is completely wrong about the actual true relation between position and time. According to law  $A$ :

- The movement of the particle is not linear.
- The particle has an oscillating acceleration.
- The position of the particle depends to some degree on the irrational numbers  $e$  and  $\pi$ .
- The position is, also to some degree, inversely proportional to time.
- The net sum of the forces acting on the particle is never zero neither constant.

All those facts are radically false regarding the actual *way* in which the variables are related in the world. Law  $A$ , although being highly accurate, is in a sense completely wrong about *how the world is*, about how position and time are actually related. Therefore, *accuracy* seems not enough to measure truthlikeness on deterministic laws. The  $Tr$  of a law should also take into account something like the “shape” of the law, its “nomicity”, its degree of “lawlikeness”, how well it captures the “actual causal structure” of the world or the actual “way” in which the variables are related in reality.

My proposal is that the *derivative* can solve this problem. The derivative of a point measures the *behaviour* of the function in the proximities of that point. It tells us *how* the variables are related near the point: in the proximities of point  $c$ , if we increase variable  $x$  by  $dx$ , then variable  $y$  will increase by  $a \cdot dx$ , being  $a$  the value of the derivative for that point. If two corresponding points of two functions have the same derivative that means that in the proximities of these points both functions “behave” equally. Finally, if two functions have a very similar derivative in each of its corresponding points that means that they define a very similar behaviour between the variables across all the domain, and so that their ‘shapes’ are very similar.

My proposal, then, is to judge the *similarity* ( $d^{id}$ ) between two laws  $A$  and  $B$  according to two parameters: *accuracy* and *nomicity*.

$$d^{id}(A, B) = F(\text{accuracy}, \text{nomicity})$$

Accuracy is well measured by the quantitative distance given by the Minkowski metric. Nomicity can be measure by the square root of the squared difference of the derivatives<sup>2</sup>. Opting for  $k = 2$  in the Minkowski metric (to maintain elegance with the other factor) we obtain (always for some interval  $C$ ):

$$d^{id}(A, B) = \gamma \left( \int (f^A(x) - f^B(x))^2 dx \right)^{\frac{1}{2}} + (1 - \gamma) \left( \int (f^{A'}(x) - f^{B'}(x))^2 dx \right)^{\frac{1}{2}}$$

Which satisfies the four requirements for metric functions:

- $d^{id}(A, B) \geq 0$
- $d^{id}(A, B) = 0 \Leftrightarrow A = B$
- $d^{id}(A, B) = d^{id}(B, A)$
- $d^{id}(A, C) \leq d^{id}(A, B) + d^{id}(B, C)$

Variable  $\gamma$  ( $0 \leq \gamma \leq 1$ ) is a measurement of the relative importance of each parameter, introduced to allow the possibility that they don’t have the same weight if some *a priori* arguments can be given in favour of *accuracy* or *nomicity*. In the absence of those arguments we can ignore it or assign it  $1/2$ .

As we are working with infinite spaces, the normalization of  $d^{id}$  can be in many cases impossible (although we could always proceed by a pragmatic normalization establishing some contextual limits). For simplifying, we can define the degree of truthlikeness of a law  $A$  as:

$$Tr(A) = d^{id}(A, T) = F(\text{accuracy}, \text{nomicity})$$

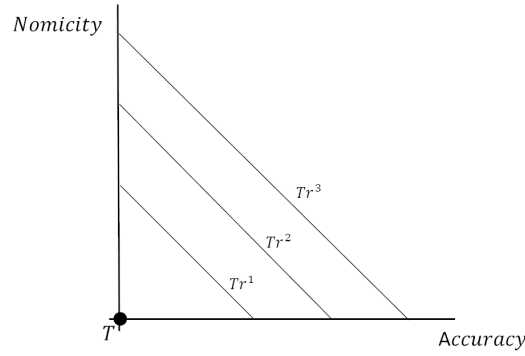
Where:

---

<sup>2</sup>A natural way to measure nomicity will be by the absolute value of the difference of the derivative:  $\int |f^{A'}(x) - f^{T'}(x)|$ . However, it does not work because of possible symmetries.

- $Tr(T) = d^{id}(T, T) = 0$
- $Tr(A) > Tr(B) \leftrightarrow d^{id}(A, T) < d^{id}(B, T)$

We can represent all possible laws as points in a two dimensional space according to their values on *accuracy* and *nomicity*. It's easy to see that for a given degree of truthlikeness we would have theoretically infinite laws, as there are infinite different value combinations of accuracy and nomicity that produce the same degree of truthlikeness. This can be represented with some level lines  $Tr^1$ ,  $Tr^2$ , etc in decreasing degree of truthlikeness. At the origin we find the true law in question  $T$ :



### Numerical example

Consider:

- $y^T = x + 5$
- $y^A = 0,98x$
- $y^B = x + \sin x + \frac{e}{(x+5)^{10}} - \pi \cdot 10^{-3}$

Range:  $[0, 30]$

According to the available approach:

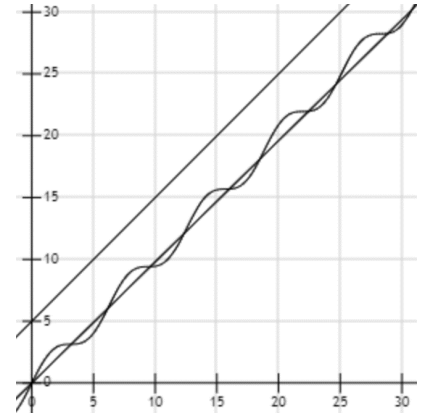
- $Tr(A) = d^{eu}(A, T) = 29,04$
- $Tr(B) = d^{eu}(B, T) = 27,52$

Therefore,  $Tr(B, T) > Tr(A, T)$ , against our intuitions.

According to the present proposal:

- $Tr(A) = d^{eu}(A, T) = 29,04 + 0,11 = 29,15$
- $Tr(B) = d^{eu}(B, T) = 27,52 + 3,86 = 31,38$

Therefore,  $Tr(A, T) > Tr(B, T)$ , which yields the correct result.



## References

- Hilpinen, R. (1976). "Approximate truth and truthlikeness", in Przelecki, et al. (eds.), *Formal Methods in the Methodology of the Empirical Sciences*. Dordrecht: Reidel, 19–42.
- Kieseppä, I. A. (1996). *Truthlikeness for Multidimensional, Quantitative Cognitive Problems*.
- Kuipers, T. (1987). *What is closer-to-the-truth? A parade of approaches to truthlikeness*. (Poznan Studies in the Philosophy of the Sciences and the Humanities, Volume 10), Amsterdam: Rodopi.
- Liu, C. (1999). "Approximation, idealization, and laws of nature", *Synthese*, 118(2):229–256.
- Miller, D. (1974). "Popper's Qualitative Theory of Verisimilitude", *British Journal for the Philosophy of Science*, 25(2):166–177.
- Niiniluoto, I. (1982). "Truthlikeness for Quantitative Statements", *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1982*, 208–216.
- Niiniluoto, I. (1982). *Truthlikeness*. Dordrecht: Reidel.
- Niiniluoto, I. (1998). "Verisimilitude: The Third Period", *British Journal for the Philosophy of Science*, 49(1).
- Oddie, G. (1986). *Likeness to Truth*. Reidel.
- Popper, K. (1972). *Conjectures and Refutations: The Growth of Knowledge*. 4<sup>th</sup> edition, London: Routledge & Kegan Paul.
- Thom, R. (1975). *Structural Stability and Morphogenesis: An Outline of a General Theory of Models*. Reading, MA: Addison-Wesley.
- Tichý, P. (1974). "On Popper's Definitions of Verisimilitude", *British Journal for the Philosophy of Science*, 25(2):155–160.
- Van Fraassen, B. (1970). "On the Extension of Beth's Semantics of Physical Theories", *Philosophy of Science*, 37(3):325–339.
- Weston, T. (1992). "Approximate Truth and Scientific Realism", *Philosophy of Science*, 59(1):53–74.

# La teoría relacional del tiempo como teoría constructiva y su poder explicativo

Mikel Henda Gómez de Segura

La idea del tiempo como medida del cambio, ya la sugería Mach (1912: 189), cuando afirma que “el tiempo es más bien una abstracción a la cual llegamos por la variación de las cosas, debido a que no está señalada ninguna medida determinada, por estar todas las cosas vinculadas entre sí. Llamamos uniforme a un movimiento en el cual a iguales incrementos de camino corresponde iguales incrementos de camino de un movimiento de comparación”. Efectivamente, cuando aportamos mediciones temporales como, “hace tres horas que  $x$  ocurrió” no estamos midiendo el tiempo, sino que medimos el cambio y lo hacemos a través de otro sistema que cambia. En realidad, lo que estamos haciendo con esa medición es afirmar que en nuestro reloj de referencia la manecilla que marca los minutos ha pasado ya en tres ocasiones por la misma posición. El tiempo es la medida del cambio al igual que el espacio es la medida de la extensión actual o hipotética de los objetos materiales y sus posiciones relativas (geometría).

Voy a defender que la concepción del tiempo relacional puede ayudar a aportar una explicación satisfactoria de los fenómenos relativistas de la dilatación temporal y la contracción de las reglas. Para ello, debemos tener en cuenta la capacidad explicativa de los diferentes modelos teóricos que se presentan. Balashov y Janssen (2003) distinguen dos tipos, las “teorías de principios”, que son aquellas que se construyen basándose en ciertas generalidades empíricas bien confirmadas a las que se les da carácter de postulados (en la TRE por ejemplo consta de dos postulados: el principio de la relatividad y el de la constancia de la velocidad de la luz en el vacío). En este tipo de teorías, la explicación consiste en mostrar que los fenómenos a explicar se dan necesariamente en el mundo como consecuencia de los postulados asumidos. En oposición está la “teoría constructiva”, que trata de explicar los fenómenos a través de leyes o de procesos subyacentes. Estas teorías ofrecen una serie de modelos de una parte de la realidad física (por ejemplo, la teoría cinética de los gases que explica la temperatura de un gas por la energía cinética de las moléculas que lo componen), por lo que la teoría es explicativa en tanto que aporta modelos empíricamente adecuados de la realidad. Así pues, la elección entre la preferibilidad de un modelo explicativo u otro, se basará en lo que entendamos por “explicar” o qué es una buena explicación.

Si trasladamos esta cuestión al análisis del tipo de explicación que ofrece la teoría de la relatividad acerca de los fenómenos temporales, nos encontramos con que es una explicación de principios. Tal y como Balashov y Janssen (2003: 331) reconocen, a la hora de explicar el fenómeno relativista de la contracción de reglas y la dilatación temporal, lo que esta teoría nos dirá es que en todo

universo en el que se cumplan tanto el postulado de la velocidad de la luz como el principio de relatividad, se darán estos efectos relativistas, sin embargo, no tendrá nada que decir sobre la realidad subyacente de estos fenómenos. A pesar de este hecho, consideran que este modo de explicación es el preferible apoyándose en dos argumentos de autoridad:

First, we note that theory-of-principle explanations appear to fit the once-standard covering law model of explanation much better than constructive-theory explanations. That, however, might just be another nail in the coffin of the covering law model. More importantly, we want to emphasize that (ii) [que las teorías de principios son explicativamente deficientes] also applies thermodynamics. That in and of itself, we submit, places the relativity interpretation in very good company (Balashov y Janssen 2003: 332).

La cuestión fundamental en último término es qué aceptamos como explicación: a) los relojes y los metros se comportan de una manera determinada debido a que es necesario que así sea por los postulados de la teoría, o b) bien, los relojes y los metros se comportan de un modo que resulta ser consistente con el principio de la relatividad y como consecuencia aportan siempre la misma medida de la velocidad de la luz para todos los observadores. Considero, junto con otros autores como Brown y Pooley (2006) y Pooley (2013), que la situación más adecuada de entender los fenómenos es de un modo constructivo y no a modo deductivo de principios generales. La geometría del espacio-tiempo no nos ayuda a comprender por qué los objetos materiales se comportan de un modo determinado, contrayéndose o ralentizando su evolución en función del entorno espacio-temporal en el que se encuentren. Incluso el propio Einstein prefería una explicación teórico constructiva por su poder explicativo:

... when we say we have succeeded in understanding a group of natural processes, we invariably mean that a constructive theory has been found which covers the processes in question (Einstein, 1982: 228).

Y es que, cuando logramos una teoría de principios, hemos conseguido relacionar de un modo estructurado y lógico las diferentes variables de los fenómenos observados, consiguiendo de este modo una herramienta altamente generalizable y potente a la hora de aplicarla a diferentes situaciones y realizar predicciones. Sin embargo, no nos aporta justificación del comportamiento de dichos sistemas, para ello no nos queda más remedio que recurrir a sus propiedades dinámicas. Consideremos la siguiente situación a modo de ejemplo: a la hora de explicar el comportamiento de los gases, podríamos proceder estableciendo ciertos principios y relaciones entre presión, temperatura, volumen y cantidad de gas. Si conseguimos relacionar estas variables adecuadamente y de un modo lógico, podremos conseguir una teoría a través de la cual realizar predicciones acertadas. Pero esta teoría no podrá decirnos por qué cuando aumentamos o reducimos la temperatura, la presión del gas se ve afectada. Para ello necesitaríamos una teoría constructiva, como la teoría cinética de los gases, que nos dijese que el calor es la energía cinética de las moléculas del gas, por lo que cuanto mayor temperatura tiene un gas, mayor energía tiene y por lo tanto mayor capacidad para ejercer fuerza sobre las paredes del recipiente que lo contiene. De un modo similar, al acelerar un coche, podemos establecer matemáticamente la relación



entre la presión ejercida en el pedal de aceleración y la velocidad que el coche pueda alcanzar en una situación ideal (en una superficie plana, sin resistencia del viento, etc.). Sin embargo, estaríamos dejando fuera todo el mecanismo interior del coche, que nos explicaría no solo cómo se relacionan ciertas variables (velocidad y presión en el acelerador), sino que, además, explicaría el cómo. Así pues, si nos conformamos con una explicación de principios, podríamos estar dejando fuera de nuestro marco de análisis e investigación una buena parte de la realidad.

Con esto no quiero decir que las teorías de principios sean inferiores a las teorías constructivas, muy por el contrario, las primeras tienen un gran poder de predicción y de deducir consecuencias que emanan de los propios postulados de la teoría debido a su sistematicidad. Además, en muchas ocasiones los postulados de estas teorías han tenido gran contrastación científica. Más bien, considero que cada una tiene su función, de modo que, si lo que queremos es disponer de un fuerte aparato deductivo que nos permita realizar predicciones novedosas y altamente fiables, probablemente, el modelo de teoría de principios sea la adecuada. Sin embargo, si lo que pretendemos es comprender en profundidad una parcela de la realidad, considero que en ese caso la teoría constructiva es la adecuada. Por lo tanto, tal vez no sería necesario verlas como opciones opuestas, sino como dos vías complementarias en la investigación científica.

Considero, que una interpretación del tiempo como cambio en los sistemas, puede ayudar a la construcción de una teoría constructiva que ayude a comprender de un modo más profundo los fenómenos relativistas. De este modo, se abre la puerta a la interpretación de que la dilatación temporal en los sistemas físicos obedece a una interacción entre diferentes sistemas físicos y no únicamente un fenómeno necesario por la relación lógico-matemática de nuestros postulados teóricos. Si entendemos en su dimensión práctica, el tiempo no es otra cosa que la medida del cambio, entonces lo que queda por explicar no es otra cosa que la razón por la que los sistemas ven alterados sus ritmos de cambio. Desde este punto de vista, nada nos impediría lanzar la hipótesis de que podría ser debido a la interacción del sistema con los campos gravitatorios. Sabemos que la influencia de los campos gravitatorios dilata la medición de los relojes y que cuanto más intenso es el campo gravitatorio tanto más fuerte es este efecto. Dado que los relojes lo único que hacen es servir como patrón de medición para otros cambios, no parece descabellada la idea de que los sistemas físicos estén interactuando (a nivel atómico o subatómico) con el campo gravitatorio de tal modo en el que este lastrase su cambio. Los sistemas sufrirían, por así decir, una especie de rozamiento que ralentiza su cambio o evolución. Así mismo, la TRG iguala los efectos de un campo gravitatorio con los de la aceleración, por lo que aquellos sistemas que son sometidos a aceleración también sufrirán una dilatación temporal, esto es, su ritmo de cambio se verá afectado y se ralentizará. Si tomamos la idea de Mach (1912) de que las fuerzas inerciales surgen debido a la aceleración respecto de la media de la masa del universo y la combinamos con la afirmación de Robelli (2006) de que el espacio-tiempo no es otra cosa que el campo gravitatorio de toda la masa del universo, tenemos que cuando un objeto acelera, lo hace respecto de la influencia de la combinación de todos los campos gravitatorios del universo. Por ello, la aceleración genera las mismas fuerzas que los campos gravitatorios, y es que no dejan de ser una interacción con los mismos, un cambio de velocidad relativa a estos que hace que estas fuerzas se

manifiesten de tal modo que al acelerar respecto de estos campos la influencia de éstos aumentase, generando esa especie de rozamiento entre el sistema y los campos gravitatorios que provoca que la evolución de cambio de los sistemas se vea alterada y ralentizada.

Con esto no pretendo haber dado una hipótesis que sea coherente con la física actual, sino que pretendo mostrar que tal vez una formulación de este tipo sea deseable debido a sus ventajas. Con una interpretación relacional del espacio y del tiempo (entendiendo el tiempo como medida del cambio), se podría conseguir una descripción de la realidad en la que no necesitásemos entidades intermediarias que explicasen las relaciones entre sistemas físicos, sino que podríamos hacerlo únicamente haciendo referencia a correlaciones entre sistemas observables. De este modo, ya no necesitamos ni al espacio, ni al tiempo, ni al espacio-tiempo substantivos, ya que explicamos la realidad física a través de las relaciones espaciales y el cambio que se da en los sistemas. El cambio en los sistemas, a diferencia de la dimensión temporal o el paso del tiempo, es algo observable y medible directamente en todos los sistemas conocidos; el tiempo, por el contrario, se mide a través de relojes que no dejan de ser sistemas cambiantes. De modo que, a primera vista, con un sistema teórico de este tipo tendríamos en principio dos ventajas fundamentales de las teorías relacionales: 1.) su mayor simplicidad teórica y 2.) una ventaja epistemológica; y es que las relaciones son siempre entre sistemas físicos observables y no sobre entidades teóricas como los son el espacio y el tiempo o el espacio-tiempo.

Por otro lado, vemos que las explicaciones geométricas y las explicaciones de las teorías de principios sobre la dilatación temporal y la contracción espacial nos dicen poco, más allá de los fenómenos que suceden y el grado en el que suceden. Sin embargo, el adoptar una interpretación del tiempo como medida del cambio, abre las puertas a preguntarse qué es lo que ocurre en el sistema para que este vea alterado su ritmo de cambio, pudiendo de este modo elaborar diferentes teorías que expliquen el mecanismo que actúa sobre los sistemas y provoca alteraciones en su ritmo evolutivo. Esto es, entender el tiempo como cambio nos permite la elaboración de teorías constructivas que nos permitan comprender mejor la naturaleza de los fenómenos, como la dilatación y contracción relativistas, aportando explicaciones dinámicas de la evolución atómica o subatómica de los sistemas a través de sus interacciones con otros sistemas físicos, como podría ser el conjunto de todos los campos gravitatorios.

Tampoco creo que sea posible llegar a una conclusión acerca de la corrección al adoptar una posición relacional o substancialista. Esta cuestión cae sobre el ámbito filosófico, sobre todo, ya que depende de nuestra postura epistemológica acerca de la deseabilidad de la inclusión de conceptos no observacionales. También juegan un papel muy importante en esta decisión otro tipo de valores, como la preferencia por la simplicidad ontológica y teórica, así como posiciones filosóficas como el realismo o el pragmatismo respecto de los conceptos teóricos. De modo que se podrían interpretar ciertos conceptos de una teoría, como la geometría del espacio-tiempo, como algo que forma parte de la realidad del universo, o interpretarlo simplemente como una sofisticada herramienta que nos ayuda a manejar y a acotar la realidad para poderla describir y hacer predicciones; podríamos entender el espacio-tiempo como una gran red teórica que nos ayuda a pescar la realidad, tal y como decía Popper (1962).

A pesar de lo anterior o debido a ello, considero que una teoría relacional es preferible por su simplicidad ontológica y teórica, por la reducción de entidades no observables y por su mayor poder explicativo a la hora de entender la naturaleza del espacio y del tiempo (del cambio de los sistemas) a través de comprender la complejidad de la interacción entre los sistemas. Además, si miramos atrás en la historia veremos que lo que se creía fijo, ha cambiado. Puede, que la interpretación que damos a los conceptos de espacio y tiempo, y el espacio-tiempo corran la misma suerte:

Geometry has repeatedly been mistaken for an a priori feature of reality. Euclidean geometry was erroneously thought of as necessary. Later, Riemannian geometry as well has been erroneously considered necessary. However, there is no a priori reason for which reality has to be understood as a continuum with metric properties. Nor, for that matter, as a continuum at all (Robelli, 2006: 32).

## Bibliografía

- Balashov, Y. and Janssen, M. (2003). Presentism and Relativity. *The British Journal for the Philosophy of Science*, 54(2):327–346.
- Einstein, A. (1982). *Ideas and opinions*. New York: Crown Publishers, Inc.
- Brown, H. R. and Pooley, O. (2006). Minkowski Space-Time: A Glorious Non-Entity. In: D. Dieks (ed.), *The Ontology of Spacetime* (pp. 67–89). Netherland: Elsevier.
- Mach, E. (1912). *Desarrollo Histórico Crítico de la Mecánica*. Buenos Aires: Espasa Calpe, (1959).
- Pooley, O. (2013). Substantialist and Relationalist Approaches to Spacetime. In: R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics* (pp. 522–586). New York: Oxford University Press.
- Popper, K. (1962). *La Lógica de la Investigación Científica*. Madrid: Tecnos.
- Rovelli, C. (2006). The Disappearance of Space and Time. In: D. Dieks (ed.), *The Ontology of Spacetime* (pp. 25–36). Netherland: Elsevier.

# Functional kinds as universal natural kinds in the life sciences

**Margarida Hermida**

MARE, Marine and Environmental Sciences Centre / CIIMAR-Madeira,  
Interdisciplinary Centre of Marine and Environmental Research of Madeira /  
Oceanic Observatory of Madeira (OOM)  
Email: margarida.hermida@ciimarmadeira.org

Science divides the world into kinds of things in order to study them. This can be done in a purely utilitarian manner, for purposes of convenience, or in a way which has real metaphysical import, i.e., which recognizes groups of things which really exist in the world, and not merely in the human mind. The latter are called *natural kinds*, and are held to play important roles in scientific practice, serving as the basis for classification and inference.

Philosophical accounts of natural kinds in science have had a hard time accommodating biological practice. Standard accounts of natural kinds have tended to be essentialist, defined by essences, which are unchanging sets of necessary and sufficient conditions for things to belong to a certain kind. The problem is that essentialist accounts cannot incorporate evolutionary change. As Richards (2016) points out, not even pre-Darwinian naturalists really thought of species in this way, but it is after Darwin (1859) that this view became untenable, although prominent philosophers such as Kripke (1980) continued to attempt to fit biological kinds into the essentialist shoe. The essentialist view still has contemporary supporters, e.g. Devitt (2008), but this view seems to arise from mistaking for essential properties what are merely identification criteria, namely genetic and other properties which are used by biologists to identify whether or not an organism belongs to a certain taxon which is however not defined by these properties, but by relatedness.

Since the essentialist view of biological kinds has been shown as inadequate in the face of evolution, several other hypotheses have been suggested. Ghiselin (1974) and Hull (1976) proposed that species should not be thought of as kinds at all, but as individuals, a view favoured by many biologists as well as philosophers of biology, which highlights the spatio-temporally restricted nature of species. Since then, other philosophical explanations of natural kinds were developed with the aim of incorporating the historical nature of biological taxa without abandoning the natural kinds framework altogether.

The most popular of these is homeostatic property cluster (HPC) theory, which defines kinds by a collection of similar properties which are instantiated to some degree by the members of a kind and maintained by some homeostatic mechanisms (Boyd 1990). Crucially, the kind is defined by some or all of the properties and some or all of the homeostatic mechanisms, to account for the vagueness

found in nature (Boyd 1990). There are some problems with this account, however. First of all, it can only be applied once species have already been identified, otherwise it is likely to group organisms in the same species when they belong to different species (in the case of cryptic species), and vice-versa (in the case of distinct morphotypes of the same species). In addition, while it seems to be a very good description of a species under stabilizing selection, natural selection can also be, and often is, disruptive. Neither evolutionary change over time, nor the maintenance of stable polymorphisms, are ultimately well accommodated in this framework; the crux of the matter, though, is that HPC theory comes into conflict with scientific practice in privileging similarity over relatedness, or shared properties over historical relations (Boyd 1999), where biology always follows the opposite route: morphological similarity and genetics might be used as proxies for relatedness, but when in conflict, relatedness always trumps similarity (Ereshefsky & Reydon 2015).

One alternative is to conceive of biological taxa as natural kinds that are defined by essential properties after all, but the essential properties are relational, and not intrinsic (Griffiths 1999). While this might work for species, and possibly higher taxa, phylogenetic taxa are not the only kinds of biology. For instance, functional kinds play important roles in evolutionary biology, ecology, and biomedical science. Unlike species and other taxa, these kinds are not essentially historical, nor are they necessarily spatio-temporally restricted. In fact, they are potentially universal biological kinds. Functional kinds are not adequately picked out by HPC theory because they do not map well onto clusters of properties. For instance, while there might be functional properties common to all predators, the kind is too heterogeneous to constitute a cluster of properties in the HPC sense, and the causal mechanisms producing them might also be extremely variable (Ereshefsky & Reydon 2015 attribute this example to Elliot Sober); the same can be said of the group of fast-swimming aquatic animals sharing the trait “fusiform shape” (Reydon 2009). Yet this sort of broad generalizations are the most useful for discovering universal laws in ecology (Mikkelsen 2003), as well as for investigating macro-evolutionary patterns in evolutionary biology (Powell & Mariscal 2015).

While biological classification reflects phylogenetic history, taxonomy is not the only goal of biology, nor are phylogenetic kinds the only natural kinds which play important roles in the life sciences. Functional kinds are somewhat underrated in philosophy of science due to their multiple realizability, which supposedly make them epistemically less valuable for making inductive predictions. Thus the folk biological kind “fish”, which used to include whales, is often used as an example of a functional kind based on shallow similarities, which accordingly did not allow many inferences to be drawn, because the group was very heterogeneous; whales are very different from fish and similar to other mammals, e.g. in being homeothermic, giving birth to live young, producing milk, etc; the whaleless taxon “fish” is better (more natural) because its members are more similar, and thus it supports more inferences. Likewise, the monophyletic clade “teleost” (bony fishes) allows more inferences about its members than the polyphyletic group “fish” which includes sharks, teleosts, and sarcopterygians, the ancestors of terrestrial vertebrates. Yet the virtues of phylogenetic classification cannot be reduced to its better support of inferences. In fact, in all taxa there are bound to be exceptions to most of the generalizations that can be made; e.g.

teleosts do not have the capacity to regulate their body temperature, except for tunas; teleosts breathe water through gills, except for lungfishes, and so on. Even within a species, generalizations can fail to obtain, e.g. not all blackbirds are black. Phylogenetic classification is to be preferred in that it captures the real historical pattern of the tree of life, and is thus a more natural classification.

On the other hand, it is perfectly possible to draw inferences from functional kinds, as long as they are rigorously defined, and inferences are restricted to the particular function under analysis. For example, functional kinds are important in evolutionary biology, which not only studies diversification but also cases of convergent evolution, for example of locomotion devices (e.g. legs, wings), sense organs (e.g. eyes, mechanoreceptors), or reproductive abilities (e.g. viviparity) in widely separated animal lineages. Functional kinds are also important in ecology, which needs to be able to compare fundamentally different organisms which nevertheless share similar traits which allow them to fulfil comparable roles in ecosystems (Mikkelsen 2003). Measures of functional diversity are increasingly used in ecological studies alongside traditional taxonomic diversity measures, in recognition of the importance of the diversity of roles played by organisms for ecosystem functioning, regardless of the particular taxonomic position of the organisms (Cadotte et al. 2011).

Contrary to what is often assumed, it is not true that the biological sciences lack universal natural kinds, although it is quite correct that species are not it. While species and other phylogenetic kinds are indeed spatio-temporally restricted, historical kinds (or, alternatively, not kinds at all but individuals), there are also universal natural kinds in biology. Just like a physical particle, such as an electron, can emerge anywhere in the universe, so can a photosynthetic organism, or a predator. That they have not usually been recognized as universal is due to the circumstance that we have not yet managed to detect a single example of independently evolved extra-terrestrial life. Yet the functional kinds that are already used in ecology and evolutionary biology are universal in the sense that, unlike species and higher taxa, they would apply to any organisms and ecosystems, anywhere in the universe, facing similar physical, chemical or biotic constraints. These include functional kinds used in the detection of convergent evolution of widely separated lineages on Earth, which would equally apply to the detection of convergent evolution in truly independent (extra-terrestrial) lineages, and ecological kinds such as trophic mode, feeding guilds, or life history strategies, as well as functional kinds which effectively overlap the two—for instance, in animals with eyes, predators are more likely to have both eyes facing forward, whereas prey are more likely to have laterally located eyes (which provide them with a broader visual field, at the expense of stereoscopic vision).

None of this is to say that, if alien life were to be discovered, classification should not follow phylogeny in this other tree of life; quite the contrary, that would be the most adequate classification. However, much would be learned from comparison to Earth organisms and ecosystems, and these comparisons would have to be made in terms of functional kinds, which are universal biological kinds.

## Acknowledgements

M.H. was financially supported by a post-doctoral grant from ARDITI (Regional Agency for Development of Research, Technology and Innovation of Madeira), Project M1420-09-5369-FSE-000001. This study was partially supported by the Oceanic Observatory of Madeira Project (M1420-01-0145-FEDER-000001-Observatório Oceânico da Madeira-OOM).

## References

- Boyd, R. (1990). Realism, approximate truth, and philosophical method. In C. W. Savage, ed. *Scientific Theories*, pp. 355–391, University of Minnesota Press.
- Boyd, R. (1999). Kinds, complexity and multiple realization. *Philosophical Studies*, 95(1-2):67–98.
- Cadotte, M. W., Carscadden, K., & Mirotchnick, N. (2011). Beyond species: functional diversity and the maintenance of ecological processes and services. *Journal of Applied Ecology*, 48(5):1079–1087.
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. Murray.
- Devitt, M. (2008). Resurrecting biological essentialism. *Philosophy of Science*, 75(3):344–382.
- Ereshefsky, M., & Reydon, T. A. (2015). Scientific kinds. *Philosophical Studies*, 172(4):969–986.
- Ghiselin, M. T. (1974). A radical solution to the species problem. *Systematic Biology*, 23(4):536–544.
- Griffiths, P. E. (1999). Squaring the circle: natural kinds with historical essences. In R. A. Wilson, ed. *Species: New Interdisciplinary Essays*, pp. 209–228. MIT Press.
- Hull, D. L. (1976). Are species really individuals? *Systematic Biology*, 25(2):174–191.
- Kripke, S. (1980). *Naming and Necessity*. Wiley-Blackwell.
- Mikkelsen, G. M. (2003). Ecological kinds and ecological laws. *Philosophy of Science*, 70(5):1390–1400.
- Powell, R., & Mariscal, C. (2015). Convergent evolution as natural experiment: the tape of life reconsidered. *Interface Focus*, 5(6):20150040.
- Reydon, T. A. (2009). Do the life sciences need natural kinds? *Croatian Journal of Philosophy*, 9(2):167–190.
- Richards, R. A. (2016). *Biological classification: a philosophical introduction*. Cambridge University Press.

# Más allá de Kitcher: la reformulación del contrato social de la ciencia

Abraham Hernández<sup>1</sup>  
Universidad de La Laguna<sup>2</sup>

Philip Kitcher ha argumentado recientemente que la idea de un “contrato social para la ciencia” se sustenta en una imagen de la ciencia basada en el mito de la pureza. En esta imagen, las investigaciones pueden ser evaluadas al margen de sus implicaciones o consecuencias. En nuestro trabajo, profundizaremos en las críticas de Kitcher y añadiremos algunos nuevos elementos.

La enorme confianza que la sociedad tiene depositada en la ciencia no es en modo alguno gratuita. En gran medida ha sido proporcionada por el éxito de sus investigaciones. Sin embargo, desde el siglo XX existen una serie de ejemplos que permiten cuestionar esta imagen benévola y optimista de la ciencia. Esto es, precisamente, lo que plantean autores como Philip Kitcher, por ejemplo, en *Truth, Science and Democracy* (2011). El Proyecto Manhattan, una investigación científica impopular, es utilizado por Kitcher como caso paradigmático para explicar cómo ciertas investigaciones están respaldadas por una imagen excesivamente optimista de la ciencia que él denomina “mito de la pureza”.

El análisis del Proyecto Manhattan muestra lo necesario que es reflexionar acerca de los límites y objetivos de la investigación científica. Para Kitcher el problema reside, justamente, en el esquema valorativo a través del cual se juzga la investigación científica. La imagen que se tenga de la ciencia no solo determina la agenda de la investigación científica, sino que determina también el esquema valorativo a través del cual analizar su producción.

El “mito de la pureza” consistiría en hacer una interpretación dicotómica entre lo epistémico y lo práctico (o su equivalente, entre la ciencia básica y la ciencia aplicada). La pureza quedaría del lado del conocimiento científico puro. Por el contrario, el desarrollo de las aplicaciones tecnológicas cargaría con todas las responsabilidades morales, sociales y políticas. A menudo, se justifican investigaciones impopulares diciendo que el objetivo de la ciencia es ante todo la búsqueda de la verdad en la naturaleza sin relación alguna con el contexto humano. Si el conocimiento es en sí mismo valioso, es decir, significativamente epistémico, puede ser evaluado al margen de sus implicaciones y consecuencias. Lo cual quiere decir que el desarrollo de ciertas investigaciones puede entrar en contradicción con los intereses de la sociedad.

---

<sup>1</sup>Trabajo cofinanciado por la Agencia Canaria de Investigación, Innovación y Sociedad de la Información de la Consejería de Economía, Industria, Comercio y Conocimiento y por el Fondo Social Europeo (FSE) Programa Operativo Integrado de Canarias 2014-2020, Eje 3 Tema Prioritario 74 (85 %)

<sup>2</sup>Este trabajo ha sido realizado en el marco del proyecto *Puntos de vista, disposiciones y tiempo. Perspectivas en un mundo de disposiciones* (FFI2014-57409-R), financiado por el Ministerio de Economía y Competitividad del Gobierno de España).



El Proyecto Manhattan generó una alianza entre ciencia y política sin precedentes. Inmediatamente después a la Segunda Guerra Mundial apareció la figura de un “contrato social para la ciencia”. Dicho contrato se basaría en un pacto entre la comunidad científica y la política que tenía como fin la puesta en práctica de políticas científicas que fomentaran el desarrollo social y económico de la sociedad (en su origen específico, de la sociedad norteamericana). Una de las cuestiones que hicieron que dicho contrato tuviera una acogida exitosa fue sin duda esa supuesta pretensión baconiana de que el fin último de la ciencia era promover el bienestar social.

La idea principal fue plasmada en el informe “Ciencia, la frontera sin fin” que Vannevar Bush elaboró, en 1945, a petición del presidente Roosevelt. Se trataba de que la ciencia y la política siguieran trabajando juntas, como así ocurrió en la Segunda Guerra Mundial, pero ahora en un contexto de paz, donde la producción científica no fuera exclusivamente militar y pudiera revertir en la sociedad.

Más concretamente, el contenido del informe giraba en torno a cuatro puntos: 1) la seguridad militar, 2) la lucha contra la enfermedad (sanidad), 3) apoyo a programas de investigación científica y 4) programa para fomentar el talento de jóvenes científicos norteamericanos. A pesar de que en la carta que Roosevelt envía a La Oficina de Investigación y Desarrollo Científico se puede observar un claro interés por los beneficios prácticos de la ciencia, el modelo que Vannevar Bush le remite estaba basado en la “ciencia básica”. Este tipo de ciencia no solo era, según él, la que producía la mayoría de los nuevos conocimientos, sino la que generaría la “ciencia aplicada”, procurando así el verdadero desarrollo. El estado debía, pues, mejorar y apoyar económicamente a las universidades y grupos de investigación porque era allí donde se desarrollaba ese tipo de ciencia. Se conseguía así contrarrestar el interés estatal por la generación inmediata de conocimientos prácticos de cara a fines estratégicos mediante la incorporación de un principio que salvaguardara la autonomía de la investigación.

En esta comunicación, utilizaremos la noción de “programa científico” para continuar desarrollando la línea crítica de Kitcher a la hora de mostrar los inconvenientes de seguir manteniendo esa imagen de la ciencia basada en el mito de la pureza. Y añadiremos algunos elementos más que deben ser tenidos en cuenta. Fenómenos actuales como 1) la globalización en todas las esferas, 2) la producción científica en el seno de grandes corporaciones con capital exclusivamente privado (esto es, sin financiación pública) y 3) el fenómeno reciente de contestación reivindicativa por parte de las comunidades científicas, respecto a las ingerencias de la política y a una excesiva burocratización de su trabajo, sugieren nuevas complejidades que deben ser abordadas a la hora de plantear la revisión de un “contrato social para la ciencia”.

En conclusión, el “contrato social para la ciencia”, pese a haber sido desarrollado con la intención de maximizar el desarrollo social y económico, se sustentaba en una imagen de la ciencia basada en el mito de la pureza, donde todo gira en torno a la ciencia básica. Esta forma de concebir la ciencia, donde las investigaciones pueden ser evaluadas al margen de sus implicaciones o consecuencias, no puede ser adecuada de cara a conseguir ese supuesto objetivo último de la ciencia de mejorar las condiciones de vida.

## Referencias bibliográficas

- Brooks, H. (1968). *The government of science*. The MIT Press.
- Fuller, S. (2010). *The philosophy of science and technology studies*. New York: Routledge.
- Longino, H. (2002). “Science and the Common Good: Thoughts on Philip Kitcher’s Science, Truth, and Democracy”. *Philosophy of Science*, 69(4):560–568.
- Kitcher, Ph. (2001). *Science, truth and Democracy*. Oxford University Press.
- Kitcher, Ph. (2010). “Science in a Democratic Society”, en González, W.J (ed), *Scientific Realism and Democratic Societies: The philosophy of Philip Kitcher*, Poznan Studies in the philosophy of Science and Humanities, Rodopi, Amsterdam, pp. 95–112

# The enhanced indispensability argument for Platonism and the circularity problem

Jan Heylen  
Lars Tump

Mathematical platonists affirm the existence of mathematical entities, whereas mathematical nominalists deny it. A major argument for mathematical platonism is the so-called ‘Quine-Putnam indispensability argument’ (Colyvan, 2001, p. 11):

**P1** Mathematical entities are indispensable to our best scientific theories.

**P2** We ought to be ontologically committed to all and only those entities that are indispensable to our best scientific theories.

**C**  $\therefore$  We ought to be ontologically committed to mathematical entities.

This is the backdrop against which Baker (2005) advanced the so-called ‘Enhanced Indispensability Argument’. We only ought to be ontologically committed to those entities that are indispensable for *explanations*. Common ground in this debate is that if there are mathematical entities that fulfill the same kind of explanatory role as theoretical entities in scientific theories, then mathematical entities are on par with theoretical entities and mathematical platonism prevails. The challenge is to find an example of a mathematical explanation of a physical phenomenon. Baker (2005) purports to provide such an example. Furthermore, he uses it to further boost the cause of platonism (Baker, 2017).

The example treats of the Periodical Cicada, a genus of the Cicada family, of which the species have either a 13- or a 17-year life cycle (D). To prevent too much repetition, only the first datum will be used. One of the questions biologists are concerned with, is why the life cycles of these species are prime (the explanandum). For instance, whether it might be the case that this would give them some evolutionary advantage because prime periods minimize intersection. Explanation (E) thus states that life cycles are likely to be prime because they minimize intersection. Extended explanation (E\*) is then (E) with added biological constraints, and yields an explanation of (D) (Baker, 2005, p. 233).

However, given the assumption that the explanandum is true because of the use of an inference to the best explanation, and the presence of the mathematical property of primeness in the explanandum, it would be the case that one is ontologically committed to mathematical entities, namely prime numbers. But then it would be the case that mathematical platonism is established beforehand, and as such the argument would indeed be circular (Bangu, 2008).

So, a modification is needed. The reference to numbers is paraphrased away in first-order logic with identity, for reasons of circularity:

$$(D_{fol}) \quad \begin{aligned} &\exists x_1 \dots \exists x_{13} (Fx_1 \wedge \dots \wedge Fx_{13} \wedge x_1 \neq x_2 \wedge \dots \wedge x_{12} \neq x_{13} \wedge \\ &\quad \forall x_{14} (Fx_{14} \leftrightarrow (x_{14} = x_1 \vee \dots \vee x_{14} = x_{13}))) \end{aligned}$$

This paper questions the feasibility of (E\*) yielding an explanation of (D) within the context of the Enhanced Indispensability Argument. Note that, on the one hand, (D<sub>fol</sub>), treats of a *plurality*, that is, the collection of objects expressed by the first-order logic sentence. But (E), on the other hand, is about an *entity*, which has the property of being prime. This means that there is a gap between (D<sub>fol</sub>) and (E\*): how can the plurality of (D<sub>fol</sub>) end up as entity ('13') about which (E\*) claims that it is prime?

We need some sort of bridge principle, so as to be able to logically relate the explanans to the explanandum by connecting pluralities to entities. Without a logical connection, there is no explanation of the phenomenon. In light of Frege's work (Frege, 1893) and the neo-Fregean research program (Burgess, 2005), a natural suggestion would be to use Hume's Principle (HP), which states that "for any concepts F, G, the number of F is identical to the number of G if and only if F and G are equinumerous" (Shapiro, 2000, p. 110). The principle HP connects the plurality of objects that have properties F and G to entities (numbers). If one combines HP with dyadic second-order logic, then Frege's and Dedekind's Theorems say that the Peano postulates can be deduced, given Frege's definitions of zero, successor and natural number (Burgess, 2005, pp. 26, 29). Once we have that, we can start proving results about prime numbers.

Frege and the neo-Fregeans think that HP is analytic, whereas others think that it is not, because it is not ontologically neutral (Boolos, 1998). In response to that charge a conditional version of HP has been proposed: *if* numbers exist, then HP is true (Field, 1984, p. 661). This leads to a dilemma the Enhanced Indispensability Argument faces when it is intended as an argument for mathematical platonism: the predicate '\_\_\_ is prime' can either not latch on to the first-order logic sentence without the latter being interpreted as a mathematical statement, resulting from the supposed analyticity of HP, or the predicate can not latch on to the first-order logic sentence with identity without making it conditional upon the existence of numbers, as is the case with a conditional version of HP. In either case no ontological conclusion can be drawn from it. In the first case, it would make the argument again question begging against the mathematical nominalist. And in the second case, it would make the argument ontologically neutral between the platonist and nominalist. The defenders of the Enhanced Indispensability Argument might, of course, try to find an alternative way of bridging the gap without falling into the trap of the foregoing dilemma, but it is up to them to provide such a way.

## Referencias bibliográficas

Baker, A. (2005). Are there genuine mathematical explanations of physical phenomena? *Mind* 114(454), 223–238.

- Baker, A. (2017). Mathematical spandrels. *Australasian Journal of Philosophy* 95(4), 779–793.
- Bangu, S. I. (2008). Inference to the best explanation and mathematical realism. *Synthese* 160(1), 13–20.
- Boolos, G. (1998). *Logic, Logic, and Logic*. Harvard University Press.
- Burgess, J. P. (2005). *Fixing Frege*. Princeton University Press.
- Colyvan, M. (2001). *The Indispensability of Mathematics*. Oxford University Press.
- Field, H. (1984). Frege's Conception of Numbers as Objects by Crispin Wright. *Canadian Journal of Philosophy* 14, 637–662.
- Frege, G. (1893). *Grundgesetze der Arithmetik, Begriffsschriftlich abgeleitet*, Volume 1. Jena: Verlag Hermann Pohle.
- Shapiro, S. (2000). *Thinking about mathematics: the philosophy of mathematics*. Oxford University Press.

# How our mind enables and constrains the theories we develop

**Alexander Krauss**

London School of Economics

The limits of our mind shape the limits of knowledge and the sciences. A common assumption among scientists is that our mind has evolved well for scientific reasoning and theorising. I argue that this common assumption oversimplifies the cognitive and sensory constraints we face in formulating highly complex and precise scientific theories about the world, and especially about phenomena we cannot directly perceive. I show this combining evidence and perspectives from the evolution of our cognition and empirical research on adults, children and other animals. The paper thereby addresses the broad question of how our cognition and senses enable and also constrain the particular theories we can develop. Implications are then discussed on how scientists need to label different types of knowledge accordingly in order not to oversell their reliability.

# Patterns of theory change in physics

Ladislav Kvasz

Theory change was not always part of our image of science. Philosophers of enlightenment and positivism viewed science as a continuously growing field that liberates our thought from prejudice, religious superstition, and metaphysical speculation. Theory change was for the first time seriously discussed by a generation of outstanding philosophers-scientists: Ernst Mach, Henri Poincaré, and Pierre Duhem. Mach in his famous *Die Mechanik in ihrer Entwicklung* (Mach 1893) showed that scientists sometimes use metaphysical concepts that have no empirical content. Newton's concept of absolute time is, according to Mach, such a concept:

This absolute time can be measured by comparison with no motion; it has therefore neither a practical nor a scientific value; and so no one is justified in saying that he knows aught about it. It is an *idle metaphysical conception*. (Mach 1893, p. 273, my emphasis)

This conceptual criticism of Newton's mechanics is one of the most outstanding examples of philosophical criticism, and was appreciated even by Einstein. So scientific theories are also subject to change, but this change consists in the elimination of "idle, metaphysical conceptions". It seems that the logical positivists largely accepted Mach's diagnosis, and so even though they witnessed the most radical cases of theory change in physics—the creation of general relativity and quantum mechanics—they were mostly interested in devising methods for identifying and replacing the empirically unverified parts of theories, and thus securing the growth of science.

Thomas S. Kuhn has radically challenged this picture. In *The Structure of Scientific Revolutions* (Kuhn 1962) he advanced a view of science in which theory change plays a central role. His theory was at the center of heated discussions, and there are several good expositions of his views (e.g. in Horwich ed. 1993). We can distinguish three stages in the development of Kuhn's views (see Gonzalez 2011). The *initial period* (roughly until 1968) was built around the duality of normal science and revolutionary science, and his thesis of the incommensurability of paradigms. The period of *revision, clarification and enlargement* of that initial thesis was marked by the publication of the *Reflections on My Critics* (Kuhn 1970) and *Second Thoughts on Paradigms* (Kuhn 1974; first presented in 1969). During that second period Kuhn introduced his notion of a disciplinary matrix and thereby softened his initial radical thesis of incommensurability. The third and final period was marked by Kuhn's *turn towards language*. The main change here was a new emphasis on taxonomic lexicons. While Kuhn still maintained his incommensurability thesis, which he now closely tied to the impossibility of translation between taxonomic lexicons, he admitted that scientists can become bilingual. Kuhn's views at different periods opened the door for developments in different directions.

The fact that during his first period Kuhn focused on the scientific community and its changes of belief, and that during his second period he discussed objectivity and value judgment, opened the door for closer contact with Bayesianism and its theory of belief revision. Indeed, two prominent Bayesians published papers exploring this connection. Wesley C. Salmon in *Rationality and Objectivity in Science, or Tom Kuhn Meets Tom Bayes* (1990) and John Earman in *Normal Science, Scientific Revolutions, and All That: Thomas Bayes versus Thomas Kuhn* (1992) discuss how Bayesian belief revision theory could help in developing some aspects of theory change.

On the other hand, as Gonzalez (2011) shows, Kuhn's stress on language and taxonomic lexicons during the last period of his thought allowed his theory of scientific revolutions to be developed toward an analysis of conceptual change, as done by Paul Thagard in his *Conceptual Revolutions* (Thagard 1992). Thagard challenged the Bayesian account of theory change on the grounds that it requires a stable conceptual framework. He questioned how far conceptual change can be modeled by methods of belief revision, and instead of Bayesian theory he based his reconstruction on methods of artificial intelligence, which opened entirely new perspectives.

So we have at least three frameworks for the analysis of theory change related to the work of Kuhn: the historiographical, the Bayesian, and the connectionist. It is interesting to notice, with Woosuk Park, that these frameworks do not interact: "What is curious is that Thagard's interesting question has failed to attract any serious response from belief revision theorists" (Park 2010, p. 121). Most of the approaches to theory change are subject to a criticism formulated by John Earman:

The philosophy of science is littered with methodologies of science, the best known of which are associated with the names of Popper, Kuhn, Lakatos, and Laudan. ... I have two common complaints. The first stems from the fact that each of these methodologies seizes upon one or another feature of scientific activity and tries to promote it as the centerpiece of an account of what is distinctive about the scientific enterprise. The result in each case is a picture that accurately mirrors some important facets of science but only at the expense of an overall distortion. The second common complaint is that these philosophers, as well as many of their critics, are engaged in a snark hunt in trying to find The Methodology of Science. (Earman 1992, p. 203)

Maybe Earman would disagree, but I am inclined to level his first complaint also at the Bayesian belief revisionist approach and Thagard's theory of conceptual change. The aim of the present paper is to offer a picture of theory change that would address Earman's complaints. This picture is based on the observation that changes occur in science at different levels. If we separate these levels, study them in isolation, develop for each of them the necessary methodological tools for their analysis, and then devise a theory that incorporates them all, we can reduce the degree of distortion mentioned by Earman. The conflicting accounts of science, to which he calls our attention, can be seen to result from the fact that authors such as Kuhn, Lakatos, Popper, and Laudan (as well as Salmon and Thagard) each focus on one of these levels while distorting the others.



But let us start with *The Structure of Scientific Revolutions* (Kuhn 1962), which was undoubtedly one of the most influential interpretations of the evolution of science in the twentieth century. Nevertheless, from the point of view developed in this paper Kuhn's theory can be likened to a composite image made by superimposing photographs of three different faces. Each of the original three photographs is sharp and rich in specific detail. By superimposing them, however, the details are lost, and what remains is the gross structure of the face: the overall contours, dark spots instead of eyes, and a blot instead of the mouth. Similarly, when Kuhn superimposed the "photographs" of the three different kinds of scientific revolutions, he lost the details of the *cognitive dynamics* specific to each kind. What remained were only the gross features common to all three types of revolution: the *social dynamics* of the scientific community's response to change.

The notions of *anomaly*, *crisis*, and *revolution* are not specific to science. They can equally be used to describe a group's wandering in the woods. From some observations (*anomalies*) the group concludes that it went astray; for some time, several suggestions of which way to take are discussed, without any consensus (*crisis*); finally one of the suggestions is accepted (*revolution*); and the group walks in the accepted way (*new period of normal wandering*). The social dynamics described by Kuhn are not a characteristic of the development of science, but only a consequence of mixing different kinds of revolutions. Mixing "cancels" the *cognitive dynamics*, which is different in every kind, and "strengthens" the *social dynamics*, which is always the same.

We can discriminate four kinds of theory change: three of them revolutionary and one close to normal science. I will study the linguistic changes that accompany them. Thus in a sense I will follow Kuhn's view that "violation or distortion of a previously unproblematic scientific language is the touchstone for revolutionary change" (Kuhn 2000, p. 14). Below I introduce each kind of theory change on the basis of a few examples.

## First kind of theory change: changing the aim of scientific research

A total change of the aim of scientific research is a change of the greatest possible scope. Perhaps the typical example of this kind of change is *The Scientific Revolution*, which separated ancient science that tried to reveal an *unchanging order in nature*, from modern physics that looks for *dynamic laws beyond the phenomena*. This change can be illustrated by comparing Kepler with Newton (for details see Kvasz 2012).

Kepler was one of the last eminent scientists to follow the ideal of science from classical antiquity: he searched nature for eternal, unchanging forms. His law of the elliptical shape of the planetary orbits is a typical law of this kind. From a Newtonian point of view, we can say that Kepler was really lucky. The tables left behind by Tycho Brahe were *sufficiently precise* to discriminate the elliptical shape of Mars's orbit from a circular shape, but they were *sufficiently inaccurate* not to reflect the perturbations of Mars's orbit caused by the other planets.

Only because of that could Kepler assert (in agreement with the classical ideal of science) that Mars's orbit *has the form* of an ellipse. In fact, this orbit, just like any other orbit in the Solar system, has no pre-given form, which could be described by means of geometry. The motion of each planet is lawful, not in the sense that its trajectory would have a particular geometric shape (as ancient science believed), but rather in the sense that it is governed by a law describing the generation of its trajectory by forces. The trajectories of the planets are not lawful in the sense that they would *reveal some eternal and unchangeable geometric shape* that should be discovered (as Kepler and, before him, all ancient scholars understood the aim of science), but rather in the sense that it is possible to find a *dynamic law describing their generation* (as Newton and, after him, the majority of scientists see the aim of science). A further step on this level leads us to Darwinian science as a new kind of research aiming to discover evolutionary *lineages*, genetic *variation*, *fitness* and *selection*.

*The Scientific Revolution* led to the introduction of a general linguistic framework common to every sufficiently general physical theory. It consists of: (1) determination of the *measurable quantities*; (2) description of the *state of a physical system*; and (3) choice of the *differential equation* describing the *temporal evolution of the state*. Theories like Newton's mechanics, Maxwell's electrodynamics, Schrödinger's quantum mechanics, and many others share this common linguistic structure. Thus when, in the next subsection, we turn to examples of theory change of the second kind, we will be following developments that occurred in this common framework.

## Second kind of theory change: changing the instrumental practice

The second kind of change is of lesser scope than entirely changing the aims of one's research. An example of this second kind of change in physics would be the introduction of field theory, or of quantum mechanics. The birth of field theory was due to Volta's discovery of the battery, which gave physicists a sufficiently stable source of electric current to experiment with. Shortly an entirely new instrumental practice emerged, based on the use of batteries, conductors, magnets, and all the rest of the equipment used by physicists such as Hans Christian Oersted, Georg Simon Ohm, and Michael Faraday. The work of these and several other scientists finally led Maxwell to discover field theory. Nevertheless, we have to remember that this discovery occurred in the common linguistic framework described above. On the one hand, this common framework prescribes that every physical theory must have *a* list of measurable quantities, *a* description of the system's state, and *a* differential equation describing the temporal evolution of the state. On the other hand, it doesn't prescribe *which particular* measurable quantities, *which particular* state description, or *which particular* equation to choose.

Theory change of this second kind consists in a *change* of the set of measurable quantities (corresponding to the introduction of new instrumental techniques, that make it possible to measure quantities that were hitherto immeasurable);

a *change* of the description of a physical system's state (incorporating into the description new aspects of reality that were hitherto not included); and a *change* of the fundamental equations of the theory.

*Newtonian mechanics* has as measurable quantities *time*, *position*, *velocity*, *acceleration*, and *mass*; while the state is given by two vectors—the *vector of position* and the *vector of momentum*—for each particle; and the dynamical equation is *Newton's second law*. In this respect Newtonian mechanics differs from *field theory* that introduces further measurable quantities, such as *electric charge*, *electric current*, and *electric field*; while the state of a system is given by two vector fields—the *electric field* and the *magnetic field*; and the dynamical equations are *Maxwell's equations*.

### Third kind of theory change: changing the conceptual framework

Changes of the entire conceptual framework are changes of lesser scope than changes of the instrumental practice. They take place within the framework constituted by an instrumental practice. As examples of this third kind of change in classical mechanics (i.e. in the Newtonian framework described above) we can take *Newton's mechanics*, *Lagrange's mechanics*, and *Hamilton's mechanics*. Despite their similarity—they share a common instrumental practice, so they describe the same set of empirical data—there are interesting differences between these systems. To illustrate these differences, consider the description of the system of the Earth and the Moon.

*Newton's mechanics* (1687) describes the Earth-Moon system as the motion of two bodies in a three-dimensional space, by six second-order differential equations. *Lagrange's mechanics* (1788) describes the Earth-Moon system as the motion of a single body in a six-dimensional configuration space (whose first three coordinates determine the position of the Earth, and the remaining three coordinates the position of the Moon), and the differential equations remain of the second order. *Hamilton's mechanics* (1835) describes the Earth-Moon system as the motion of a single body in a twelve-dimensional phase space (whose first three coordinates determine the position of the Earth, another three coordinates the position of the Moon, another three coordinates the momentum of the Earth, and the last three the momentum of the Moon), and the differential equations are twelve first-order equations. We should recognize not only the unity of these three theories (they deal with the same system of measurable quantities, and the same empirical data), but also their differences, which consist of the different conceptual frameworks (Euclidean space, configuration space, phase space; or force, kinetic potential, energy) by means of which they describe the same empirical data.

## Fourth kind of theory change: discovering new facts

Discoveries of new facts are changes of the smallest scope. They form the content of *normal science*, when scientists find what the present paradigm foresaw. Nevertheless, I would like to include here also *unexpected discoveries of limited extent*, like the discovery of the planet Uranus by William Herschel in 1781, which do not change the conceptual framework in any substantial way. Such changes are discontinuities; in our example, a new technical term, ‘Uranus’, was introduced and its presence distinguishes the papers and books published before and after the discovery.

## Acknowledgments

The author expresses his deep gratitude for the generous support of the project Formal Epistemology – the Future Synthesis in the framework of the program *Praemium Academicum* of the Czech Academy of Sciences. I would like to thank my colleagues on the project, Joan Bertran-San Millán, Aldo Filomeno, and Vera Matarese for comments and criticism of previous versions of this paper.

## References

- Earman, J. (1992): Normal Science, Scientific Revolutions, and All That: Thomas Bayes versus Thomas Kuhn. In Earman, J.: *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. The MIT Press, Cambridge Massachusetts, pp. 187–205.
- Gonzalez, W. J. (2011): The Problem of Conceptual Revolutions at the Present Stage. In: Gonzalez, W. J. (ed.): *Conceptual Revolutions: From Cognitive Science to Medicine*. Netbiblo, La Coruña, pp. 7–38.
- Hamilton, W. R. (1835): On a General Method in Dynamics; by which the Study of the Motions of all Free Systems of Attracting and Repelling Points is Reduced to the Search and Differentiation of one Central Relation or Characteristic Function. In: *Phil. Trans. Roy. Soc. London*.
- Horwich, P. (ed. 1993): *World Changes. Thomas Kuhn and the Nature of Science*. Cambridge: The MIT Press.
- Kuhn, T. S. (1962): *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.
- Kuhn, T. S. (1970): Reflections on My Critics. In: Lakatos, I. and Musgrave A. (eds): *Criticism and the Growth of Knowledge*, Cambridge University Press, Cambridge, 1970, pp. 231–278.
- Kuhn, T. S. (1974): Second Thoughts on Paradigms. In: Suppe, F. (ed.): *The Structure of Scientific Theories*. University of Illinois Press, Urbana, pp. 459–482.

- Kuhn, T. S. (2000): *The Road Since Structure*. Chicago: The University of Chicago Press.
- Kvasz, L. (2012): Galileo, Descartes, and Newton – Founders of the Language of Physics. *Acta Physica Slovaca*, 62:519–614.
- Lagrange, J. L. (1788): *Mécanique Analytique*. Paris. English translation by A. Boissonnade and V. N. Vagliente: *Analytical Mechanics*. Boston Studies in the Philosophy of Science, Kluwer 1998.
- Mach, E. (1893): *Die Mechanik in ihrer Entwicklung*. Brockhaus, Leipzig. English translation: *The Science of Mechanics*. The Open Court, Chicago 1902.
- Newton, I. (1687): *The Principia*. A New Translation by I. Bernard Cohen and Anne Whitman, Proceeeded by *A Guide to Newton's Principia*, University of California Press, Berkeley 1999.
- Park, W. (2010): Belief Revision vs. Conceptual Change in Mathematics. In: Magnani, L. et al. (eds.): *Model-Based Reasoning in Science and Technology*. Springer, Berlin, pp. 121–134.
- Salmon, W. C. (1990): Rationality and Objectivity in Science or Tom Kuhn Meets Tom Bayes. In: Savage, C. W. (ed.): *Scientific Theories*. University of Minnesota Press, Minneapolis. pp. 175–204.
- Thagard, P. (1992): *Conceptual revolutions*. Princeton: Princeton University Press.

# Can functionalism save emotions from elimination?

**Juan R. Loaiza**

Berlin School of Mind and Brain  
Humboldt-Universität zu Berlin

In the past decades there has been a growing skepticism about the natural kind status of emotions and, consequently, about their place as objects of scientific study in psychology and neuroscience. This skepticism takes two forms. One is eliminativism, prominently defended by Griffiths (1997). According to Griffiths, even though emotion concepts are natural kind concepts, they do not refer to a single natural kind. Instead, different theories of emotions capture a subset of the whole category, without any given theory explaining all emotions under one common framework. As a result, Griffiths proposes eliminating emotions as scientific concepts.

The second form of skepticism lies at the heart of psychological constructionism (Barrett, 2006; 2014; 2018; Russell, 2003; 2009). According to psychological constructionists, emotions are a heterogeneous set of processes at the behavioral, physiological, and neural levels. In other words, there are not specific nor consistent patterns corresponding one-to-one to emotion categories. Consequently, emotions do not form natural kinds, and thus their status as scientific concepts must, at the very least, be reevaluated.

In either of these forms, a common motivation behind skepticism is that empirical findings have not been able to establish physiological or neural correlates that are either unique to emotions as a whole, or specific or consistent to individual emotions in particular. At the neural and physiological levels, all we see is a myriad of processes that are also present for a number of cognitive and affective phenomena besides emotions, and that do not consistently map onto particular emotion categories. This leads Griffiths to claim that emotions should be eliminated from the vocabulary of science, and constructionists to think of emotions as arbitrarily constructed categories with which we describe some of our core affective states. In either case, emotions lose their scientific and explanatory value as a distinct category.

In my talk, I present arguments to accept the premise but resist the conclusion. I show why emotion research must deal with the fact that emotions do not form kinds at the neural or physiological levels, but I will claim this should not lead to skepticism. I argue that we can offer a functionalist account of emotions that cashes out neural and bodily variability as the multiple realization of a group of functional kinds. By presenting emotions as functional kinds, I show how we can account for their epistemic and methodological value as proper objects of study while remaining consistent with empirical evidence.

To argue for this claim, first, I contend that part of the problem leading to skepticism lies in the notion of *natural kind*. It is the assumption that the kinds science is concerned with must be defined in a uniform way across disciplines that leads to the claim that all scientific objects must conform to a single pattern of inference. Be it inferences in virtue of a common essence (Kripke, 1972; Putnam, 1975) or a homeostatic mechanism (Boyd, 1991; 1999), traditional accounts of natural kinds have overstated specific patterns that support inferences in different scientific disciplines. If we accept that there is a variety of inferential patterns across disciplines, and hence a variety of ways in which scientific theories refer to kinds, then the question is not whether emotions conform or not to essentialism or the HPC account, as skeptics have argued. Instead, the question becomes whether scientists do inferences using emotion categories that are fruitful for their discipline and how do these patterns work.

In my view, there are at least four patterns with which scientific theories make reference to kinds. In other words, there are at least four ways in which scientists project properties from one member of a kind to other unobserved members. One is the one referred to by essentialist accounts, i.e. inferences in virtue of a common essential property. Presumably, inferences across chemical elements and physical entities follow this pattern. I call such kinds *essentialist* kinds. A second pattern of inference are inferences due to a common history. These *historical* kinds of this sort include, for example, biological species. Third, there are inferences in virtue of functional profiles, which not only include entities in psychology and neuroscience, but also certain physical entities (e.g. machines). Kinds of this type are *functional* kinds. Lastly, some inferences are supported due to common social practices and facts, hence constituting social kinds. Examples of this type are kinds such as currency, marriage, and other social entities.

With this four-way taxonomy in mind, I argue that emotions conform to the third pattern of inference, that is, that they are functional kinds. To support this claim, I begin by presenting recent empirical evidence showing the lack of unique, specific, and consistent neural and bodily markers of emotions, i.e. the Problem of Variability. In the last two decades, a number of meta-analyses and reviews (Hamann, 2012; Kragel & LaBar, 2016; Kreibig, 2010; Lench, Flores, & Bench, 2011; Lindquist et al., 2012; Murphy, Nimmo-Smith, & Lawrence, 2003; Phan et al., 2002; Siegel et al., 2018; Vytal & Hamann, 2010) have shown conflicting evidence about the nature of emotion kinds. On one hand, a number of studies suggest that emotions are too heterogeneous to form kinds at all. These are the sort of studies that lead psychological constructionists to skepticism. Studies in this direction report different brain regions activating for a single emotion category, as well as significant overlap between the underpinnings of different emotions and lack of homogeneity in physiological responses. Yet, on the other hand, there is also evidence implying that there are identifiable patterns of brain and autonomic activity at other levels. More recently, these studies have proposed the use of multivariate pattern analyses to unveil specific and consistent networks that may correspond to emotion categories. Nevertheless, evidence is still inconclusive.

I submit that there are two reasons for such inconclusiveness. One is the lack of more sophisticated imaging and statistical methods that enable us to observe

patterns at the right level. However, I believe there is yet a deeper reason for this stalemate. In my view, studies already presuppose a particular theoretical framework that brings along assumptions about what emotions ought to be. This makes the interpretation and comparison of results problematic, since the experimental designs themselves already presuppose that emotions must be located in specific brain regions, correspond to consistent patterns of physiological activity, and so on. Consequently, to solve this issue, we must take a step back and reflect on how we construct scientific categories of emotion that are then testable through empirical methods.

In the last part, I then consider different methodologies to construct a scientific framework for the study of emotions. I contrast three options that have been explored in the literature. One is eliminativism, either in its more general form as defended by Churchland (1981), or in its specific form regarding emotions as proposed by Griffiths. I reject eliminativism on the grounds that emotions have still been useful categories in a number of research projects in psychology and neuroscience, and hence we should consider other alternatives before elimination. Second, I examine explication, that is, the idea that we can explicate folk-psychological categories to form scientific categories. I contend that explication presupposes that folk-psychological categories ought to be theoretical themselves (as explication transforms folk categories into scientific ones), an assumption that I reject. Lastly, I consider reconstitution as proposed in the literature on mechanistic explanation (Bechtel, 2008; Bechtel & Richardson, 2010). Doing reconstitution means going back to the phenomenon to update scientific categories. In contrast to explication, reconstitution does not require that we identify folk categories with ill-defined scientific ones. Rather, it separates these two types of concepts and claims that scientific categories should be informed by the phenomena that constitute the explananda of a given scientific project. I claim that by reconstituting emotions in this sense, we can work towards an account of emotion kinds that does not risk changing the subject, does not eliminate emotions, and respects folk categories in their own right.

After proposing reconstitution as the methodology to follow, I compare different possible models of emotion kinds. By using the taxonomy of emotion kinds developed in the first part, I claim that only an account of emotions as functional kinds can be compatible with the empirical evidence reviewed in the third chapter while showing how science can do generalizations and projections on emotion categories. With this in mind, I formulate a functional account of emotions that, going back to folk categories and taking them as a starting point, abstracts away from them to formulate multiply realizable kinds. In this sense, emotions are proposed to be relatively automatic and rapid reactions to ecologically salient objects related to the well-being of an organism. These reactions may be realized in different ways in the brain and body, which is consistent with the observed heterogeneity of neural and physiological processes presented before.

This leads me to conclude that we need a reevaluation of some of the empirical findings in the light of a new functionalist perspective, showing that many of the disputes between different theories can be resolved or dissolved with this new approach. Furthermore, I show that the functional account can shed light on how to offer a theory of emotions that overcomes Griffiths's theoretical challenge



(i.e. that explains all emotions under the same framework) and that dissolves Barrett's empirical challenge, as lack of specificity, uniqueness and consistency is no longer problematic.

## References

- Barrett, L. F. (2006). Are Emotions Natural Kinds? *Perspectives on Psychological Science*, 1(1), 28–58.
- Barrett, L. F. (2012). Emotions are real. *Emotion*, 12(3), 413–429.
- Barrett, L. F. (2018). *How Emotions Are Made*. London: Pan Books.
- Bechtel, W. (2008). Mechanisms in Cognitive Psychology: What Are the Operations? *Philosophy of Science*, 75, 983–994.
- Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: decomposition and localization as strategies in scientific research*. Cambridge, Mass: MIT Press.
- Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, 61, 127–148.
- Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, 78(2), 67–90.
- Griffiths, P. (1997). *What emotions really are: the problem of psychological categories*. Chicago: The University of Chicago Press.
- Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in Cognitive Sciences*, 16(9), 458–466.
- Kragel, P. A., & LaBar, K. S. (2016). Decoding the Nature of Emotion in the Brain. *Trends in Cognitive Sciences*, 20(6), 444–455.
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3), 394–421.
- Kripke, S. A. (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological Bulletin*, 137(5), 834–855.
- Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, 35(3), 121–143.
- Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience*, 3(3), 207–233.

- Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI. *NeuroImage*, 16(2), 331–348.
- Putnam, H. (1975). The meaning of ‘meaning’. In *Mind, Language and Reality: Philosophical Papers*, Volume 2 (pp. 215–271). New York: Cambridge University Press.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1), 145–172.
- Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition & Emotion*, 23(7), 1259–1283.
- Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., ... Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin*, 144(4), 343–393.
- Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *Journal of Cognitive Neuroscience*, 22(12), 2864–2885.

# Is there room for a new foundation for the force interpretation?

Víctor J. Luque

The force interpretation of evolutionary theory was firstly proposed by Elliott Sober (1984). Sober argues that evolutionary theory is a theory of forces because, in the same way that different forces of Newtonian mechanics cause changes in the movement of bodies, evolutionary forces cause changes in gene and/or trait frequencies. As a result, selection, drift, mutation and migration would be the main forces or causes of evolution. The force interpretation is intended to expose the causal structure of the evolutionary theory. This is what Maudlin (2004) calls “quasi-Newtonian” theories. These are characterized by shaping them in a similar form to Newtonian mechanics. The main purpose of building quasi-Newtonian theories is to identify the causes that affect a particular system.

A theory of forces is structured by three different types of laws: zero-force laws, source laws, and consequence laws. Each law plays a specific role inside a theory. The zero-force law says what happens to a system when no forces act on it. Sober stipulates the Hardy-Weinberg law as the zero-force law of evolutionary systems. According to the Hardy-Weinberg law a diploid and ideal infinite population, where there is random mating (panmictic population) and whose individuals are viable and fertile, will remain or return to equilibrium (i.e. gene and genotype frequencies will remain stable) if no force acts on it, since Mendelian inheritance alone cannot change the allele frequencies. Thus, the Hardy-Weinberg law (like the law of inertia in classical mechanics), tell us how the system would behave if nothing disturbed it, assuring a neutral substrate where we can introduce forces.

Source laws tell us when forces will be present. Thus, the different formulations of distinct evolutionary forces describe the circumstances that produce such forces. For example, the force of migration appears when individuals move between populations. On the other hand, consequence laws allow us to compute the change in a system produced by forces. In the original formulation by Sober, the mathematical models of population genetics played this role.

Nevertheless, the appropriateness of the force interpretation has been challenged in the last years by several authors. On one hand, some authors (Walsh *et al* 2002, Matthen and Ariew 2002 and 2009, Pigliucci and Kaplan 2006) argued for a new view, the *statistical view*, where core evolutionary concepts like selection, drift, or fitness, denote mere statistical outcomes, not causal influences. The so called evolutionary forces should be conceptualized as statistical population-level tendencies, abandoning any causal role for them. On the other hand, other authors (Brandon 2006, McShea and Brandon 2010) accept the framework postulated by the force interpretation but disagree with the content of some specific parts. Essentially, they argue that, contrary to the classical view of the force interpretation, the Hardy-Weinberg law is not a good zero- force law

for evolutionary theory. Instead, they propose that the default state of evolutionary systems is not the lack of change but constant change. Thus, evolutionary systems would be always under drift, establishing the zero-force law of them.

I propose a new formalization of the force interpretation that avoids these critiques. The main strategy of this approach is to postulate a unique consequence law for all evolutionary systems (analogous to Newton’s second law of motion in classical mechanics): the Price equation. This equation describes the evolution of a population from one time interval (usually one generation) to another in a simple algebraic language (Frank 2012). Price’s equation is expressed in terms of covariances and expectations for describing evolution. In its usual form in biological literature, it is represented as follows:

$$\bar{w}\Delta\bar{z} = Cov(w, z) + E(w\Delta z)$$

where  $\Delta\bar{z}$  is the change in average value of character  $z$  from one time interval to another,  $w$  is the absolute fitness of an individual (number of descendants), and  $\bar{w}$  is the average fitness. The first term on the right-hand side is the covariance between fitness  $w$  and character  $z$ , and represents the change due to the contribution of character  $z$  to differential survival and reproduction. Usually, this term is interpreted as representing natural selection. The second term on the right-hand side is the expected value (the average) of the quantity  $\Delta z$ , that is, the change in the values of characters from the parents to the offspring due to processes involved in reproduction. This expected value is also called the *transmission bias* and can be caused by different factors such as mutation, recombination, selection at a lower level of organization, and so on. In conclusion, the Price equation defines evolutionary change as the sum of two terms (which can also be expanded).

The Price equation has been recently identified as the fundamental equation of evolution (Luque 2017; Queller 2017) because, among other reasons, it requires the fewest assumptions and all the other historical fundamental equations of evolution can be easily derived from it with additional assumptions. In addition, I have defended (Luque 2017) that the Price equation can play the same role in evolutionary theory that Newton’s second law of motion plays in classical mechanics. The idea here is that, much as Newton’s second law of motion is not supposed to provide a full description of the workings of physical systems but rather emphasises that a (potentially quite complicated) description of that system can be provided in terms of force, mass and acceleration, the proper use of the Price equation is to motivate the development of more detailed evolutionary equations that use its same basic language and logic. So, both Newton’s second law and the Price equation are what Thomas Kuhn named “generalization-sketches” (Kuhn 1970). These generalization-sketches are not models but sketches or schemas that provide a unifying framework in order to develop specific models. Moreover, The Price equation works as the single consequence law of evolutionary theory because it describes how evolutionary forces produce changes in a population, but do not determine how many causes exists, how these causes are, and so on. In addition, the Price equation encompasses all the traditional population genetic models as special cases, unifying all of them in a single mathematical framework.

Regarding the default state of evolutionary systems, I have proposed (Luque 2016) a new zero-force law for evolutionary theory that is not tied to any genetic system and therefore overcomes the limitations of the Hardy-Weinberg law. I have called it “the Principle of Stasis” which could be formulated as follows:

An evolutionary system where there is no selection, drift, mutation, migration, etc., so there is no evolutionary force, will not undergo any change (it will remain in stasis).

This formulation resembles to the law of inertia, more than the Hardy-Weinberg law, in two ways. First, it establishes a universal default state of no change for all evolutionary systems (and not only to diploid populations). Secondly, in the same way that we can get to the law of inertia through Newton’s second law of motion (if  $F = 0$  then inertia), we can also get to the actual default state of evolutionary systems if every term in the Price equation equals to zero ( $\bar{w}\Delta\bar{z} = 0$ ), i.e. no evolutionary forces. So, the Price equation supports the Principle of Stasis as the zero-force law of evolution.

In conclusion, the critiques to the force interpretation are not counterexamples to the force interpretation itself, but only to the original formulation elaborated by Elliott Sober. In opposition to the statistical view, the Price equation shows how to decompose the total evolutionary change due to each evolutionary force. Moreover, contrary to Brandon and McShea, the correct default state of evolutionary systems is stasis, no change. This new formalization of the force interpretation avoids these critiques. In addition, this new formalization has the benefit of being more closely allied with contemporary practice in evolutionary biology than the traditional force interpretation.

## References

- Brandon, R. (2006). The principle of drift: biology’s first law. *Journal of Philosophy*, 103:319–335.
- Frank S. A. (2012). Natural selection. IV. The Price equation. *J Evol Biol*, 25:1002–1019.
- Kuhn, T. S. (1970) Second thoughts on paradigms. In: Suppe F (ed) *The structure of scientific theories*. University of Illinois Press, Urbana, pp. 459–482.
- Luque, V. J. (2016). The Principle of Stasis: Why drift is not a Zero-Cause Law. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 57:71–79.
- Luque, V. J. (2017). One equation to rule them all: a philosophical analysis of the Price equation. *Biology and Philosophy*, 32(1):97–125.
- Matthen, M. & Ariew, A. (2002). Two ways of thinking about fitness and natural selection. *Journal of Philosophy*, 99:55–83.
- Matthen, M. & Ariew, A. (2009). Selection and causation. *Philosophy of Science*, 76:201–224.

- Maudlin, T. (2004). Causation, counterfactuals, and the third factor. In: J.D. Collins, N. Hall, L.A. Paul (Eds) *Causation and counterfactuals* (pp. 419–443). Massachusetts: The MIT Press.
- McShea, D. & Brandon, R. (2010). *Biology's first law: the tendency for diversity and complexity to increase in evolutionary systems*. Chicago: The University of Chicago Press.
- Pigliucci, M. & Kaplan, J. (2006). *Making Sense of Evolution: The Conceptual Foundations of Evolutionary Biology*. University of Chicago Press, Chicago, IL.
- Queller, D. (2017). Fundamental theorems of evolution. *American Naturalist*, 189(4):345–353.
- Sober, E. (1984). *The Nature of Selection*. Chicago: University of Chicago Press.
- Walsh, D., Lewens, T. & Ariew, A. (2002). The trials of life—natural selection and random drift. *Philosophy of Science*, 69:452–473.

# The enhanced indispensability argument. Where are we?<sup>1</sup>

Concha Martínez-Vidal

Navia Rivas de Castro

Universidad de Santiago de Compostela

The Quine-Putnam indispensability argument prompted several reactions against it. Field or Hellman, among others (Hellman 1989; Field 1980) have tried to prove that mathematics is descriptively dispensable to science. Others have contended that even if mathematics is descriptively indispensable for science, it does not follow that mathematical entities exist. Thus, Maddy (1992) has objected to the argument by pointing out that scientists do not take a holist viewpoint towards their theories. For instance, scientists posit idealized entities such as frictionless slopes and continuous fluids without thinking of them as existent objects.

Back in 2005, Alan Baker proposed a new version of the indispensability argument, one that does not include holism as one of its premises and that infers the existence of mathematical entities from their ‘genuine’ explanatory indispensability. The Enhanced Indispensability Argument (EIA) goes as follows:<sup>2</sup>

**EIA1.** There are genuinely mathematical explanations of empirical phenomena.

**EIA2.** We ought to be committed to the theoretical posits in such explanations.

**EIAC.** We ought to be committed to the entities postulated by the mathematics in question.

Different authors have advanced several cases of purported ‘genuine’ mathematical explanation (the cicadas, the Königsberg bridges, the honeybee comb...) to conclude, by inference to the best explanation, that there are mathematical objects. However, nominalists (Saatsi 2011, Knowles and Saatsi, Forthcoming) contend that—even if mathematics is expressively indispensable—it is not explanatory indispensable; at least not in a way that conveys a commitment to mathematical objects. A debate has been going on since i) there is no agreement on the character of purported genuine mathematical explanation; and, ii) as Marcus<sup>3</sup> has noted, the argument “leaves open the question of how one is supposed to determine the commitments of an explanation. EI[A] refers to the theoretical posits postulated by explanations, but does not tell us how we are supposed to figure out what an explanation posits. If the commitments of a scientific explanation are found in the best scientific theory used in that explanation,

---

<sup>1</sup>This work was supported by the Spanish Ministry of Economy and Competitiveness under projects FFI2013-41415-P, FFI2017-82534-P and FEDER.

<sup>2</sup>Mancosu 2008, 137; Baker 2005; Lyon and Colyvan 2008; Baker 2016 are some of the various places where we can find it.

<sup>3</sup>Marcus 2018.

then EI[A] is no improvement on QI. If, on the other hand, EI[A] is supposed to be a new and independent argument, its proponents must present a new and independent criterion for determining the commitments of explanations.”

Baker (2016) intends to make the debate advance by providing some hints as to how to figure out what an explanation posits; in particular, he advances that mathematics contributes to scientific explanation scope generality and, above all, topic neutrality. This contribution of mathematics makes it the best explanation available, and hence mathematical explanation commits us to the objects that figure in it. Knowles and Saatsi (Forthcoming) acknowledge that the nominalists have neither answered ‘the challenge from explanatory generality,’ nor provided an account of explanatory generality and of how mathematics contributes to it. But, they counter argue that the debate is still at an impasse. They claim that in order to supersede it, the answer to the question about the ontological commitments of the mathematics in distinctive mathematical explanation has to be given independently of the different intuitions at stake. Hence, they articulate an answer in terms of the counterfactual theory of explanation. They acknowledge that mathematics plays an indispensable role, in explanation but contend that the role of mathematics in those cases is just to provide “cognitive salience.” They conclude that cognitive salience does not allow for any ontological commitment.

Baron (2016, 2017), aims to characterize genuine mathematical explanation in terms of some of the well-established theories of explanation available (Baron 2016). In his 2017, he puts forward a theory of explanation, the deductive-mathematical theory of explanation, that, purportedly, allows for the distinction between genuine and descriptive mathematical explanation, the kind of explanation that proves mathematics to be explanatory indispensable and that should allow the EIA to get through.

In this contribution, we intend to show that, even if some progress has been made in order to get a better understanding of how mathematics explains, the debate still remains in a standoff.

From a certain perspective, the situation seems to be that Marcus’ dilemma has not been answered. While it is clear that explanations formulated in mathematical terms add to explanations formulated in terms of empirical facts, it is far from clear whether its contribution is to be understood in an ontologically committing way (Baron’s position) or just as a matter of epistemic convenience (Knowles and Saatsi, Forthcoming). Alternatively, it might be the case that cognitive salience suffices for light-weighted platonism, a possibility that Baron contemplates (2016).

## References

- Baker, Alan, 2005. ‘Are there Genuine Mathematical Explanations of Physical Phenomena?’ *Mind*, 114:223–238.
- Baker, Alan, 2017. ‘Mathematics and Explanatory Generality.’ *Philosophia Mathematica*, 25(2):194–209.



- Baron, Sam, 2016. 'Explaining Mathematical Explanation.' *The Philosophical Quarterly*, 66(264):458–480. <https://doi.org/10.1093/pq/pqv123>
- Baron, Sam, 2017. 'Mathematical Explanation by Law.' *The British Journal for the Philosophy of Science*, forthcoming. <http://philsci-archive.pitt.edu/13370/>
- Colyvan, Mark. 2010. 'There is No Easy Road to Nominalism.' *Mind*, 119(474)(April 2010):285–306.
- Field, Hartry, 1980. *Science Without Numbers*. Princeton: Princeton University Press.
- Hellman, Geoffrey, 1989. *Mathematics Without Numbers*. New York: Oxford University Press.
- Knowles, Robert and Saatsi, Juha, Forthcoming. 'Mathematics, Explanatory Generality, and Ontological Commitment.' [https://juhasaatsidotorg.files.wordpress.com/2018/02/mathematics\\_and\\_explanatory\\_generality\\_single\\_names.pdf](https://juhasaatsidotorg.files.wordpress.com/2018/02/mathematics_and_explanatory_generality_single_names.pdf)
- Lyon, Aidan and Colyvan, Mark, 2007. 'The Explanatory Power of Phase Spaces.' *Philosophia Mathematica*, 16(2):227–243.
- Maddy, Penelope, 1992. 'Indispensability and Practice.' *The Journal of Philosophy*, 89:275–289.
- Mancosu, Paolo, 2008. 'Mathematical Explanation: Why It Matters.' In P. Mancosu (ed.), *The Philosophy of Mathematical Practice*, Oxford: Oxford University Press.
- Marcus, Russell, 2018. 'The Indispensability Argument in the Philosophy of Mathematics.' At *Internet Encyclopedia of Philosophy*, <https://www.iep.utm.edu/indimath/>
- Saatsi, Juha, 2011. 'The enhanced indispensability argument.' *The British Journal for the Philosophy of Science*, 62:143–154.

# Structural and organisational conditions for machine-likeness

Guglielmo Militello

The concept of ‘machine-like system’ has been extensively employed in the conceptual framework of neo-mechanistic accounts, since many biological mechanisms have been regarded as the functional components of a system which behaves like a machine (Bechtel and Richardson 1993 (2010); Glennan 1996). Moreover, the analogy between machines and certain biological macromolecular structures plays a pivotal role in nanotechnology, as some kinds of macromolecules are artificially reproduced by considering them machine-like systems. Yet, it has been recently argued (Moore 2012; Nicholson 2013; Skillings 2015) that molecular devices (biological and synthetic as well) are not machines since they are subject to physicochemical forces that are different from those of macroscopic machines.

Despite this, the structural and physicochemical conditions that allow both macroscopic machines and microscopic devices to work and perform new functions, through a combination of elemental functional parts, have not yet been examined. In order to fill this void, this talk has a threefold aim: first, to clarify the structural and organisational conditions of macroscopic machines and microscopic devices; second, to determine whether the machine-like analogy fits nanoscale devices; and third, to assess whether the machine-like analogy is appropriate for describing the behaviour of some biological macromolecules.

In order to address these issues, the presentation will be divided into three parts. Firstly, the criticisms (Moore 2012; Nicholson 2013; Skillings 2015) levelled at the machine-likeness of molecular devices will be discussed. Particular attention will be devoted to the question of whether or not the different structures and the different physicochemical behaviour of macroscopic and molecular machines prevent us from employing the term machine at the molecular level. Secondly, the structural and physicochemical conditions underlying both macroscopic (e.g. mechanical machines) and microscopic (biological as well as synthetic) devices will be examined. Finally, a comparative analysis of synthetic (e.g. artificial DNA architectures) and biological (e.g. myosin, dynein, and  $F_0F_1$ ATPase) machines will be carried out so as to appreciate their differences and the distinctive character of biological molecular devices.

In line with Militello and Moreno (2018), this study suggests that, even though macroscopic and molecular machines exhibit different component parts, a distinct design, and different physical laws (Newtonian mechanics underlying macroscopic machines and quantum mechanics governing microscopic ones), both kinds of devices share a common organisation which is the ontological basis for being a machine: they are sets of functional components that harness a flow of energy so as to do work and perform function(s). This essential characteristic

stresses that, contrary to what have been claimed by Moore (2012), Nicholson (2013), Skillings (2015), a number of microscopic devices can rightly be regarded as machines.

It will be argued that the machine-likeness of certain biological macromolecules was of paramount importance in prebiotic and biological evolution, because it opened up a new domain of functional diversification: new forms of mechanistically-complex functions could be achieved through different combinations of parts. Furthermore, although biological and synthetic molecular devices share common physicochemical properties, they exhibit different organisations, since biological macromolecular machines are integrated with other macromolecules within the cellular biochemical network by determining a ‘closure of molecular machines’.

## References

- Bechtel, W. and Richardson, R. C. (1993 (2010)). *Discovering Complexity: Decomposition and Localisation as Strategies in Scientific Research*. MIT Press, Cambridge (Mass)
- Glennan, S. (1996). Mechanisms and the Nature of Causation. *Erkenntnis*, 44:50–71.
- Moore, P. B. (2012). How Should We Think About the Ribosome? *Ann Rev Biophys*, 41:1–19.
- Nicholson, D. J. (2013). Organism  $\neq$  Machines. *Stud Hist Phil Biol Biomed Sci*, 44:669–678.
- Skillings, D. J. (2015). Mechanistic Explanation of Biological Processes. *Philos Sci*, 82:1139–1151.
- Militello, G. and Moreno, Á. (2018). Structural and organisational conditions for being a machine. *Biol Phil*, 33:35. <https://doi.org/10.1007/s10539-018-9645-z>

# Los mundos posibles de Lewis son entrópicos

Carlos Montojo  
Miguel García Valdecasas  
Universidad de Navarra

**Abstract.** I argue that David Lewis' possible worlds are entropic unless they are worlds where the events are decidable *more geometrico*, in which case this world does not necessitate at all a theory of possible worlds, and the theory of possible worlds ceases to be a Best System Account and can be discarded. But the theory of possible worlds is necessary because objective chance is an observable feature of reality. Chance events introduces the need to analyze events by comparison of possible worlds and this comparison is necessary to determine which is our actual world. Such comparison is impossible or undecidable without the principle of maximum entropy. Hence, entropy is a feature of all of Lewis' possible worlds.

La teoría de mundos posibles (TMP) de Lewis predica la existencia real (como eventos) de todos los mundos posibles. Se trata de mundos que son topológicamente medibles en función de su semejanza (Lewis 1986a, p. 2 ).

En el marco de su teoría de la probabilidad, el presente trabajo defiende esta tesis: solo pueden ser denominados “mundos posibles” aquellos mundos a los que se les puede aplicar el principio de máxima entropía en la descripción de eventos no decidibles *more geometrico*.

El razonamiento se construye sobre una consideración introductoria y tres proposiciones:

La consideración introductoria es que el concepto de “azar” (*chance*) de Lewis hace referencia a una realidad objetiva y que dicha realidad hace posible tanto la TMP como la teoría de la probabilidad de Lewis. Según Lewis el azar es probabilidad subjetiva objetivada es una característica de la realidad que nos permite establecer nuestra creencia (*credence*) sobre cuál es el mundo actual en el que vivimos, mediante la consideración de (y la comparación con) todos los mundos posibles. Si estuviera absolutamente especificado cuál es el mundo en el que estamos, perfectamente descrito en su formalidad, no sería precisa la teoría de la probabilidad, el concepto de azar sería redundante y la TMP sería superflua pues habría otras teorías más simples para conocer nuestro mundo.

Por eso el primer paso es determinar que la teoría de la probabilidad de Lewis, fundamentada en su TMP es un *Best System Account*. Es fundamental aceptar para ello que es una teoría necesaria porque no todos los conjuntos de eventos pueden ser descritos mediante leyes que se ajustan perfectamente a las observaciones, mediante proposiciones analíticas de exactitud geométrica. Si así fuera

la TMP no sería un requerimiento epistemológico, sería epistemológicamente superflua. Pero no lo es pues en la realidad (el espacio de eventos que percibimos) existe una característica que hace necesaria su utilización para analizar los eventos que observamos. Sostengo que dicha característica es la indecidibilidad *more geométrico* de sistemas de eventos. Si todos los eventos fueran decidibles mediante su descripción instrumentada por leyes o ecuaciones descriptivas; entonces no sería preciso analizar mundos posibles. Estos mundos podrían existir y la teoría podría ser una teoría interesante; pero otra más simple, la descripción de la realidad *more geométrico*, ocuparía su lugar. Pero no es así, existen eventos indecidibles salvo mediante la consideración del azar, o en otras palabras, mediante una teoría de la probabilidad.

El segundo paso del razonamiento supone explicar que la teoría de la probabilidad de Lewis, en el marco de la TMP, precisa de la comparación de mundos. Sabemos que un mundo es el contrafactual y no el actual (mirando al pasado) y creemos que es un mundo posible que no será nuestro mundo actual (en el presente), porque podemos determinar su configuración y compararla con el mundo actual para determinar si estamos hablando del mismo mundo o de mundos diferentes. Esta afirmación se fundamenta en el principio de Superveniencia Humeana, que sostiene que el mundo es un mosaico de eventos puntuales relacionados y que toda la verdad que podemos conocer del mundo es precisamente la localización espaciotemporal de estos puntos y sus relaciones. En palabras de Lewis, la Superveniencia Humeana:

It is the doctrine that all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another.  
(Lewis 1986b: ix)

Por eso, la comparación de mundos es fundamental para determinar cuál es nuestra combinación de puntos, nuestro mundo.

El tercer paso supone defender una hipótesis: la comparación entre mundos es irrealizable sin la aplicación del principio de máxima entropía de la información. El principio de máxima entropía en la información supone que en aquellos casos en los que la información no permite decidir exhaustivamente la configuración de un sistema, se debe afirmar que dicha configuración se despliega maximizando la entropía, es decir, maximizando la función  $H = -\sum p(x) \cdot \log(p(x))$  (Shannon 1963, Pierce 1980).

Dado que la TMP requiere de comparación entre mundos y que dicha comparación requiere del principio de entropía máxima en la información, no es posible determinar que un mundo es posible sin el principio de máxima entropía. Este principio, es un principio objetivo, *a posteriori*, no subjetivo o apriorístico que se predica del mundo actual y de los mundos posibles. En consecuencia, estamos obligados a mantener que los mundos posibles de Lewis o bien son mundos entrópicos, o no podemos afirmar que sean posibles, pues para medir la distancia entre mundos precisamos determinar la distribución de propiedades mediante curvas de distribución que no se pueden decidir si no se determinan mediante el principio de máxima entropía.

Se podría argüir contra esta conclusión que Lewis (Lewis 1970, p. 177) habla de mundos posibles como mundos *concebibles*. Por tanto, se podría negar la

hipótesis 2 de este análisis y decir que los mundos posibles de Lewis no son comparables mediante la determinación de distancias entrópicas.

La respuesta a esta objeción se fundamenta en el principio de superveniencia humeana. Este principio (Lewis 1986a p. 14 ) tiene dos implicaciones principales: (1), que toda la verdad sobre el mundo sobreviene a la distribución de propiedades naturales en dicho mundo, (2) que dichas propiedades son intrínsecas a objetos de dimensión puntual y relaciones espaciotemporales (Brian Weatherston *David Lewis* 2014). Suponer la mínima formalización implica suponer la máxima entropía. De esta manera, a falta de un principio entrópico podemos afirmar que *cualquier* distribución puede subyacer bajo cualquier sistema, y *cualquier* distribución significa *infinitas* posibles distribuciones. En dicho escenario, sería imposible comprobar si el mundo que hemos concebido es un mundo posible o, quizás, el mundo actual en el que vivimos, al no poder trazar una comparación de distribuciones de probabilidad (Lewis 1980, p. 279). Por lo tanto, no sabemos si estamos ante mundos posibles o actuales, y la teoría de mundos posibles deja de ser eficiente.

## Bibliografía sucinta

- Lewis D. *On the Plurality of Worlds*, Blackwell Publishing, (1986a) pp. 2–27.
- Lewis D. Anselm and Actuality, (1970) *Nous*, Vol. 4 (May) pp. 175–188.
- Lewis D. *A Subjectivist's Guide to Objective Chance*, Berkeley Ed. *Studies in Inductive Logic and Probability*. (1980) pp. 278–280.
- Lewis D. *Philosophical Papers*, Vol. II, Oxford: Oxford University press (1986b).
- Shannon C. *The Mathematical Theory of Communication*, (1963) Illinois University, p. 53.
- Pierce J. R. *An Introduction to Information Theory. Symbols, Signals and Noise*, (1980) Dover Publications Inc.

## Interés de la ponencia

Una limitación de mundos posibles a aquellos que son entrópicos encuadra el método científico en unos límites estrictos acordes con la metodología experimental utilizada en la práctica.

La inflación de mundos posibles atribuida a David Lewis no lo sería tanto y la re-interpretación de su Teoría de Mundos Posibles permitiría la fundamentación filosófica de métodos estadísticos como el método de máxima verosimilitud. Finalmente, se establecería un puente preciso e indiscutible entre la teoría de la probabilidad de Lewis y la objetividad del método, basado en una característica necesaria de la información empírica.

# The dappled world of evolutionary biology: Scientific pluralism and the extended evolutionary synthesis

Íñigo Ongay de Felipe

This paper examines the present state of the art in Evolutionary Biology with a view of showing how the sort of scientific pluralism Nancy Cartwright argues for in *The Dappled World* casts lights on some debates pertaining to the field.

There are winds of change blowing in the domain of Evolutionary Biology. Much debate has recently arisen within the territory of Evolution with regard to a variety of new developments in EvoDevo (short for Evolutionary Developmental Biology), Epigenetics, Phenotypic and Developmental Plasticity, Behavioral Biology, Social Transmission of information and Niche Construction and Ecological Inheritance to name just a few areas of concern. While there is no doubt that these research threads are polymorphic in character and a general articulation of such variety of research programs is to date lacking, there is still something they all share in common, namely: they all seem to point out to a much more multidimensional conception of Evolution than that envisioned by the architects of the Modern Synthesis in the first half of the 20th century. It is worth considering that although very few of those involved in the debate would argue that the new research directions call for a complete and wholesale debunking of the Modern Synthesis in its entirety, many do insist that the construction of an new and extended evolutionary paradigm is required: one which while allowing for mechanisms and evolutionary factors long neglected in the traditional neo-Darwinian framework, leaves ample room for the reassessment of some pivotal tenets of the Modern Synthesis which have proved deficient in empirical support. In a nutshell, whereas the Modern Synthetic framework rests on a view of evolution in which the organisms react passively to the change in relative genetic frequencies and the action of natural selection, many of the proponents of the Extended Synthesis favor a vision of evolution where organismal activity is accorded a more active role in accounting for evolutionary change. Such a shift of focus has led to a vindication of a pluralistic theory of inheritance, which goes in many ways beyond the demands of the Weismann barrier and the central dogma in Molecular Biology. While an increasingly growing array of relevant researchers working in the fields of Epigenetics, Development and Ecology (from Eva Jablonka to Kevin Laland, Gerd Müller or Denis Noble just to list but a few) are presently pushing this process ahead with exultance, other more traditionally inclined biologists have –perhaps unsurprisingly– welcome the news with a foreseeable mixture of disdain and skepticism contending that very little is truly novel in these novelties.

It is not my objective to choose sides in this current scientific discussion. Even if there are undoubtedly many worth exploring philosophical ramifications to

such a thriving state of affairs in Biology, the goal of this paper is not to analyze all the implications of the Extended Evolutionary Synthesis for philosophy of science in general or philosophy of biology in particular. What I propose instead is more modest in nature: I will put the focus on the relevance this new scenario bears to a pluralistic view of science. In that regard, the scope of the paper is threefold. Firstly, the author shall draw a comparison between the theoretical structure of the Modern Synthesis and the Extended Evolutionary Synthesis to show how the later is not only much broader in its explanatory scope but also encompasses a number of evolutionary factors that the proponents of the Modern Synthesis had largely excluded from their (comparatively more one-sided) conception of evolution. Secondly, the paper goes on to argue that the sort of scientific pluralism advocated by philosophers such as Nancy Cartwright, John Dupré or Gustavo Bueno constitutes a philosophical account of scientific practice that casts light on much of what is presently going on in the milieu of Evolutionary Biology. In this respect, contrary to other more traditional views in Philosophy of Science ever since the advent of the discipline with the Vienna Circle, the kind of pluralism these authors and some others defend helps understand the ways in which scientific growth often promotes plurality at the cost of parsimony and unification. Conversely, by reflecting on some ideas advanced by H. Chang and P. Kyle Stanford, it will be shown that an excessively stringent emphasis on epistemic simplicity may sometimes go hand in hand with the undesired result of hampering scientific development. Thirdly, the point will be made that such an epistemically pluralistic conception of science comes with a number of ontological implications about both the *disunity of science* and the *disorder of things* in nature. In particular the author will show that much like there are many more things in evolution than dreamt of in the Modern Synthesis, nature does sometimes show a disconcerting taste for diversity, promiscuity and *the pomp of superfluous causes*. The upshot of the analysis resonates with clarity in this connection to the effect that there is something Nancy Cartwright and others got entirely right: *it is, indeed, a dappled world*.



# Acciones y motivaciones para la constitución de lo común: Una propuesta desde la ciencia ciudadana

**Maite Pelacho**

Universidad del País Vasco/Euskal Herriko Unibertsitatea

Fundación Iberoamericana

maite.pelacho@gmail.com

**Palabras clave:** ciencia ciudadana, bienes comunes, motivaciones, cooperación.

La ciencia no es un sistema abierto ‘per se’ ni mucho menos un sistema sostenible. Es un hecho que los recursos del conocimiento científico-tecnológico se gestionan en muchos casos de un modo ineficiente y/o no plenamente accesible. La incorporación de elementos del mercado en la gestión de la ciencia se ha ido incrementando como supuesta solución a los diversos problemas de su gestión y mantenimiento. Sin embargo, tal mercantilización se ve frecuentemente contestada aludiendo a la naturaleza de la ciencia como un bien público (Callon, 1994; Vega, 2001; Novotny 2005), aportando argumentos diversos que deberían impedir o frenar su privatización. Como alternativa para hacer frente a este discutible dilema, que afecta tanto a la generación de conocimiento científico como a su difusión –con implicaciones epistemológicas y ético-políticas–, se ha abierto en la última década toda una línea de estudios y activismo que ha extendido la noción de bienes comunes o ‘recursos de uso común (RUC)’ (Ostrom, 2000) más allá de los recursos físicos. Entre los ‘nuevos comunes’ se encuentran los ‘comunes del conocimiento’ (Hess y Ostrom, 2016) y, entre éstos, el conocimiento científico.

De acuerdo con los estudios ostromianos, los bienes comunes (pesquerías, lagos, bosques, sistemas de irrigación...) se constituyen y preservan como tales como consecuencia de unas determinadas acciones y reglas de uso acordadas dentro de las diversas comunidades o sistemas de autogobierno. Más de cuatro décadas de investigación empírica en muy diversos lugares del mundo revelan el hecho de la existencia multisecular –sostenible y accesible– de este tipo de recursos y las correspondientes comunidades, poniendo de manifiesto que la acción colectiva no sólo no está necesariamente condenada al fracaso, sino que ha conducido a modos de gestión altamente eficientes y equitativos. Junto a la propuesta de la acción colectiva bajo determinadas reglas de uso, es también crucial la conclusión observacional de la existencia de motivaciones y capacidades para resolver dilemas sociales en torno a los recursos ligadas a intereses que van más allá del personal (Ostrom, 2009) en los individuos y grupos que gestionan recursos de uso común, siendo claves también el cultivo de los ámbitos de la confianza, la reciprocidad y la cooperación (Hess y Ostrom, 2016). El redescubrimiento de los bienes comunes o RUC pone en entredicho el modelo de ‘elección racional’

que presupone una naturaleza humana individualista, así como la clasificación dual y estática de los bienes como públicos o privados, regidos respectivamente por dos únicas formas institucionales, el Estado y el mercado.

En mi investigación propongo comprender la ciencia como un recurso de uso común, lo cual implica entender cómo deberíamos gestionarla para lograr efectivamente su constitución como tal tipo de recurso –sostenible y accesible– junto a la constitución de las comunidades de prácticas que hacen que esto sea posible. Con la expresión ‘lo común’, y siguiendo a Laval y Dardot (2015), quiero hacer referencia al principio político que alude tanto a una cualidad del actuar (el actuar común) como a aquello a lo que este actuar da lugar (los bienes comunes). Bajo este principio, pretendo entonces entender qué acciones relativas a la ciencia pueden ser prácticas idóneas para la constitución de ‘lo común’. Mi propuesta es que una de estas prácticas, que ha ido surgiendo –o renaciendo– en las dos últimas décadas, es la ciencia ciudadana, entendiendo esta noción como la participación del público general en la investigación científica, al contribuir activamente los ciudadanos a la ciencia bien a través de su esfuerzo intelectual y conocimiento o con sus herramientas y recursos (European Commission, 2014). El fenómeno creciente de la ciencia ciudadana en las dos últimas décadas se ve reflejado en la multiplicación de proyectos y laboratorios ciudadanos, junto a la de asociaciones, fundaciones, plataformas on-line y observatorios nacionales y locales en todo el mundo, todo ello reflejado, al menos parcialmente, en convocatorias de proyectos, y en el creciente número de publicaciones científicas indexadas. Las denominaciones son muchas y diversas –participación del público en la investigación científica, ciencia participativa, ciencia cívica, ciencia amateur, ciencia de crowdsourcing, ciencia 2.0, entre otras– así como las tipologías y definiciones. Como rasgo definitorio y común a las diversas definiciones puede señalarse el hecho de la cooperación, más allá de las fronteras institucionales, entre los científicos profesionales y el público en las actividades que conforman el proceso científico (Haklay, 2015).

Pretendo entonces, a través del conocimiento del desarrollo y resultados de diversos proyectos, comprender cómo y por qué la ciencia ciudadana puede dar lugar a una ciencia más sostenible y abierta, y a la vez a comunidades más corresponsables e inclusivas. Mi hipótesis es que la ciencia ciudadana, al superar diversas y ficticias fracturas –entre ciencia y sus entornos, entre expertos y legos– promueve la cooperación en muy distintos niveles y ámbitos y, por tanto, una mejor gobernanza del sistema científico-tecnológico orientada a la constitución de lo común, con la consiguiente reducción de los conflictos –no sólo económicos– asociados al reduccionista esquema ‘público vs. privado’.

Aun con todo, lograr el doble objetivo –la constitución de los bienes comunes como resultado de la acción común junto a las comunidades que los hacen posibles– a través de la ciencia ciudadana dependerá significativamente de que las motivaciones de los diversos actores estén fuertemente relacionadas con la corresponsabilidad respecto del interés general. Conocer las motivaciones entre los distintos grupos de interés (usuarios, científicos, gestores, políticos) involucrados en diferentes proyectos de ciencia ciudadana en España es un objetivo de esta investigación: afirmar que las prácticas de ciencia ciudadana son constituyentes de lo común requiere que tales motivaciones están realmente ligadas a la consecución de intereses generales junto a los intereses personales. Con el fin

de comprender mejor las motivaciones en ciencia ciudadana, esta investigación incluye la realización de un estudio empírico de alcance nacional, realizada junto con la Fundación Ibercivis, en un momento en que el número de prácticas se está multiplicando, a la vez que desde las instancias políticas locales y nacionales ha empezado a reconocerse explícitamente su potencial científico y social (MINEICO, 2017). Mediante ese estudio empírico se busca asimismo promover la reflexión y deliberación compartida respecto de las posibles y diversas implicaciones epistémicas, éticas y políticas de la práctica creciente de la ciencia ciudadana.

## Referencias

- Callon, M. (1994). Is Science a Public Good? Fifth Mullins Lecture, Virginia Polytechnic Institute, 23 March 1993. *Science, Technology, & Human Values*, 19(4):395–424.
- European Commission. (2014). Socientize consortium. *White Paper on Citizen Science for Europe*. En: <http://www.socientize.eu/?q=eu/content/white-paper-citizen-science>
- Haklay, M. (2015). Citizen Science and Policy: A European Perspective. Washington, DC: Woodrow Wilson International Center for Scholars.
- Hess, Ch., Ostrom, E. (2016). *Los bienes comunes del conocimiento*. Quito: IAEN-Instituto de Altos Estudios Nacionales del Ecuador. (Original en inglés, 2007).
- Laval, C., Dardot, P. (2015). *Común. Ensayo sobre la revolución en el siglo XXI*. Barcelona: Gedisa. (Original en francés, 2014).
- MINEICO (2017). Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020. En: <http://www.idi.mineco.gob.es/stfls/MICINN/Presencia/FICHEROS/2018/PlanEstatalIDI.pdf>
- Nowotny, H., Pestre, D., Schmidt-Aßmann, E., Schulze-Fielitz, H., Trute, H.-H. (2005). *The Public Nature of Science under Assault. Politics, Markets, Science and the Law*. Berlin/Heidelberg: Springer.
- Ostrom, E. (2000). *El gobierno de los bienes comunes. La evolución de las instituciones de acción colectiva*. México: México UNAM-CRIM-FCE (Original en inglés, 1990).
- Ostrom, E. (2009). Beyond Markets and States: Polycentric Governance of Complex Economic Systems. En: [http://www.nobelprize.org/nobel\\_prizes/economic-sciences/laureates/2009/ostrom\\_lecture.pdf](http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2009/ostrom_lecture.pdf)
- Vega, J. (2001). Ciencia privada, conocimiento público. *ISEGORÍA*, 25:247–261.

# The science of ethics & evolutionary moral realism

João Pinheiro

In this talk we start by introducing recent developments in the Science of Ethics, which has seen its empirical and scientifically study substantially developed over the last few decades, becoming itself a cross-disciplinary field of Applied Science, with import from such a myriad of disciplines as Evolutionary Biology, Ethology, Evolutionary, Developmental, Social, and Moral Psychology, Cognitive Neuroscience, Sociocultural Anthropology, Ethnography, etc.. When brought together, the insights from these disciplines allow us to model Ethics as a facet of human behavioural ecology [e.g. Flanagan 1996, and Flack & Krakauer 2009], where moral cognition is subject to EvoDevo constraints (not unlike any other biological trait), and plays a role in the construction of our social niche, for which it can be causally selected.

Whence the origin of its biological function. The biofunction of moral systems is to control interactions between individuals (and, possibly, groups), so as to maximize individual (and, possibly, group) fitness, which will consist, by and large, in (potential) social conflict management, and therefore in boosting cooperation, which has been shown [e.g. Hardin 1968, Axelrod & Hamilton 1981, & Nowak 2006] to be the behavioural strategy that best benefits all individuals' (and groups') fitness, were all the interactors to cooperate.

However, despite the proliferation of scientific approaches to Ethics stimulating methodological naturalism in Metaethics, there remains serious doubts regarding the tenability of a fully naturalistic and realist programme, especially because moral facts, understood realistically (i.e. as independent from our evaluative attitudes), and values, are still widely believed to be non-analysable in scientific terms.

The pseudo-dilemma posed by “Street-style” evolutionary debunking arguments of moral realism [Street 2006] are exemplary of this scepticism [cf. Copp 2008, & Kahane 2011]. These arguments, founded on evolutionary theory, depend on the truth-value of one fundamental empirical & epistemic premise: moral cognition is truth-tracking in virtue of its natural history. Naturalists sceptical about moral realism (as previously conceived), claim that, while denying the truth of this proposition, the realist is left blind as to the possibility of a naturalistic explanation of moral realism; whereas while accepting it, the realist is left with the difficult task of providing us with a cogent “just-so-story” of truth-tropic cognitive evolution.

Until quite recently, most naturalists would bite the bullet and agree that the empirical hypothesis that there's only positive selection for truth-tropic moral cognition is false, thus recognizing the frailty of the moral realist's position. Here we'll survey a novel scientific research programme/metaethical theory which

might just turn the tides. According to this theory, which we'll call *evolutionary moral realism*, moral truths are grounded in facts about cooperation [e.g. Brosnan 2011, and Sterelny & Fraser 2016] or sociability [e.g. Copp 2007 & 2009], or, if you prefer, there are facts against which moral propositions can be judged to be true or false, and these are «objective facts about the practices and norms that would promote stable cooperation» [Sterelny & Fraser 2016:5].

This theory establishes a version of moral realism compatible with the thesis that «the primal scene of morality (...) is (...) one in which we do something together» [Korsgaard 1996:275] (fitting nicely in the social niche construction theory of morality as part of human's behavioural ecology). What is more, by equating payoffs with fitness values and strategies with behaviours according to norms, evolutionary game theoretical approaches formally predict which moral practices will become evolutionary stable strategies in specific environments, for normative strategies that maximize utility will thrive, meaning that they'll enjoy of higher fitness-values [cf. Hammerstein 2012, el Moulden et al. 2012, & Sterelny 2012].

As Street would argue, one should accept the empirical hypothesis that natural selection is blind to truth-tracking as being more parsimonious. But, in fact, epistemic rules of inferences to the best explanation actually recommend we pick the loveliest of *alternative* explanations [Lipton 1991 & 2000], and, according to evolutionary moral realism, evolutionary explanations of moral cognition in terms of fitness-tracking and in terms of truth-tracking are *not* alternatives, but, rather, *complementary* explanations at different levels of analysis (at “the consequence level”, and “the ecological level”, respectively) [cf. Goode & Griffiths 1995, and Griffiths & Wilkins 2015], such that one may safely infer *both*.

However, and notwithstanding the truth-tropic nature of our moral cognition, it remains to be shown whether «in favourable cases there is a reliable causal connection between moral opinion and these facts» [Sterelny & Fraser 2016:17] about which behavioural strategies are evolutionarily stable. For example, if the evolutionary psychologists are right in that our cognition is adapted to the environment of the Pleistocene [Richerson & Boyd 2000], then, given the role played by kin & group selection in human evolution, and because we inhabit substantially different environments from the Pleistocene, we're unlikely to produce as many adaptive moral beliefs *qua* moral truths when inter-group conflict and cooperation with strangers are in case [Robinson 2013:75]. As such, our moral cognition might not be reliable in our present and global state of affairs, and moral error theorists [e.g. Joyce 2016] may still rejoice in the fact that moral error is rife.

# Extremal axioms in mathematics

**Jerzy Pogonowski**

Department of Logic and Cognitive Science  
Adam Mickiewicz University in Poznań, Poland

The term *extremal axiom* was introduced in the paper (Carnap, Bachmann, 1936). These authors tried to develop a general approach to axioms which were supposed to characterize in a unique way intended models of some important mathematical theories. They took into account: Hilbert's completeness axiom from his *Grundlagen der Geometrie* (Hilbert, 1899), Peano's axiom of induction in arithmetic (Peano, 1889) and Fraenkel's axiom of restriction in set theory (Fraenkel, 1928). They worked in type theory and tried to express in that language the fact that an extremal axiom is either a *maximal axiom* (like Hilbert's completeness axiom, which was later replaced by the continuity axiom for real numbers) or a *minimal axiom* (like axioms of induction or restriction, mentioned above). It may be worth recalling here the Hilbert's axiom of completeness which stated that: *To a system of points, straight lines, and planes, it is impossible to add other elements in such a manner that the system thus generalized shall form a new geometry obeying all of the five groups of axioms.*

Some critical remarks concerning the project in question can be found in (Hintikka, 1986) and (Fraenkel, Bar-Hillel, Levy, 1973). An earlier work (Carnap, 1930) containing a flawed proof of his *Gabelbarkeitssatz* and in a sense related to the project under consideration is discussed in (Awodey, Carus, 2001). The work (Lindenbaum, Tarski, 1936) solves similar problems, among others.

Our talk will be devoted mainly to historical remarks concerning extremal axioms. Because extremal axioms are supposed to determine models uniquely, they are connected with such properties as *categoricity*, *categoricity in power* and *completeness*. Obviously, categoricity implies completeness, but the converse implication in general does not hold – cf. e.g. (Tarski, 1940).

Limitative theorems obtained in the 20th century (Löwenheim, Skolem, Gödel, Rosser, Tarski, Church, Turing) have shown possibilities and restrictions in the characterization of such fundamental notions as: consistency, completeness, categoricity, decidability. They also shed new light on the role of extremal axioms in mathematics – cf. e.g. (Awodey, Reck, 2002) and (Scanlan, 1991) where, among others, the influence of the American Postulate Theorists (like Veblen and Huntington) on the emergence of the concepts of categoricity and completeness is discussed.

Categorical characterizations of natural numbers (Peano), Euclidean geometry (Hilbert) and real numbers (as a unique completely ordered field – here the extremal axiom of continuity is essential, cf. (Dedekind, 1872)) are all well recognized in mathematics. A little bit different situation has emerged in the case

of set theory. Here both kinds of extremal axioms: minimal as well as maximal have been taken into account. The already mentioned Fraenkel's axiom of restriction (roughly: there are no more sets than those whose existence follows from the axioms of set theory) belongs to the first category. The famous Gödel's axiom of constructibility is also an axiom of restriction, i.e. a minimal extremal axiom (Gödel, 1940). It should be stressed that Gödel himself advocated later a possibility of extending set theory by maximal axioms, in analogy with Hilbert's axiom of completeness. An interesting, but less known proposal of a precise formal explication of the axiom of restriction in set theory is the axiom of canonicity introduced in (Suszko, 1951). The monograph (Fraenkel, Bar-Hillel, Levy, 1973) contains a sharp critique of set theoretical restriction axioms and presents arguments (but to a high degree pragmatical ones!) in favor of maximal axioms in set theory. They are axioms of the existence of large cardinal numbers – cf. e.g. (Kanamori, 1994). It seems that such axioms were for the first time proposed in (Zermelo, 1930) where the author argued that the spirit of set theory requires consideration of a transfinite hierarchy of strongly inaccessible cardinals.

We will also compare the role of extremal axioms with some important isomorphism theorems in algebra (Frobenius, Ostrowski, Hurwitz, Pontriagin) which characterize up to isomorphism some fundamental structures (like real numbers  $\mathbb{R}$ , complex numbers  $\mathbb{C}$ , quaternions  $\mathbb{H}$  and octonions  $\mathbb{O}$ ). Finally, we will point to chosen results in modern model theory related to the dependencies between categoricity and completeness.

The terms *intended model* and *standard model* are sometimes used interchangeably in the literature. However, we would like to propose the following distinction. By the *intended model* we will mean the structure which has been investigated for its own sake, typically for a long time, so that we have collected a large amount of data and have proven many fundamental theorems about it. One could say that intended models are such structures which became *domesticated*, easily accessible cognitively and responsible for the basic mathematical intuitions. One can see that this is only an intuitive characterization of the term *intended model* which, in turn, also is an intuitive term. Natural number series (with arithmetical operations) may serve as an example of intended model understood in this way.

One can talk about the *standard model* only after one has obtained a developed formal theory, ultimately an axiomatic one. In such a case one can establish the class of all possible models of the theory in question. Then the standard model of such a theory is its model most closely related to intended one. It may happen that the standard model of a theory obtains a precise characterization, e.g. in terms of isomorphism or elementary equivalence; for instance, Tennenbaum's theorem says that the standard model of first order Peano arithmetic is the only recursive model of that theory. Examples of standard models in the proposed sense are: the standard model of first order Peano arithmetic, the completely ordered real field  $\mathbb{R}$ , the model of Euclidean geometry (in Hilbert's axiomatization).

A theory may obviously have also *non standard* models which do not resemble the intended model at all or resemble it only to a certain degree. For instance, the first order Peano arithmetic has a lot of non standard models; all its countable

non standard models have the same order type  $\omega + (\omega^* + \omega) \cdot \eta$  but they are differentiated with respect to the arithmetical structure.

In general, standard models can be determined neither syntactically nor semantically. It is our epistemic choice to call a given model the standard one. That choice is of course influenced by the research practice of professional mathematicians – accumulation of the results characterizing the intended model.

The work on this abstract has been sponsored by the National Scientific Center research grant 2015/17/B/HS1/02232 *Extremal axioms: logical, mathematical and cognitive aspects*.

## References

- Awodey, S., Carus, A.W. (2001). Carnap, completeness, and categoricity: the *Gabelbarkeitssatz* of 1928. *Erkenntnis*, 54, 145–172.
- Awodey, S., Reck, E.H. (2002). Completeness and Categoricity. Part I: Nineteenth-century Axiomatics to Twentieth-century Metalogic. *History and Philosophy of Logic*, 23, 1–30.
- Carnap, R. (1930). Bericht über Untersuchungen zur allgemeinen Axiomatik. *Erkenntnis*, 1, 303–307.
- Carnap, R., Bachmann, F. (1936). Über Extremalaxiome. *Erkenntnis*, 6, 166–188.
- Dedekind, R. (1872). *Stetigkeit und irrationale Zahlen*. Braunschweig: Friedr. Vieweg und Sohn.
- Fraenkel, A.A. (1928). *Einleitung in die Mengenlehre*. Berlin: Verlag von Julius Springer.
- Fraenkel, A.A., Bar-Hillel, Y., Levy, A. (1973). *Foundations of set theory*. Amsterdam London: North-Holland Publishing Company.
- Gödel, K. (1940). The consistency of the axiom of choice and of the generalized continuum hypothesis with the axioms of set theory. Princeton: *Annals of Mathematics Studies*, 3.
- Hilbert, D. (1899). *Grundlagen der Geometrie*. Leipzig: B.G. Teubner.
- Hintikka, J. (1986). Extremality Assumptions in the Foundations of Mathematics. *Philosophy of Science Association*, 2, 247–252.
- Kanamori, A. (1994). *The Higher Infinite. Large Cardinals in Set Theory from Their Beginnings*. Berlin: Springer-Verlag.
- Lindenbaum, A., Tarski, A. (1936). Über die Beschränktheit der Ausdrucksmittel deduktiver Theorien. *Ergebnisse eines mathematischen Kolloquiums*, 7, 1934–1935, 15–22.
- Peano, G. (1889). *Arithmetices principia, nova methodo exposita*. Torino: Bocca.



- Scanlan, M. (1991). Who were the American Postulate Theorists? *The Journal of Symbolic Logic*, Volume 56, Number 3, 981–1002.
- Suszko, R. (1951). Canonic axiomatic systems. *Studia Philosophica*, IV, 301–330.
- Tarski, A. (1940). On the Completeness and Categoricity of Deductive Systems. In: Mancosu, P. (2010). *The Adventure of Reason. Interplay between Philosophy of Mathematics and Mathematical Logic, 1900–1940*. Oxford: Oxford University Press, 485–492.
- Zermelo, E. (1930). Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicae*, 16, 29–47.

# Theoretical innovation. A new look at creativity in the natural sciences

Andrés Rivadulla, a. H. Prof.<sup>1</sup>

Department of Logic and Theoretical Philosophy  
Complutense University of Madrid, Spain

The making anything known which was unknown before is an innovation in knowledge: and if all such innovations had been forbidden, men would [not] have made a notable progress in the arts and sciences.  
(George Berkeley, “Third Dialogue between Hylas and Philonous”)

## Introduction

The task of accounting for observations in science is understood in two different ways: either in form of an explanation—and, if possible, a true explanation—or simply trying to save the phenomena. Much of the epistemological debate in the history of Western philosophy has focused on elucidating whether, from a position known as *realist*, science must seek true explanations of the observations, or rather, from an *instrumentalist* position, it must conform to *save the appearances*.

Well, if we were convinced that *true* explanations, *causal* explanations, were not possible, this would not imply that science should be limited to saving phenomena. Even if saving the appearances allowed us to successfully predict and intervene in Nature, it would leave us quite dissatisfied, because science is not only limited to account for what is observed; it also anticipates theoretical novelties susceptible of being tested empirically.

There is indeed a way that recommends that science focus on the search for *theoretical* explanations. These are explanations relative to a theoretical frame but disconnected from any pretension of reference to true causes. Theoretical explanations are contextual.

The case is, however, that, if the philosophy of science considered that scientific activity should only focus on the theoretical explanation of observations, it would be neglecting a no less important part of scientific practice as well: *theoretical innovation*, a new way of thinking and naming *scientific creativity*.

---

<sup>1</sup>Complutense Research Group 930174 and Project FFI2014-52224-P of the Ministry of Economy and Competitiveness of the Kingdom of Spain.

Theoretical innovation is a form of scientific discovery. Therefore each innovation has to comply with the demands of novelty and theoretical value. Although concepts of *scientific discovery* can be as many as thinkers have dealt with the subject, I favour Saso Dzeroski's (2007: 3–4) definition as “the process by which a scientist creates or finds some hitherto unknown knowledge, such as a class of objects, an empirical law, or an explanatory theory (...) A defining aspect of discovery is that the knowledge should be new and previously unknown.”

Theoretical innovations take place primarily at the intra-theoretical level, that is to say, in the context of a theory. The most common way to describe these innovations is as ‘predictions’, although there is a qualitative difference with the predictions that occur in the use of a particular theory. For the *intra-theoretical innovations* suppose the incorporation into the theoretical stock—and without resorting to other theories—of novelties hitherto unknown.

But innovations also occur at the inter-theoretical level making use of *preductive reasoning* (I made a presentation of this concept in my contribution to the *Proceedings of the VII Conference of the SLMFCE*, University of Santiago de Compostela 2012). Revolutionary discoveries in the field of relativistic quantum mechanics, some of which I will present briefly here, are an excellent example of theoretical innovations via *preductive reasoning* in contemporary theoretical physics.

There is a common element to both types of theoretical innovation: intra-theoretical *predictions* and inter-theoretical *predutions*. Both are the result of the application of deductive reasoning to the context of creativity or scientific discovery!

Of course there are other ways for the incorporation of theoretical novelties: induction, abduction and analogy as I have shown in Rivadulla (2008) and elsewhere, but scientific innovation via *predution* is a complete novelty in the methodology of science.

The three main elements of my paper are the following:

1. The first task historically entrusted to science is to account for the facts of experience. After two and a half millennia, it remains a matter of philosophical debate if this task should consist of *explaining* the observations, that is, making the reasons for things explicit, or *saving* the appearances, that is, adjusting the observations to a theory. In the first case, it is still under discussion whether the *scientific explanation* must be *causal* or not.
2. But to question the existence of causal explanations does not mean giving up explanations at all: The alternative is *theoretical explanation*, that is, explanation by reference to theoretical frameworks, which helps to alleviate the tension between the supporters of causal explanation and those of saving the phenomena.
3. There is finally a main component of science to which contemporary philosophy has not paid the attention it deserves. I call it *theoretical innovation* and it consists of devising unexpected, sometimes surprising and in any case useful proposals for the advancement of science and culture.

## Does the Newtonian derivation of Kepler's Laws provide a causal explanation of planetary motions?

According to Herschel (1830: 302) “In this great work [*Principia*, A.R.], Newton shows all the celestial motions known in his time to be consequences of the simple law, that every particle of matter attracts every other particle in the universe with a force proportional to the product of their masses directly, and the square of their mutual distance inversely, and is itself attracted with an equal force.”

For Comte (1998: 393) “we say that the general phenomena of the universe are explained, as far as they can be, by the law of Newtonian gravitation.”

Stuart Mill (1970: 303–304) claimed that “it was very reasonable deemed an essential requisite of any true theory of the causes of the celestial motions that it should lead by deduction to Kepler's laws; which, accordingly, the Newtonian theory did.”

And Einstein (1973: 254) maintained that “Kepler's empirical laws of planetary movement,...,confronted him [Newton, A.R.], and demanded explanation. These laws gave, it is true, a complete answer to the question of how the planets move round the sun ... But these rules do not satisfy the demand for causal explanation.”

Finally Hempel (1965: 173–174) certified that “the principles of Newtonian mechanics...explain certain ‘general facts’, i.e., empirical uniformities such as Kepler's laws of planetary motion; for the latter can be deduced from the former.”

As it seems Newtonian mechanics explains Kepler's empirical laws deductively. For its deduction, it is simply a matter of imagining a binary system consisting of two bodies of masses  $M$  and  $m$  respectively and separated by a distance  $d$ , between which there is a balance  $G_N \frac{M}{d^2} = \frac{4\pi^2}{P^2} d$  between gravitational attraction and centrifugal force, from which Kepler's 3<sup>rd</sup> law  $P^2 = (\frac{4\pi^2}{G_N M}) d^3$  results<sup>2</sup>.

The big question is now: Does this *theoretical explanation* of Kepler's Third Law in the context of Newtonian Celestial Mechanics offer a *causal explanation* of planetary movements? Is Stuart Mill right when he says that this explanation is causal? The answer is: definitely not! Two reasons. First: Kepler's third law is also derivable in the context of general relativity theory (*GRT*) and, paraphrasing Mill, it should be recognized as very reasonable to consider an *essential requirement* of *GRT*, as a true theory of the causes of celestial movements, that it leads *by deduction* to Kepler's laws. Consequently, we would have two different causal explanations, the Newtonian and the Einsteinian, of the same phenomenon.

Second: *NM* and *GRT* are *incompatible* with each other, *both* at the level of their respective fundamental postulates: “the [relativistic] picture of the four-dimensional pseudo-Euclidean Universe contradicts that of the Newtonian three-dimensional Euclidean Universe” (Rivadulla 2016: 529) *and* at the level of the theoretical entities that deny each other: according to *GRT*, gravity is not the

<sup>2</sup>This law also results by combining the value of angular momentum with the area of the ellipse and Kepler's second law.

result of real forces, but of the curvature of space-time; there are no gravitational forces and the objects ‘float’ freely, following the geometry of space-time.

*Two causal explanations incompatible with each other is unbearable for scientific rationality!* Conclusion: It seems reasonable to relinquish to causal explanations in theoretical physics.

## ***Apparentias salvare* and theoretical innovation in science**

Citing Simplicius’s *Commentary*, Duhem (1969: 5) states that Plato posed the following question: “What circular movements, uniform and perfectly regular, must be accepted as a hypothesis so that it is possible to save the appearances presented by the planets?”

Rheticus (1959: 136) attributed to Copernicus “that the motion of the earth could produce most of the appearances in the heavens, or at any rate save them satisfactorily.”

And Cardinal Bellarmino advised Foscarini and Galileo “to content himself with speaking *ex suppositione* and not absolutely, as I have always believed that Copernicus has spoken. Because to say that, supposing that the Earth moves and the Sun is still, all appearances are saved better than with putting the eccentric and epicycles it is very well said and there is no danger, and this is enough for the mathematician” (Cortés Pla 1952: 109).

Finally Duhem (1969: 117) maintained that “the hypotheses of physics are mere mathematical contrivances devised for the purpose of saving the phenomena.” And Bas van Fraassen (1980: 4) affirmed categorically that “the belief involved in accepting a scientific theory is only that it ‘saves the phenomena’, that is, correctly describes what is observable”.

Nonetheless scientific methodology can not be content exclusively with saving phenomena. If it did, it would stop giving full account of scientific practice. Indeed, if science were merely concerned with the observable, we would have to suppose that it lacks the creative capacity, anticipatory of novel theoretical developments, sometimes surprising, generally unsuspected, and always useful for scientific progress. So, my thesis is: *Theoretical innovation is, together with the ability to account for observations, the most important task of science. Any theoretical anticipation must meet the requirements of novelty and theoretical usefulness.*

## **Theoretical predution**

*Predution* (Rivadulla, 2008) is the form of reasoning by virtue of which, by mathematical deduction compatible with dimensional analysis, the combination of previously accepted results—but not necessarily as true—of different theories and/or disciplines of physics, allows the *anticipation* of new theoretical products,

which should be submitted to the trial of experience. Preductive reasoning serves both for *innovation* and *explanation* in the methodology of physics. To illustrate this point of view, I present two examples, one of each case:

- the innovative discovery of anti-matter by Dirac in relativistic quantum mechanics,
- the theoretical explanation of the existence of neutron stars in astrophysics.

### ***An example of preductive discovery: antimatter***

Relativistic quantum mechanics, which, as its name suggests, is a theory resulting from the combination of quantum mechanics with relativity theory, anticipates the existence of antimatter. Its main equations are Klein-Gordon and Dirac. The first one describes free particles without spin, that is  $s = 0$ . The second one describes free particles with spin  $s = 1/2$ , and its non-relativistic limit is the Schrödinger equation, making the physical *fiction*  $c = \infty$ .

To derive these equations, we start from the relativistic equation of the total energy of a particle:

$$E^2 = p^2 c^2 + m^2 c^4$$

If we interpret both members as operators, then we have:

$$E^2 \Phi = (p^2 c^2 + m^2 c^4) \Phi$$

that is:

$$-\hbar^2 \partial_t^2 \Phi(x_i, t) = (-\hbar^2 c^2 \nabla^2 + m^2 c^4) \Phi(x_i, t)$$

which is the equation that Schrödinger and Klein-Gordon propose as an equation of relativistic evolution.

Now, this equation admits exact solutions with negative energy. That is why Paul Dirac looked for a similar equation that excluded the possibility  $E < 0$ . The equation he found<sup>3</sup> authorizes to interpret the absence of electrons as the presence of positrons, and, therefore, anticipates the existence of positive electrons, that is, antimatter. Which is a theoretical innovation of the highest order.

### ***Predictive theoretical explanation. The neutron stars theoretical model***

The derivation of Kepler's law in the context of Newtonian celestial mechanics offers clearly a theoretical explanation of planetary motions.

---

<sup>3</sup> $(i\partial - \mu)\psi(x) = 0$ .

In physics we offer a *theoretical explanation* of an event when the mathematical expression of the event in question is recovered deductively in a given theoretical framework. Susceptible to theoretical explanation are also the theoretical laws and even theories themselves.

Examples of theoretical explanation are the deduction of Balmer's empirical formula in the framework of Bohr's atomic model, or the explanation of Planck's radiation law in Bose-Einstein's quantum statistical mechanics (Rivadulla 2005: 170–172, 175–177), among many others. These explanations are intra-theoretical, that is, they occur within the framework of a single theory. The scientific explanation thus coincides with the derivation of the explanandum in a given theoretical context, following Popper-Hempel's deductive-nomological model.

Contemporary astrophysics seriously questions the idea that science only deals with observables. Indeed, celestial objects and phenomena such as stellar atmospheres and interiors, novae and supernovae, white dwarfs, neutron stars, black holes, etc., whose internal processes, naturally, are not observable, would be left unexplained. To deal with them, astrophysicists build theoretical models. These models result from the combination of accepted results of different theories and disciplines of physics by application of *predictive reasoning*. The explanations that these models provide are, therefore, inter-theoretical explanations. Let us take the example of neutron stars to illustrate this concept.

Neutron stars are, together with black holes and white dwarfs, the most compact objects in the universe; although their origin is very different. For while the first two are remnants of supernova explosions, the white dwarfs are the rest, bright by its internal residual heat, of a small or medium star in the final stage of its life, when it no longer generates nuclear reactions in its interior. White dwarfs usually have a mass comparable to that of the Sun but a size like that of Earth. Its density is much lower than that of a neutron star. Thus, for example, the white dwarf par excellence, Sirius *B*, compared to a typical neutron star, about 10 km in radius, is 550 times larger.

The explanation of the internal structure of a neutron star can only come from a theoretical model that, by means of a *predictive* way of reasoning, by combining results from Newtonian mechanics, relativity theory, quantum mechanics, nuclear physics, particle physics, etc. gives us a picture of this kind of objects.

A neutron star model should incorporate both the history of the physical process of producing a supernova and the theoretical explanation, comparable to that of the white dwarfs, with which it shares being supercompact objects. In the case of white dwarfs, gravity itself is supported by electron degenerate pressure. In the case of neutron stars, its gravity is supported by the pressure of the degenerate neutron gas that is produced inside it.

Indeed, for densities higher than  $10^{12}$  gr.cm<sup>-3</sup> neutron degeneracy pressure dominates electron degeneracy pressure due to the electronic capture process by which protons are transformed into neutrons. The number of electrons is reduced and the number of neutrons increases, and practically the interior of the stars is a superfluid of neutrons plus other elementary particles.

According to Ostlie & Carroll (p. 604) “a very general argument involving the general theory of relativity shows that the maximum mass possible for a neutron

star cannot exceed about  $3 M_{\odot}$ ... If a neutron star's mass exceeds  $3 M_{\odot}$ , it cannot generate pressure quickly enough to avoid collapsing. The result...is a black hole."

As a result of the collapse of the supergiant star that produces the supernova and leaves as a remnant the neutron star, it can rotate very quickly, dragging in each rotation its enormous magnetic field. It becomes a pulsar.

## Conclusion

I have started by asking if scientific explanation must be causal explanation. To answer this question I have presented the argument of inter-theoretical incompatibility between *NM* and *GRT* to conclude that *NM* can not give a causal explanation of the planetary motions, nor does *GRT*.

Next I have wondered if 'saving the phenomena' is a reasonable alternative to the commitment to causal explanation. After concluding that this 'alternative' does not take into account the work of theoretical innovation that the scientific practice involves, I conclude that the theoretical explanation *more preductivo* (which does not pretend to be causal) does seem to do justice to the double task of science: explanation and innovation.

By means of two examples from *RQM* and Astrophysics I have tried to show that explanation and innovation go hand in hand, that both constitute the joint goal of scientific theorization: The *revolutionary innovation* of Dirac offers a *theoretical explanation* of the existence of antiparticles, the *theoretical explanation* of neutron stars 'reveals' the internal structure of this kind of compact bodies, i.e. it is *innovative* as well.

## References

- Comte, A. (1998). *Cours de Philosophie Positive*. Presentation et notes par Michel Serres *et al.* Paris: Hermann.
- Duhem, P. (1969). *To Save the Phenomena. An Essay on the Idea of Physical Theory from Plato to Galileo*. Chicago: University Press.
- Dzeroski, S., et al. (2007). "Computational Discovery of Scientific Knowledge". In S. Dzeroski and L. Todorovski (eds.), *Computational Discovery of Scientific Knowledge*. LNAI 4660, Berlin: Springer.
- Einstein, A. (1973). "The Mechanics of Newton and their Influence on the Development of Theoretical Physics". In A. Einstein, *Ideas and Opinions*. London: Souvenir Press Ltd., pp. 253–261.
- Hempel, C. G. (1965). *Aspects of Scientific Explanation*. New York: Free Press, 2. print.
- Herschel, J. (1830). *A Preliminary Discourse on the Study of Natural Philosophy*. Chicago: Univ. Press, 1987. (Facsimile edition)



- Mill, J. S. (1970). *A System of Logic Ratiocinative and Inductive*. London: Longman.
- Ostlie, D. & Carroll, B. (1996). *Modern Stellar Astrophysics*. Reading, MA: Addison-Wesley Publ. Co., Inc.
- Pla, C. (1952). *Galileo Galilei*. Buenos Aires: Espasa-Calpe Argentina, S.A.
- Rheticus (1959). “Narratio Prima”. In E. Rosen (ed.), *Three Copernican Treatises*. New York: Dover Publications.
- Rivadulla, A. (2005). “Theoretical Explanations in Mathematical Physics”. In G. Boniolo et al. (eds.), *The Role of Mathematics in Physical Sciences*, 161–178. Dordrecht: Springer.
- Rivadulla, A. (2008). “Discovery Practices in Natural Sciences: From Analogy to Predution”. *Revista de Filosofía*, 33(1):117–137.
- Rivadulla, A. (2016). “Models, Representation and Incompatibility. A Contribution to the Epistemological Debate on the Philosophy of Physics”. In J. Redmond et al. (eds.), *Epistemology, Logic and the Impact of Interaction*. Switzerland: Springer, 521–532.
- Van Fraassen, B. (1980). *The Scientific Image*. Oxford: Clarendon Press.

# Diagramatización en física: aproximación semiótica al alcance representativo de los diagramas de Feynman

Carmen Sánchez Ovcharov

## Resumen:

El objetivo de esta comunicación es aportar un punto de vista semiótico al debate acerca de la relación de representación que guardan los diagramas de Feynman con los procesos físicos que se manejan a través de ellos. Se propone abordar esta cuestión desde la noción de “diagrama”, en tanto que signo, y determinar el *modo específico y propio* de la representación diagramática y, solo posteriormente, reflexionar sobre la relación *diagrama de Feynman – proceso físico*. El signo “diagrama” es aquél que guarda una relación de semejanza estructural –isomorfía– con determinadas relaciones entre los componentes de su objeto de representación. En el caso de la herramienta ideada por Feynman, al menos para el caso específico de la interacción electromagnética, la semejanza estructural de los diagramas es con una expresión formal matemática; aquella con la que se calcula la amplitud de probabilidad de darse un proceso de interacción concreto. En consecuencia, la pregunta por la relación de *representación diagramática* entre los diagramas de Feynman, y los procesos físicos, queda trasladada a la pregunta por la relación de *representación simbólica* que guarda la expresión matemática citada con los procesos físicos cuyas probabilidades arroja. Y esta es la cuestión que debe resolverse.

## Palabras clave:

Diagrama, Feynman, signo, isomorfía, estructura, semejanza, representación, representación diagramática.

Existe disparidad de opiniones en torno a cuál es la función representativa (y si la hay) que desempeñan los diagramas de Feynman en aquellos apartados de la física que los emplean como método de simbolización diagramática, como son: la teoría cuántica de campos (para simbolizar aspectos de las interacciones fuerte, débil y electromagnética), la física del estado sólido no relativista (especialmente en el apartado de física de muchos cuerpos) y la física estadística. Se discute si los diagramas de Feynman son representaciones pictóricas, modelos visuales, representaciones científicas, convenciones, leyes o meras herramientas de cálculo. Con el fin de evitar sacar conclusiones a partir de generalizaciones, nos vamos a restringir a las teorías cuánticas de campos, del modelo estándar, y, más específicamente, a la electrodinámica cuántica. En este ámbito concreto, ¿cuáles

son las relaciones epistémicas entre: elementos de la teoría, elementos de los diagramas y elementos constitutivos del proceso físico?

Para avanzar en esta cuestión y poder valorar algunas de las interpretaciones citadas, la propuesta de esta comunicación refleja la necesidad de centrarnos primeramente el aspecto semiótico: el significado e implicaciones sémicas del término *diagrama* y, solo posteriormente, investigar las relaciones que pueda tener un diagrama concreto (p.e., un diagrama de Feynman) con aquello que representa diagramáticamente. Tomemos algunas definiciones:

“Un diagrama es un conjunto de objetos en el plano que denotan objetos en una situación [relación, estructural], cuyas mutuas relaciones espaciales y gráficas denotan relaciones en aquella estructura.” (Lemon & Pratt, 1997).

“Un diagrama es un ícono [un signo que guarda relación de semejanza con el objeto de su representación] de un conjunto de objetos racionalmente relacionados. [...] El diagrama no solo representa los correlatos vinculados, sino también, y de manera mucho más definida, la relación entre ellos.” (Peirce, NEM IV: 316)

Podemos ver que la peculiaridad de la relación que un diagrama guarda con lo representado (lo simbolizado diagramáticamente) es que se trata una relación de *semejanza estructural*, una semejanza entre relaciones exclusivamente. Por lo tanto, las relaciones que guardan entre sí los objetos del diagrama representan relaciones externas al diagrama. En este sentido, los diagramas se pueden considerar como *isomorfos* con su objeto de representación. Teniendo en cuenta esto, la función representativa del diagrama es *hacer visible* (permitir visualizar) o *mostrar* la estructura de las relaciones que guardan los componentes de su objeto; una estructura que suele quedar oculta o diluida en representaciones de otro tipo (comúnmente de carácter secuencial), como las lingüísticas o las formales (matemáticas).

Ahora podemos reformular nuestra cuestión: cuando nos planteamos cuál es función y alcance representativos de los diagramas de Feynman, lo que queremos saber es con qué son *isomorfos* los diagramas, esto es, con qué guardan una relación de semejanza estructural.

Originariamente, Feynman presentó, en 1949, su método diagramático, precisamente, como una técnica para organizar visualmente los *componentes matemáticos* de la expresión formal de la amplitud mecánico-cuántica de un proceso electrodinámico dado (en el marco de la teoría de las perturbaciones (TP)), permitiendo así mostrar o hacer visible la estructura de las relaciones entre los componentes de la citada expresión. Para llegar a esta última son necesarios cuatro pasos de teorización (probablemente, además, de *modelización*): *modelo estándar de partículas e interacciones*, *teoría cuántica de campos*, *electrodinámica cuántica* y *teoría de las perturbaciones*. Sin analizar los componentes de estas teorías y modelos no podremos desentrañar el alcance representativo de los diagramas de Feynman, puesto que no sabemos con qué establecer la relación de semejanza estructural. Es más, presentar los componentes y las reglas de los diagramas, y discutir su función epistémica carece de sentido sin establecer el marco teórico en el que se construyen las expresiones matemáticas cuyas relaciones estructurales vamos a diagramatizar. En consecuencia, acabamos de precisar

algo más nuestra cuestión inicial: en la formulación originaria de Feynman, los diagramas guardaban una relación de semejanza estructural con las relaciones entre los componentes matemáticos del cálculo de la probabilidad total del proceso de interacción estudiado, en el marco de la teoría de las perturbaciones. Consecuentemente, es sostenible afirmar que la isomorfía se da con la estructura de la representación formal (simbólica) de esa teoría y no directamente con los elementos del proceso de interacción estudiado.

Cada diagrama de Feynman simboliza *una* de las formas posibles del proceso de interacción y, como a cada forma le corresponde una probabilidad –un valor–, a cada diagrama le corresponde *un valor de probabilidad*. Así, cada diagrama de Feynman, en sí mismo, no tiene significado físico –no representa un proceso– más allá de una determinada contribución a la probabilidad total del proceso estudiado. Para obtener la probabilidad total del proceso de interacción se deben sumar los valores de cada *contribución* (cada forma posible de tener lugar el proceso) y el resultado de esa suma se eleva al cuadrado. Esta suma es una *serie*, en la que algunas formas posibles tienen mayor probabilidad de ocurrir que otras y, además, la serie es, potencialmente, infinita. Para el tratamiento de la serie se utiliza el modelo de la teoría de las perturbaciones.

Se puede mostrar que las partes integrantes de los diagramas y su estructura completa guardan una relación isomórfica con las partes integrantes de la expresión formal que permite calcular la probabilidad total del proceso estudiado. De estas consideraciones, la cuestión que llevamos planteándonos desde el inicio queda redirigida hacia otra cuestión, que es la que debería debatirse: ¿con qué es isomórfica *la estructura de la expresión formal* que los diagramas de Feynman explicitan diagramáticamente?

## Bibliografía

- Kaiser, D. (2009). *Drawing theories apart: The dispersion of Feynman diagrams in postwar physics*. University of Chicago Press.
- Legris, J. (2012). Visualizar y manipular. Sobre el razonamiento diagramático y la naturaleza de la deducción. *Aparecerá en Visualização nas Ciências Formais., comp. por Abel Lassalle Casanave & Frank Thomas Sautter. Londres, College Publications*, 89–103.
- Lemon, O., & Pratt, I. (1997). Spatial logic and the complexity of diagrammatic reasoning. *Machine Graphics and Vision*, 6(1), 89–108.
- Peirce, C. S., & Eisele, C. (1976). *The new elements of mathematics* (Vol. 4). The Hague: Mouton.
- Stöltzner, M. (2017) Feynman diagrams as models. *The mathematical Intelligencer*. Springer Science & Business Media New York DOI 10.1007/s00283-017-9716-z, pp. 46–54.
- Wüthrich, A. (2010). *The genesis of Feynman diagrams* (Vol. 26). Springer Science & Business Media.

Wüthrich, A. (2012). Interpreting Feynman diagrams as visual models. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, 6(1), 172–181.

# The creative similarity account of the *means* of representation

Julia Sánchez-Dorado

## Short abstract

The concept of similarity, if adequately characterized, can provide an account of the epistemic success of scientific representation. In this paper, I first discuss the difference between an account of the “constituents” and the “means” of representation (Suárez 2003; van Fraassen 2008). Then, I examine—and reject—the possibility of taking similarity (and its structural variants isomorphism and homomorphism) as constituents of representation (like Bartels 2006; French 2006). Next, I explore if similarity can help account for how epistemically successful representations are produced in scientific practice, that is, if similarity is a central means of representation. Three arguments against similarity as means of representation will be considered: the argument from variety, from vagueness, and from misrepresentation. None of these arguments, I claim, is conclusive enough to prevent us from trying to offer such an account. Finally, I advance my own specific account, the “creative similarity account of the means of representation”. This account has three important assets: it can adequately avoid the criticisms against similarity previously presented; it is sustained on a practical investigation of actual uses of judgments of similarity in modelling practices; and it makes similarity and relevant distortions and idealizations compatible with each other.

In this paper, I will argue that the concept of similarity, if adequately characterized, can help provide an account of the epistemic success of scientific representation. More specifically, I will claim that although it is problematic to consider similarity (and its structural variants isomorphism, partial isomorphism, homomorphism) *constituents* of representation, similarity can be considered an effective *means* of representation. First, I spell out the different implications of advancing an account of respectively the constituents and of the means of representation. Then, I examine the possibility of considering similarity a constituent of representation, and present two compelling arguments that should dissuade us from doing so, namely Goodman’s (1968) logical argument against similarity and the argument from misrepresentation. Next, I discuss the possibility of considering similarity the means (or at least a pervasive means) of representation. That is to say, I explore if referring to the notion of similarity can help account for how a relevant set of representations that we recognize for their epistemic achievements are produced in science. Three arguments against similarity as a means of representation will be considered: the “argument from variety”, the “argument from vagueness”, and the “argument from misrepresentation against the means”. The conclusion this time will be different: none of these arguments

is conclusive enough to prevent us from attempting to offer an account of the means of representation focused on similarity. Lastly, I outline a specific account of the means of representation focused on similarity, the “creative similarity account of the means of representation”. This account has three important assets: it can adequately avoid the criticisms against similarity previously presented; it is sustained on a practical investigation of actual uses of *judgments of similarity* in scientific modelling; and it defines similarity as compatible with the role of relevant distortions and idealizations in the practice of producing scientific representations.

## Two questions about representation

There is general agreement in the contemporary debate in philosophy of science that two distinct questions are involved in the so-called “problem of scientific representation”: the question about the *constituents* and the question about the *means* of representation (Suárez 2003; Contessa 2011; van Fraassen 2008). The question about the constituents asks “in virtue of what is something a representation of a target in the world?”, while the question about the means asks “in virtue of what is something an accurate or epistemically successful representation of a target in the world?”. The question about the constituents seems to demand a set of necessary and sufficient conditions to be answered. For instance, some philosophers have referred to a relation of denotation, isomorphism, homomorphism, or the combination of various of these relations to explicate what constitutes representation (French 2003; Bartels 2006; French and Bueno 2011). Meanwhile, the question about the means of representation has not been systematically discussed in the debate in philosophy of science, and it is still unclear what a satisfactory response to it requires. In this paper, I attempt to offer some clarification about the requirements of an account of the means of representation, in particular by advancing a particular account, the “creative similarity account”, and exploring its limitations.

## Similarity as means of representation

The focus of this paper is on similarity as possible means of representation. Nonetheless, I start by introducing two common arguments against similarity as constituent of representation, Goodman’s (1968, 1972) logical argument against similarity and the argument from misrepresentation, to reinforce the idea that if the role of similarity is to be rehabilitated in the debate, it should be done regarding the epistemic success of scientific models and not the mere conditions for representation.

An account of the means of representation focused on similarity would sustain a general claim like the following: similarity is the relation (or a common relation) actively employed by epistemic agents to perform successful inferences from a model about a target in the world (Suárez 2003). There are, however, three important arguments that should make us doubt whether similarity can actually be a good candidate for means of representation.

The *argument from variety* claims that, given the huge diversity of models there are, and the different ways in which epistemic agents can employ relations between vehicle and target to make inferences, similarity could be at best one among many other means of representation (Frigg 2006; Suárez 2003). To this I reply that similarity might not be the only means of representation. But this should not stop the project of attempting to develop a general type of account focused on similarity if there is evidence that, in an epistemically relevant set of representational practices, similarity works as an effective means of representation.

The *argument from vagueness* claims that similarity is an empty or utterly vague concept, given that its meaning should be specified in every particular case and context (Chakravartty 2001; Frigg 2006). To this argument I reply that having an absolute, scientifically objective definition of similarity doesn't seem a reachable goal indeed, but it should not be a desirable one either (following Giere 2004). The focus of attention should be the study of how similarity actually works in practices of representing, and how standards and general agreement about relevant similarities are established.

The *argument from misrepresentation against the means* claims that, since scientific models distort, idealize, and abstract the targets they represent, similarity doesn't seem a good candidate for means of representation. I reply to this argument that similarity and distortion, as they actually intervene in practices of representing, are neither necessarily in conflict with each other, nor pulling in different directions, but can be intertwined and complementary resources for the construction of fruitful representations.

## The creative similarity account of the means of representation

Proposing the “creative similarity account of the means of representation” has two purposes. First, it is an attempt to accommodate the role of similarity in scientific representation in a satisfactory way. Even after responding to the aforementioned arguments against similarity, it is important to emphasize that not any characterization of similarity would be adequate to sustain an account of the means of representation. An adequate characterization of similarity, I argue, has to emerge from actual practices of representing and help capture scientists' use of *judgments of similarity* when constructing successful representations. The notion of “creative similarity” can help achieve this. I characterize “creative similarity” in opposition to “bare similarity”, which is a conception of similarity that recalls a two-term relation of representation, of sharing of properties between a model and a target, and conceals an aspiration to obtain perfect copies of the object represented. Contrary to this, the notion of “creative similarity” has to be framed in a triadic conception of representation, where resourceful epistemic agents define to a great extent what relevant similarity are, depending of the goals of the representational practice. In addition, I characterize creative similarity as similarity that is compatible and combinable with distortions, based on the empirical observation that when *judgments of similarity* are used



in modelling practices, they go hand in hand with judgments of dissimilarity and distortion (i.e. simplifications, idealizations, abstractions, generalizations). Finally, I argue that what ultimately makes similarity creative is that, from its combination with relevant distortions in the practice of modelling in science, new insightful associations arise (following Cartwright 1999; Boden 2004; Elgin 2017). The use of various examples of representational practices, such as the construction of scale models in engineering, and research with model organisms in the life science will help illustrate the implications of the “creative similarity account” (Weisberg 2013; Ankeny et al. 2014).

The second purpose of advancing the “creative similarity account” is to help elucidate what exactly developing a satisfactory account of the means of representation requires. I present four desiderata that any account of the means of representation should fulfil as a result of this exploration. The first desideratum is that an account of the means of representation should develop a practical enquiry on particular modelling practices. The second desideratum is that the account should incorporate a normative component, as any genuine philosophical proposal, in connection to the descriptive component that results from the practical enquiry. The study of judgments concerning model-target comparisons in scientific practices is the key to connect the normative and descriptive components. In the case of the “creative similarity account”, specific attention is given to the analysis of scientists’ formulation of *judgments of similarity* in modelling practices. The third desideratum is that an account of the means of representation should specify how ‘epistemic success’ is understood in the account (as accuracy, truthfulness, fruitfulness, usefulness, precision), and accordingly how certain means lead to the achievement of successful representations characterized in that particular way. And the fourth desideratum is that an account of the means of representation should explicitly accommodate the potential value that idealizations and other forms of distortions might play in the epistemic success of scientific representations. The “creative similarity account” that I propose fulfils these desiderata. Alternative accounts of the means of representation, for instance focused on convention, exemplification, or isomorphism as means of representation, may be advanced, provided they fulfil them as satisfactorily as the creative similarity account.

## References

- ANKENY, R., LEONELLI, S., NELSON, N., and RAMSDEN, E. (2014) “Making Organisms Model Human Behavior: Situated Models in North-American Alcohol Research, since 1950”. *Science in Context*, Volume 27, Issue 03, September 2014.
- BARTELS, A. (2006) “Defending the Structural Concept of Representation”. *Theoria*, 21:7–19.
- CHAKRAVARTTY, A. (2001) “The semantic or model-theoretic view of theories and scientific realism”. *Synthese*, 127:325–345.
- CONTESSA, G. (2011) “Scientific models and representation”. In FRENCH,

- S. and SAATSI, J. (ed.) *The Continuum Companion to the Philosophy of Science*. London: Continuum Press.
- FRENCH, S. (2003) “A Model Theoretic Account of Representation (or, I don’t know much about art, but I know it involves representation)”. *Philosophy of Science*, 70:1472–83.
- FRENCH, S. and BUENO, O. (2011) “How Theories Represent”. *British Journal for the Philosophy of Science*, 62:857–94.
- FRIGG, R. (2006) “Scientific Representation and the Semantic View of Theories”. *Theoria*, 55:37–53.
- GIERE, R. (2004) “How Models Are Used to Represent Reality”. *Philosophy of Science*, 71, Dec. 2004.
- GOODMAN, N. (1968) *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis: The Bobbs-Merrill Company.
- SUÁREZ, M. (2003) “Scientific representation: against similarity and isomorphism”. *Philosophy of Science*, 17(3):225–244.
- VAN FRAASSEN, B. (2008) *Scientific Representation. Paradoxes of Perspective*. Oxford and New York: Oxford University Press, 2008.
- WEISBERG, M. (2013) *Simulation and Similarity*. Oxford University Press, 2013.

# A propensionist interpretation of Lewis structures

**Pedro J. Sánchez Gómez**

Dpto. Didáctica de las Ciencias Experimentales, Sociales y Matemáticas  
Universidad Complutense de Madrid  
pedros@edu.ucm.es

Lewis structures (LS's) play a crucial role in the chemical thought. LS's are schematic descriptions of the frame of a molecule in terms of the valence electrons of the atoms that make it up. These electrons are arranged in pairs, both bonding and non-bonding, according to a rather limited set of simple rules. Surprisingly enough, given the central role they play in the chemical thought, there is no consensus about the epistemological status of LS's. On the one hand, some authors, chiefly Gillespie (1991, 1996; see also Purser 1999), have considered that LS's are means for determining the three-dimensional structure of molecules. In this view, LS's must be seen as schematic planar projection of the actual molecular structure. On the other hand, back in the 1930's Linus Pauling introduced the idea that LS's are a means for defining a *best approximation* for the wave function of a molecules within an independent-electrons model (Pauling 1939). This approach is usually dubbed as the *Valence Bond (VB) method*. In this perspective, LS's have primarily to do with the electronic structure of the molecule.

In my opinion, these interpretations are both rather partial. Neither they permit to understand LS's in terms of some more basic indisputable principles nor they account for the central role they play in the chemical explanation. In this communication I present an alternative interpretive framework, grounded, on the one hand, on the probabilistic character of chemical equations, and, on the other, on a propensionist vision of these probabilities based on the idea of *mechanism of reaction*.

## Chemical propensions: reactivity

The idea of chemical reactivity implies a probabilistic vision of chemical reactions. For example, an expression such as " $A + B \rightarrow C$ " does not mean that under the right conditions some given amounts of the substances A and B react to give the mass of C that theoretically would correspond to the stoichiometry of the reaction. Instead, what is produced is a fraction, let us call it  $\alpha$ , of this theoretical amount. The percentage  $\alpha$  is usually known as the *yield* of the reaction. An  $\alpha < 1$  can be attributed to two different causes:

- 1) Impurities in the reactants, or side reactions with other substances in the environment. In other words, the assumed reactivity of A or of B, or of

both, is too limited, and thus “A + B” is only an approximated description of the initial state.

- 2) In general, given a group of reactants there is more than a possible reaction between them. Thus, in our model example, “A + B → C” does not fully describe the reactivity of A and B. It would be more accurate to write:



...

And so on to exhaust all the possible reactions between A and B. Thus, the yield  $\alpha$  for a given product also expresses the relative preponderance of its reaction.

My point is that these two acceptations of the reaction yield can be accounted for by means of a propensionist interpretation of LS. In this view, LS's are instruments to explore possible reaction paths and to determine their relative weights.

Let us see it in a standard organic chemistry example, the reactions of electrophilic aromatic substitution. In these reactions an electrophile, i.e., a substance that can accept a pair of electrons to form a covalent bond, reacts with an aromatic ring (**Fig. 1**).  $A^+$  is the electrophile (typically ethyl or an halogen atom) and B its associated nucleophile. The rearrangement of electronic pairs are shown with red curved arrows that indicate the origin and the target of the pair. The reactivity of the benzene is explained by means of a feature of its LS, namely its having electronic pairs that can be removed without altering the molecular frame. On the other hand, the electrophile is characterized as having vacant orbitals to accept a pair.

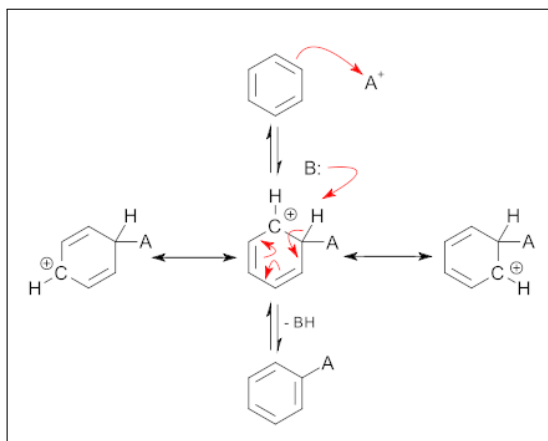


**Figure 1: Electrophilic Aromatic Substitution.**

License notice for the image: V8rik at English Wikipedia (<https://commons.wikimedia.org/wiki/File:BasicAromaticSubstitution.svg>), “BasicAromaticSubstitution”, <https://creativecommons.org/licenses/by-sa/3.0/legalcode>

This line of reasoning also have some predictive value. For example, it can be used to study the influence of a substituent group on the relative yields when more than a product is generated. For example, an *amino* group ( $-\text{NH}_2$ ) directs the reaction to the *ortho* and *para* products, by means of a resonant stabilization of the cationic aromatic intermediate (**Fig. 2**).

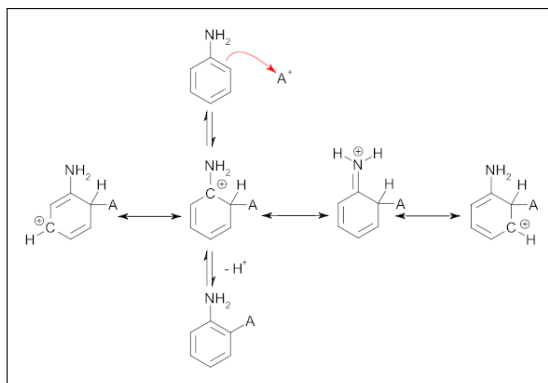


**Figure 2: Ortho electrophilic substitution on aniline.**

License notice for the image : V8rik at English Wikipedia (<https://commons.wikimedia.org/wiki/File:ElectrophilicAromaticSubstitutionOrthoDirectors.svg>), "ElectrophilicAromaticSubstitutionOrthoDirectors", <https://creativecommons.org/licenses/by-sa/3.0/legalcode>

On the other hand, in the reaction on the meta position the positive charge cannot be delocalized over the amino group, thus being less stable (**Fig. 3**).

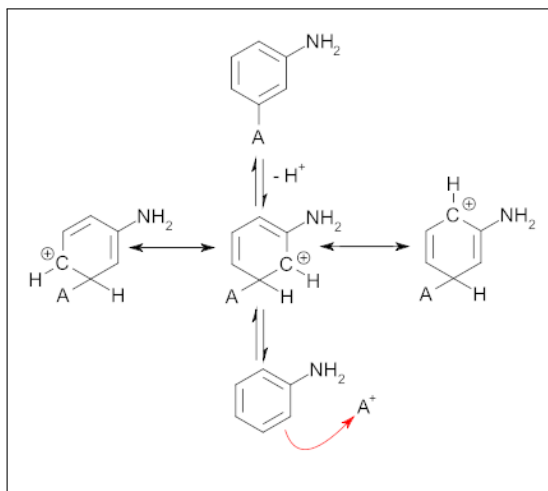


**Figure 3: Meta electrophilic substitution on aniline.**

License notice for the image : V8rik at English Wikipedia ([https://commons.wikimedia.org/wiki/File:EAS\\_meta\\_deactivation.svg](https://commons.wikimedia.org/wiki/File:EAS_meta_deactivation.svg)), "EAS meta deactivation", <https://creativecommons.org/licenses/by-sa/3.0/legalcode>

As a conclusion, the yields for the ortho and para substitutions are expected to be similar and substantially greater than that for the meta product. Such a prediction can be extended to any *activating* groups, and has been being

experimentally confirmed for more than a century.

Similar arguments can be found in any introductory textbook for any organic reaction. Moreover, a quick look on the (countless) current research publications on chemistry shows that this explanatory scheme is far from being a mere didactic introductory aid with didactic purposes.

In sum, LS's, together with a rather restricted set of rules to *operate* them, permit to explore the reactivity of a substance. In particular, the schematic chemical equation that links reactants to products, such as " $A + B \rightarrow C$ ", can be analyzed into a series of steps by means of some elementary transformations of LS's. The predictions that can be made on the basis of a mechanistic reasoning are somehow restricted, and in particular they are not quantitative, but they have a sound epistemic value. By describing the mechanism of a reaction as a sequence of consecutive LS's, the reactivity of a substance can be understood in terms of some internal properties of it. The yield of a particular reaction is thus explained in terms of dispositions of the reactants. And the tendency of some reactants to have a given reaction, and not others, has to do with formal properties of the LS's of the intermediate species that play a role in the mechanism of the reaction.

## Conclusions

As a conclusion, the nuclear role of Lewis Structures lies in that they provide a propensionist account of reactivity by means of a visual reasoning that is typical of the chemical explanation (Barradas-Solas and Sánchez Gómez, 2014). LS's provide all the theoretical implements that enable chemical thought. Ideas such as those of (single or multiple) bonding, lone pairs, (full or vacant) orbitals, resonance, electrophilicity and nucleophilicity, and so on, idiosyncratic of the chemical thought, cannot be understood outside the framework set by Lewis' structures and rules. It is thus within the framework of chemical reactivity that LS's must be studied and, in particular, where their epistemic status must be defined.

## Acknowledgments

This work has been supported by the Spanish Ministerio de Educación, Cultura y Deporte, through the research project FFI2014-57064-P. The author is member of the research group Methods of Causal Inference and Scientific Representation (MCISR), based at the Complutense University of Madrid.

## References

- Barradas-Solas, F. & Sánchez Gómez, P.J. (2014). Orbitals in chemical education. An analysis through their graphical representations. *Chemical Education Research and Practice*, 15, 311–319.

- Gillespie, R. J (1991). What is wrong with the general chemistry course? *Journal of Chemical Education*, 68, 192–194.
- Gillespie, R. J., Spencer, J. N., & Moog, R. S. (1996). Bonding and molecular geometry without orbitals – The electron domain model. *Journal of Chemical Education*, 73, 622–627.
- Pauling, L. (1939). *The nature of the chemical bond and the structure of molecules and crystals. An introduction to modern structural chemistry*. Ithaca, New York: Cornell University Press.
- Purser, G. H. (1999). Lewis Structures Are Models for Predicting Molecular Structure, Not Electronic Structure. *Journal of Chemical Education*, 76, 1013.

# Computer simulations, emergence, and scientific representation<sup>1</sup>

**Iñaki San Pedro**

Department of Logic and Philosophy of Science  
University of the Basque Country, UPV/EHU

This paper explores the challenges that computer simulations —and more especially complex systems which involve the emergence of new, unknown or unexpected properties— pose to our understanding of the notion of scientific representation, broadly understood, i.e. as an activity carried by scientist aimed at acquiring knowledge about certain aspects of the world.

Throughout the paper computer simulation shall be understood in a broad sense. This includes, following Winsberg (2010), not only the specific routine run on a computer but the whole process of choosing a specific set of theoretical models, implementing those into a computer algorithm, as well as analysing and visualising the simulation output. Arguably, computer simulations so understood entail some particular characteristic epistemic features. For instance, as claimed in Winsberg (2010) the epistemology of computer simulation is downward, meaning that it aims at justifying inferences from theory to experiment, motley since simulations typically include a diversity of sources of input knowledge and autonomous as regards the simulation validation and verification. Furthermore, computer simulations typically involve emergence, i.e. properties of the system which are not present in the original input but that the simulation output displays Humphreys (2009).

In view of such epistemic particularities, one can address the question as to what computer simulations tell us about scientific representation? Put it differently, what account of scientific representation best fits (computer) simulations of complex systems involving emergence?

It is difficult to see, for instance, how accounts of scientific representation based on similarity or structure preserving maps, e.g. isomorphisms or homomorphisms, can accommodate most of the epistemic features above because they lack the flexibility computer simulations demand in this respect. In fact, computer simulations seem to demand less strict notions of representation such as in most recent deflationary or pragmatic approaches Suárez (2004); van Fraassen (2008); Giere (2010). These accounts typically make room, either explicitly or implicitly, for pragmatic pluralism which makes easier to account for the epistemological features attributed to computer simulations. The role of the agent, for instance, becomes central in this notion of computer simulation and therefore notions such as ‘intention’, ‘perspective’, ‘purpose’, etc., become now relevant

---

<sup>1</sup>Financial support from the Spanish Ministry of Economy and Competitiveness (MINECO), project FF12014-60403-JIN, and the Basque Government’s Department of Education, Language Policy and Culture, grant IT644-13, is gratefully acknowledged.



van Fraassen (2008); Giere (2010), and the epistemic value of representations is highly related to their capacity to make sound inferences about the target system—with such and such purpose, intention, perspective, etc., in mind Suárez (2015).

When it comes to emergence, however, neither structure-preserving accounts of scientific representation, nor deflationary pragmatic approaches seem to easily account for it. In the first place, the fact that a target system features emergent properties suggest that any possible strong correspondence between model and target system is broken. This speaks against any account of scientific representation that hinges on some sort of ontological correspondence, be it in terms of either mirroring, isomorphism or homomorphism relations. Secondly, emergence somehow disrupts the capacity of making inferences we may want to attribute to a model. The argument notes that emergent properties arise a posteriori and only after intervention such as for instance experimental manipulations, and cannot in general be foreseen or predicted. Thus models of target systems that feature emergent properties provide an example of scientific representations which cannot be used to make inferences about those targets in a clear obvious way. This, once more, speaks against most pragmatic accounts, which ultimately hinge on the capacity of the model to make sound inferences about the target system.

Thus I argue that the emerging of properties demands an even deeper revision of the epistemological role and status of representations. In particular, I claim that if we are to account for emergence in scientific representational practices we must require that they have, e.g. following Ibarra and Mormann (2006), a further performative or constitutive epistemic role. In other words, I claim that scientific representation of emergent properties need not only guarantee the possibility of some degree of pragmatic pluralism but it also involves epistemic pluralism. Emergence, thus, as exemplified by computer simulations requires a shift of focus when aiming to characterise scientific representation, from objects (i.e. model and target system elements) to their relations.

## References

- Giere, R. (2010). An agent-based conception of models and scientific representation. *Synthese* 172, 269–281.
- Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese* 169, 615–626.
- Ibarra, A. and T. Mormann (2006). Scientific theories as intervening representations. *Theoria* 55, 21–38.
- Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science (Supplement)* 71, 767–779.
- Suárez, M. (2015). Deflationary representation, inference, and practice. *Studies in History and Philosophy of Science* 49, 36–47.

van Fraassen, B. C. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford: Oxford University Press.

Winsberg, E. (2010). *Science in the Age of Computer Simulation*. Chicago: University of Chicago Press.

# Usefulness (or not) of defining the concept of life in biology

Jaime Soler

The interest to find a definition of the concept of life has been increasing since the late twentieth century. This is due, largely, to the development of new disciplines where finding this definition is important, both from a theoretical perspective and from scientific practice. These disciplines include astrobiology, synthetic biology or artificial life, among other.

However, the high number of proposals has not led into a minimum consensus. Far from it, the only point of agreement is the complexity of the problem, as well as the diversity of approaches, often incompatible among themselves. Although this situation is not a conclusive evidence in favour of skepticism about the possibility of defining life, this problem has led some authors to rethink the possibility and usefulness of such a definition.

Taking into account these considerations, what is intended in this paper is to analyze the various objections to the definition of life as a concept. It is then intended to demonstrate that the reasons for this skepticism are based on an essentialist concept of life. However, there are non-essentialist ways of understanding life that are not affected by such criticisms. Finally, the aim is to show the usefulness of an instrumental and plural concept of life.

## Objections to the concept of life

Although not all authors who are skeptical of the concept of life use the same arguments, they all share two basic ideas. First of all, that the proliferation of theories about life poses a situation, at least, problematic. Secondly, they all presuppose explicitly or implicitly that defining the concept of life implies finding the essential nucleus of it. Defining a concept is equivalent to finding a set of necessary and sufficient properties. That is, life is considered as a natural kind in an essentialist sense. However, in the case of life, finding those properties does not seem to be feasible. It is possible to find these or similar ideas in Kelller (2002), Tirard, Morange and Lazcano (2010), Machery (2012) and Cleland (2012).

Keller's criticisms do not refer so much to the possibility of defining life, but to the type of entity that is concerned. Depending on this, the definition will respond to the necessary real properties or to a simple convention. Keller considers that the current situation respecting definitions shows that life is not a natural kind, but a conventional one, defined by the interests and values of human

beings. This would explain why there is no minimum consensus on necessary properties. The definition of life becomes a conventional issue.

Tirard, Morange and Lazcano (2010) try to move the focus of attention. If defining the concept of life has not yielded fruitful results, it may be more interesting to shift the question to how the associated characteristics of life are acquired. This means treating life as an empirical and historical concept whose meaning is given by a conceptual framework, rather than by a precise definition.

Machery (2010) unequivocally points out the futility of dealing with the concept of life. From his point of view, life can be understood as a folk concept or as a scientific concept. However, modern research on folk concepts does not consider them as definitions, so if life were a folk concept, it would not be definable. Therefore, to be definable it should be considered as a scientific concept. The problem is that each discipline considers the concept of life in a different way. If the different disciplines do not agree on what is the essential property (or properties) of life, a general definition cannot be found.

On the other hand, Cleland (2012) expresses its reluctance considering two different objections. The first is the confusion between the linguistic plane and the epistemic one. In second place, because of our limitations on the knowledge of different types of life. With regard to the first criticism, Cleland considers that definitions are essentially linguistic entities. As such, defining a term is a matter determined by the community of speakers. However, research on life does not study its use as a linguistic term, but as real life. From a rather exhaustive study of the theories of reference, Cleland assumes the posture defended by semantic externalism. This leads her to defend that concepts as water and life are not stipulations. In addition, Cleland considers that definitions of these type of concepts can only make sense within a broader theoretical framework.

The second criticism of Cleland is given by our limited knowledge of different types of life. If as it seems according to different theories, all current organisms are descended from a common ancestor (LUCA), then we only know one type of life. Defining a kind from only one example results, at the very least, risky. The situation is similar to define mammals by only knowing zebras. Although the stripes are common to all of them, the presence of mammary glands is much more determinant, even if only one half of the individuals possesses them.

## **Defining life as an instrumental, plural and reversible concept**

As can be seen, criticisms and arguments to reject a possible definition are very different in distinct authors. But all of them assume that, in order to define life, necessary and sufficient properties are needed.

Bich and Green (2017), in a detailed study of Machery's and Cleland's criticisms indicate that such criticisms are correct if definitions are considered in an essentialist sense. But the real situation shows that scientists use definitions in an instrumental way. Bich and Green associate essences with natural kinds, and therefore consider that what Cleland and Machery show is that life is not

a natural kind. But this does not exclude that life could be understood and defined in an instrumental sense.

However, as Diéguez (2008) shows, there are non-essential ways of considering life as a natural kind. In addition, these ways of understanding natural kinds are compatible with the instrumental character of definitions. In theories like those of Boyd (1999) or Dupré (1981) natural kinds can respond to certain interests. Therefore the objections of Bich and Green can be assumed without denying the possibility that life is a natural kind. The only thing that seems to be proven is that life is not a natural kind in an essentialist sense.

Unlike what happened with other concepts, such as the concept species, authors who consider life as a natural kind do so from a rather limited perspective. This is because they consider life as a natural kind in an essentialist sense, with a strong metaphysical commitment. As noted, Diéguez is an exception, as he opens the debate to new possibilities. In addition, there are other alternatives that have not been explored. For example, life can be considered as a promiscuous natural kind. This is the assumption done by Dupré and others in the concept of species, but that has not been addressed in the case of the concept of life. As life considered as a natural HPC, or as an instrumental concept, understanding life as a promiscuous natural kind allows us to overcome the difficulties pointed out by Keller, Machery or Cleland.

At the point where criticism of the concept of life does not seem sufficiently argued, one wonders whether a definition of life is useful for scientific research. To answer this question, two examples of scientific research are going to be analyzed. The first one is Luisi research, studied by Bich and Green. The second refers to the detection of life on the part of the Viking ship, as it appears in Benner (2010).

Bich and Green show how Luisi starts his research with an autopoietic conception of life. The important question is that this conception served as a guide for its subsequent research. Based on the theoretical model, the researchers designed experiments that improved the understanding of different aspects of the cell membrane. This introduced changes in the theoretical model, allowing the development of new lines of research.

Benner's analysis of the experiments carried out by the Viking ship is especially significant due to the way in which the results were interpreted. He points out that the results were considered negative, although there were experiments with both negative and positive results. It seems clear that in this case the negative evidence had a greater weight. Such evidence was linked to certain assumptions not always explicit. This shows an important aspect, which is that, even when explicit definitions are not formulated, researchers have certain beliefs about the concepts they work with.

Therefore, the importance of a possible definition of life does not lie in finding necessary and sufficient properties. If this were the case, criticisms done by authors as Keller or Cleland would be right. However, it is possible to understand definitions in an instrumental, plural and revisable sense. To consider life in this sense offers two interesting advantages: a) It establishes a theoretical revisable framework of reflection on which to develop scientific praxis, and b) It avoids decisions based on implicit elements that are not sufficiently justified.

## References

- Benner, S. A. (2010). Defining life. *Astrobiology*, 10(10):1021.
- Bich, L., & Green, S. (2017). Is defining life pointless? operational definitions at the frontiers of biology. *Synthese*, doi:10.1007/s11229-017-1397-9
- Boyd, R. (1999). Homeostasis, species, and higher taxa. In R.A. Wilson (Ed.), *Species: New Interdisciplinary Essays* (pp. 141–186). Cambridge: MIT press.
- Cleland, C. E. (2012; 2011). Life without definitions. *Synthese*, 185(1):125–144. doi:10.1007/s11229-011-9879-7
- Diéguez Lucena, A. J. (2008). ¿Es la vida un género natural?: Dificultades para lograr una definición del concepto de vida. *Artefactos*, 1:81–100.
- Dupré, J. (1981). Natural kinds and biological taxa. *The Philosophical Review*, 90(1):66–90. doi:10.2307/2184373
- Keller, E. F. (2002). *Making sense of life: Explaining biological development with models, metaphors, and machines*. Cambridge, Mass: Harvard University Press.
- Machery, E. (2012; 2011). Why I stopped worrying about the definition of life... and why you should as well. *Synthese*, 185(1):145–164. doi:10.1007/s11229-011-9880-1
- Tirard, S., Morange, M., & Lazcano, A. (2010). The definition of life: A brief history of an elusive scientific endeavor. *Astrobiology*, 10(10):1003–1009. doi:10.1089/ast.2010.0535

# Holobionts as units of selection: A trait-based account of holobiont individuality

Javier Suárez

**Keywords:** holobionts; units of selection; microbiome; hologenome concept of evolution; biological individuality

Holobionts are biological units that result from the symbiotic merge of a host plus all its associated microbiota, including fungi, bacteria and viruses. Examples of holobionts are very abundant in the biological world—arbuscular plants, corals, cows, etc. all represent cases of holobiosis—and holobiosis is considered to be a pervasive phenomenon among metazoans and plants (Rosenberg & Zilber-Rosenberg 2014, 2016).

Holobionts manifest two prominent features: First, they bear emergent adaptations, i.e. adaptations at the level of the holobiont as an emergent individual; second, they are very “fluid” entities, compounded by organisms from very different lineages and whose identity can (and does) change with time, disappearing and reappearing in relation to developmental stages of the host or the environmental conditions, for instance. The first of these features, i.e. the fact that holobionts bear emergent adaptations, is usually taken as the basis for considering them as “the most fundamental unit of selection” (Dupré and O’Malley 2009: 13; Lloyd 2017). However, the fluid nature of the holobiont and the fact that some of the microorganisms which are integrated in it change with time—feature two—has led some people to argue that the holobiont is not, in fact, a unit of selection, on the basis that its hologenomic identity is not stable but it is constantly changing, and thus it cannot be faithfully transmitted (Moran & Sloan 2015; Douglas & Werren 2016; Skillings 2016). Moreover, as M. O’Malley has pointed out, many of the microorganisms within the holobiont are very transient for it to be considered as a unit of selection (personal communication).

Holobionts, therefore, violate some of our intuitions about units of selection since, on the one hand, it is possible to find arguments in favour of their status as such (holobiont-level adaptations) whereas, on the other, it seems that they violate some of the basic conditions for being a unit of selection (instability across generations). How can we solve this tension?

In the first part of my talk, I will argue that, when considering the case of holobionts, we have to distinguish two different notions of stability: on the one hand, *stability of species* across generations; on the other, intergenerational *stability of traits*, or simply trait-recurrence. Holobionts, I will argue, are sometimes composed of organisms coming from recurrent lineages, and thus some holobionts exhibit stability in the lineages that compose them. A canonical example will be the eukaryotic cell, aphids and their symbiotic *Buchnera aphidicola* or bobtail

squids and their *Vibrio fischeri*. It is important to note that the distinction between horizontal and vertical transmission of the symbionts within the holobiont is independent from the distinction between the two kinds of stability: it might be the case that an example of horizontal transmission, such as the case of the squid and its *V. fischeri*, exhibits recurrence of the species that integrate the holobiont, despite horizontality.

However, stability of species is not the only kind of stability that matters, and this is especially noticeable in the case of holobionts. In the case of holobionts, what matters is the *stability of traits* that reoccur generation after generation despite the changes in the lineages of organisms that compose the holobiont. Those traits that I argue to reoccur every generation are analogous to traditional phenotypic traits (in the sense that they are observable phenotypic features that biologists recognize as independent and distinct) and they are also the result of the interactions among certain genes (plus, of course, any other inheritance factors that might be necessary for their reoccurrence). The peculiarity of these traits is that the genes that interact to produce them belong to some of the different species of organisms that compose the holobiont, and those species might vary intergenerationally provided that the genes responsible for the appearance of the trait are preserved. For instance, in most ruminants the microorganisms that compose the rumen change not only among discrete and separate populations, but also among individuals within the same population (Puniya et al. 2015). Nonetheless, despite these changes in the microorganisms that compose the rumen, there is a special kind of stability that is always preserved among the different organisms, namely: the ability to break down cellulose molecules, produce certain complex aminoacids, etc. This kind of stability has to do with the intergenerational preservation and reoccurrence of traits and, I will contend, is different from the reoccurrence of lineages within the holobiont. The stability of species can be, and in some cases is, the way in which the holobiont ensures the recurrence of the traits that it develops, but it is not the only way to guarantee it, therefore it should not be expected to be ubiquitous in the biological world: holobionts have developed different strategies to guarantee the reoccurrence of traits and that is the kind of stability that matters in this context (this is in line with the arguments presented by O'Malley 2016; Doolittle & Booth 2016; Doolittle 2017).

In the second part, I will argue that the stability of traits exhibited by holobionts is enough to solve the initial tension over their role as units of selection. Provided that the traits are given at the level of the holobiont (since they require of the host plus the specific microbiota to appear) and that they reoccur intergenerationally in a consistent way despite the changes in the lineages that compose the particular holobionts, we might be inclined to say that holobionts are units of selection, without violating the Lewontin's basic conditions (Lewontin 1970). Indeed, I will say, there are mathematical models to account for such stability in terms of constancy of epistatic interactions among different genes even if those belong to different species (Fitzpatrick 2014; Roughgarden et al 2018).

In conclusion, holobionts do not violate the criteria for being units of selection: they only violate our intuitions about the kind of stability that matters for being taken as such.



## References

- Doolittle, W. F. (2017). Darwinizing Gaia. *J Theor Biol*, 434:11–19.
- Doolittle, W. F. & A. Booth (2016). It’s the song not the singer: an exploration of holobiosis and evolutionary theory. *Biol Philos*, 32:5–24. doi:10.1007/s10539-016-9542-2.
- Douglas, A. E. & J. H. Werren (2016). Holes in the hologenome: why host-microbe symbioses are not holobionts. *mBio*, 7(2):e02099-15.
- Dupré, J. & M. A. O’Malley (2009). Varieties of living things: life at the intersection of lineage and metabolism. *Philosophy & Theory in Biology*, 1 (December). doi:10.3998/ptb.6959004.0001.003.
- Fitzpatrick, B. M. (2014). Symbiote transmission and maintenance of extra-genomic associations. *Frontiers in Microbiology*, 5:46. doi: 10.3389/fmicb.2014.00046.
- Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics*, 1:1–18.
- Lloyd, E. (2017). Holobionts as units of selection: Holobionts as interactors, reproducers, and manifestors of adaptation. In S. B. Gissis, E. Lamm & A. Shavit (eds.). *Landscapes of collectivity in the life sciences*. London, The MIT Press: 351–367.
- Moran, N. & D. B. Sloan (2015). The Hologenome Concept: Helpful or Hollow?. *PLoS Biol*, 13(12):e1002311.
- O’Malley, M. A. (2016). Reproduction expanded: Multigenerational and multi-lineal units of evolution. *Philosophy of science*, 83:835–847.
- Rosenberg, E. & I. Zilber-Rosenberg (2014). *The Hologenome Concept*. London, Springer.
- Rosenberg, E. & I. Zilber-Rosenberg (2016). Microbes drive evolution of animals and plants: the hologenome concept. *mBio*, 7(2):e01395-15.
- Roughgarden, J., Gilbert, S. F., Rosenberg, E., Zilber-Rosenberg, I. & Lloyd, E. A. (2017). Holobionts as Units of Selection and a Model of their Population Dynamics and Evolution. *Biological Theory*.
- Skellings, D. (2016). Holobionts and the ecology of organisms: Multi-species communities or integrated individuals? *Bio Philos*, 31:875–892.

# An account of fundamentality for the metaphysics of science

Abel Suñé

Logos research group - Barcelona Institute of Analytic Philosophy  
Philosophy Department - Universitat de Barcelona

A prominent view on metaphysical fundamentality, recently labelled the Ontological Minimality thesis (OM) (Tahko (2018a)), proposes to think of the fundamental as the equivalent to the axioms in a theory, or as the “minimally encompassing portion of reality that God would have to create in order to duplicate this world”. These extant renderings of OM capture significant aspects of fundamentality, but their formulation is visibly circumscribed to the metaphorical level. This paper offers an enhanced version of OM. Following Dennett (1991) and others, I put patterns at the center of ontology. The resulting account uses tools from algorithmic information theory to precisely assess the conditions that give rise to metaphysical dependence and fundamentality in the sciences.

## Patterns and algorithmic information theory

Most people are acquainted with file compression from everyday interaction with personal computers. A compressed algorithmic description is a string of symbols which causes a universal computer to print another string (the one it describes) when it receives it as input while being shorter than it. A pattern in some object, like a dataset, is the aspect in it that allows for a shorter, compressed description of it. More formally, there is a pattern in object  $D$  if there is an algorithm  $A$  such that it is shorter than  $D$  and when input to a computer, the computer outputs  $D$  (or, that is, an encoded full description of it). The length of the shortest algorithm that allows to recover  $D$  is the Kolmogorov Complexity of  $D$ —or  $K(D)$  for short.

A related relevant notion is the complexity of a string,  $Q$ , conditional on another string,  $T$ , or  $K(Q|T)$ . Here it is assumed that we can use  $T$  without cost as an input to our compression algorithm. The length of the shortest such algorithm in these circumstances is  $K(Q|T)$ .

One can quantify the amount of information an object  $x$  carries about another object  $y$ . If  $Q$  and  $T$  are strings, the algorithmic mutual information between them ( $Q:T$ ) is:

$$(Q:T) = K(Q) - K(Q|T)$$

## Ontological minimality and Kolmogorov complexity

OM seeks to identify an ontological basis satisfying (Tahko 2018b):

- *Completeness*: nothing can be added to that basis to make it more informative;
- *Minimality*: nothing can be subtracted from it without incurring in a loss of information.

Let's begin by defining OM in relation to some given domain of data. Let's call the pattern whose length determines  $K(D)$  the *K-pattern for D*.

The K-pattern for dataset D is, by definition, both complete for D, since it is among those patterns that project D, and minimal for D, as it is the shortest among the patterns projecting D.

Correspondingly, OM for domain D is defined:

- D-fundamentality: A real pattern in D is D-fundamental iff it is the K-pattern for D.

D-fundamentality provides a criterion of relativised fundamentality; it might be of interest when inquiring about the fundamental items in a given discipline or domain of relevance. But for many Ds, the D-fundamental pattern will be clearly non-fundamental in an absolute sense.

## Absolute ontological fundamentality

Let W be the dataset corresponding to the entire world. Consider the following, obvious candidate definition of absolute fundamentality:

Fundamentality (first pass): a real pattern is fundamental iff it is W-fundamental.

It follows from this definition that any pattern shorter than  $K(W)$  cannot be fundamental, as fundamentality is only granted to the K-pattern for W. As a consequence, Fundamentality (first pass) entails commitment to priority monism, i.e., the doctrine that there is exactly one fundamental pattern. Although priority monism may very well turn out to be the truth about our cosmos, it is, needless to say, not to be established by a priori means; a proper theory of fundamentality must remain noncommittal about this issue and allow for pluralism/monism to be established scientifically.

Thus, while D-fundamentality was too liberal an account for absolute fundamentality, deeming fundamental virtually every pattern, Fundamentality (first pass) is unduly constraining.

Let a dataset be independent iff it includes all and only those data-points with which it is *informationally dependent*. Any two data sets D and T are informationally dependent iff there is non-zero mutual information between them, or:  $K(D,T) < K(D) + K(T)$ . The final definition of fundamentality states:

**Fundamentality:** a pattern is fundamental iff there is an independent dataset Y for which it is Y-fundamental.

Under this definition, the problem of unbounded proliferation of “fundamental” patterns we saw in D-fundamentality does not arise, for only the K-patterns for datasets that go all the way down in any given ontological dependence chain count as fundamental.

Furthermore, the issue of dataset (in)dependence is settled empirically. Priority monism holds iff there is just one independent dataset. Otherwise, pluralism ensues. Those dependencies are brought to bear by patterns, whose discovery is a strictly empirical matter.

The complete minimal basis of the world, whether monistic or pluralistic, is constituted by the set of all fundamental patterns as defined in *Fundamentality*.

## References

- Dennett, D. (1991). “Real Patterns”. *Journal of Philosophy*.
- Tahko, T. E. (2018a). “Fundamentality and Ontological Minimality”. In Ricki Bliss & Graham Priest (eds.), *Reality and its Structure*. Oxford University Press.
- Tahko, T. E. (2018b). “Fundamentality”. *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/fundamentality/>.

# Scientific realism beyond pessimistic induction: From truth to plausibility

Sergio Urueña López

University of the Basque Country UPV/EHU

sergio.uruenal@ehu.eus

**Keywords:** pessimistic induction, no-miracles argument, truth, plausibility, meta-analysis

Although the dispute between scientific realism and anti-realism has been considered settled (e.g., Fine 1984), the debate still has a prominent role in the contemporary philosophy of science (e.g., Dicken 2016). In particular, two conspicuous arguments have remained encouraging the academic discussion, namely, the ‘no-miracles’ argument (NMA), and the pessimistic induction (PI). In broader terms, while NMA supports that the best explanation for the empirical success of our current scientific theories lies in their truth (Maxwell 1962: 18; Smart 1963: 39; Putnam 1975: 73), PI<sup>1</sup> diminishes the legitimacy of the correlation between empirical success and truth emphasised in NMA. Concretely, based on the evidence that most prior scientific theories (acknowledged as false at present) were empirically successful, PI supports that the correlation between truth and empirical success is unwarranted (Poincaré 1902: 160; Putnam 1976: 184; Laudan 1977: 126, 1981; Carrier 1991).

Since the emergence of NMA/PI, both arguments have been widely questioned and reformulated (Chakravartty 2017). The diverse criticisms and revisions that revolve around them can be classified into three specific but mutually dependent theoretical dimensions: (i) the characterization, (ii) the scope, and (iii) the justification<sup>2</sup>. The degree of discrepancy found in the present specialized literature seems to vary from one dimension to another. Although no substantial dispute have been associated to the scope of NMA/PI, a strong disagreement brews pertaining to the nature of these arguments concerning their structures and whether they are indeed valid and sound lines of reasoning (see Table 1). For instance, the latest dimension, i.e. the justificatory one, happens to be the most controversial. In order to test the validity of NMA and PI, it is common for scholars to adhere two key strategies, which are: a) to assess the robustness of the premises, and/or b) to examine the logical form of the arguments (see Table 1). Nevertheless, the question about the soundness of NMA/PI considering their underlying concepts of truth has yet to be adequately addressed. The problematization of this question may be essential to determine whether

---

<sup>1</sup>Wray (2015) identified four IPs. This paper is focused on the classical IP formulated by Larry Laudan in *A Confutation of Convergent Realism* (1981).

<sup>2</sup>The *basic structure* of the arguments affects their *scope* and *justification*. Indeed, each characterization may have different scope and would need specific evidence in order to be criticized or justified.

realism/anti-realism disagreement has arisen because of the use of differentiated concepts of truth.

The purpose of this paper is to present a meta-analysis of the realism/anti-realism debate in its justificatory dimension focused on NMA/PI's concepts of truth. The principal hypothesis outlined in this study is that despite the fact that NMA/PI are sound and valid arguments, PI is defective against any form of scientific realism in which truth is understood in terms of *plausibility*. First, the paper points out that the effectiveness of NMA/PI is dependent on the commitment towards differentiated conceptual frameworks regarding truth, i.e., both arguments are proven to work *if and only if* a concrete notion of truth is previously assumed for each case. Second, the paper identifies which one of those conceptual frameworks of truth is more suitable to comprehend scientific practice. Finally, the study concludes that NMA is reasonable beyond PI *iff* the truth of our current scientific theories is understood from the stance of *plausibility*. The relevance of this research is to shed new light on NMA/PI debate and propose *plausibility* as a novel and robust enough analytical tool which supports NMA and scientific realism.

The fact that the effectiveness of NMA/PI depends on the commitment to differentiated conceptual frameworks of truth is illustrated through a meta-analysis concerning the criticisms and counterarguments placed against (or in favour of) both the arguments (see Table 1 'Justification'). The criticism pointed out by Wray towards the notion upheld by Lewis (2001) may operate *in this extended abstract* as an example. Wray argues that, in order to disprove PI, "it is not enough for the realist to undercut the claim that success is a reliable indicator of truth" (Wray 2013: 1719). Instead, "the realist must show that *our current best theories are likely true*" (Wray 2013: 1720; original emphasis). Nonetheless, a realist could reply that the task imposed by an anti-realist (Wray) is disproportionate because it is impossible to offer *definite proofs* regarding the truth of present scientific theories; the only viable method is to offer sound and valid arguments, in which the NMA becomes an option (Popper 1989: 39; Diéguez 1998; Psillos 1999: 99).

The point emphasised here is that while the anti-realist in PI is committed to a concept of truth that is based on *certainty*<sup>3</sup> (X is true *if and only if* there is certainty that X is true), the realist in NMA may be less demanding or ambitious (*pace* Levin 1984). The realist could maintain NMA and address fallibilism without committing logical mistakes. In precise, it is not contradictory to accept the impossibility of seeking a certain truth criterion (Frege 1984: 353; Tarski 1944: 364) and maintaining that success is indeed a reliable *indicator* of truth—not a '*test*'—if "truth" is understood from the light of *probability or plausibility*. Based on recent studies (e.g., Vega 1998, 2011; Celucci 2014, 2017; Sterpetti & Bertolaso 2018), this study proposes *plausibility* as a robust epistemological criterion that may help to overcome some criticisms of *probabilistic* approaches<sup>4</sup> and defend the reasonableness of NMA and the scientific realism.

---

<sup>3</sup>The correlation of certainty-truth is also clear in Laudan (1981: 31, footnote 10).

<sup>4</sup>For example, the base rate fallacy (Howson 2000, 2013) or the "uncertainty about probabilities" (Teira 2011).

| DIMENSION                                                                                  | EXAMPLES          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                    |
|--------------------------------------------------------------------------------------------|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Characterization</b><br>‘What are the premises and the logical forms of the arguments?’ | NMA               | <p>(NMA1) The scientific enterprise is (enormously) successful.</p> <p>(NMA2) The best explanation for NMA1 is that Scientific Realism (SR) is true.</p> <hr/> <p>(<math>\therefore</math> NMA3) Therefore, by IBE (Inference to the Best Explanation):</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | <p><i>Authors who have articulated versions of NMA:</i> Maxwell (1962), Smart (1963: 39), Putnam (1975: 73), Boyd (1980), Brown (1982), McMullin (1987), Musgrave (1988: 239), Lipton (1994, 2004: 202–3), Leplin (1997), Psillos (1999), Lyons &amp; Clarke (2002: xii), Lyons (2003), Devitt (2008: 227), and Dellsén (2015)</p> |
|                                                                                            | PI                | <p>(PI1) There have been successful (<math>S</math>) but false (<math>\neg V</math>) scientific theories (<math>t</math>). <math>\exists t(S t \wedge \neg V t)</math></p> <hr/> <p>(<math>\therefore</math> PI2) The success of a theory is not a reliable <i>test</i> for its truth. <math>\neg \forall t(S t \rightarrow V t)</math></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | <p><i>Authors who have articulated versions of IP:</i> Hesse (1976), Poincare (1902: 160), Putnam (1976: 184), Laudan (1977: 126, 1981), Psillos (1996: 307; 1999: 102–3), Lewis (2001: 373), Lange (2002), Saatsi (2005: 1089), Mizrahi (2013), Ruhmkorff (2013: 410), Wray (2013, 2015)...</p>                                   |
| <b>Scope</b> ‘Are they global or local arguments?’                                         | NMA               | Sprenger (2016), Henderson (2017), Iranzo (forthcoming).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                    |
|                                                                                            | PI                | Kitcher (1993: 136, n14), Ruhmkorff (2013: 410).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                    |
| <b>Justification</b> ‘Are they valid and sound arguments?’                                 | Criticisms of NMA | <p>a) <i>Premises:</i></p> <ul style="list-style-type: none"> <li>▪ Vs. NMA1: The notion of success which is operative in science lacks diachronic stability (Wray 2013: 1725-1727).</li> <li>▪ Vs. NMA2: “There seems to be no good reason to prefer (ESR) [here (SR)] over alternative explanations” (Mizrahi 2012).</li> <li>▪ Vs. <math>\therefore</math> NMA3 (a&amp;b): <ul style="list-style-type: none"> <li>• The connection success-truth is unwarranted (Laudan 1981) (<math>\therefore</math> PI2).</li> <li>• There is no need to invoke truth to explain the success of our current theories (Wray 2010, 2013). Either it is not reasonable to believe that <i>a theory</i> is true based on its empirical success, since a better rival explanation or theory T’ with the same ‘instrumental reliability’ (Fine 1986, 1991) or ‘empirical adequacy’ (van Fraassen 1980, 1989) may always be constructed.</li> <li>• Underconsideration argument (van Fraassen 1989: 143; Lipton 1993): Our selection of ‘best explanations’ may be a ‘bad lot’ (see Khalifa 2010).</li> <li>• ANM “cannot support any realist persuasions” because “it doesn’t work empirically” (Carrier 1991: 35).</li> </ul> </li> </ul> |                                                                                                                                                                                                                                                                                                                                    |

|  |                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  |                  | <p>b) <i>Logical form</i>:</p> <ul style="list-style-type: none"><li>▪ Base rate fallacy (a&amp;b): Howson (2000, 2013).</li><li>▪ Circularity: a) NMA uses IBE to justify IBE (Lipton 2000: 194-195) / b) A hidden premise presupposes SR (van Fraassen 1980: 21).</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|  | Criticisms of PI | <p>a) <i>Premises</i>:</p> <ul style="list-style-type: none"><li>▪ Vs. PI1: Past successful theories were not completely false (e.g., Hardin &amp; Rosenberg 1981; Devitt 1984; Kitcher 1993; Leplin 1997: 141; Psillos 1999: chap. 5; Nola 2008), neither completely successful (e.g., Worrall 1989, 1994: 335).</li><li>▪ Vs. .'. P2: Exponential growth of science (Fahrbach 2011) and superiority of present theories over the past ones (Musgrave 1985: 211; Doppelt, 2007: 111, 2014; Saatsi 2009: 358; Park 2011: 80). Then, not pessimistic but 'optimistic induction' (see Nola 2008).</li></ul> <p>b) <i>Logical form</i>: Turnover fallacy (Kitcher 1993; Lange 2002) / Base rate fallacy (Lewis 2001; Magnus &amp; Callender 2004; Mizrahi 2013: 3215).</p> <p><b>This paper:</b> Empirical success is certainly not a <i>test</i> for truth, but it is a <i>plausible indicator</i> for truth. Then, NMA is reasonable.</p> |

Table 1: Dimensions of NMA/PI debate. Source: Elaborated by the author.



## References

- Boyd, Richard (1980). Scientific Realism and Naturalistic Epistemology. In Peter D. Asquith & Thomas Nickles (Eds.), *PSA 1980: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (pp. 613–662). East Lansing: Philosophy of Science Association.
- Brown, James R. (1982). The Miracle of Science. *Philosophical Quarterly*, 32(128):232–244.
- Carrier, Martin (1991). What is wrong with the miracle argument? *Studies in History and Philosophy of Science*, 22(1):23–36.
- Cellucci, Carlo (2014). Knowledge, Truth and Plausibility. *Axiomathes*, 24:517–532.
- Cellucci, Carlo (2017). *Rethinking Knowledge. The heuristic view*. Dordrecht: Springer.
- Chakravartty, Anjan (2017). Scientific Realism. *The Stanford Encyclopedia of Philosophy*.
- Dellsén, Finnur (2015). Explanatory Rivals and the Ultimate Argument. *Theoria*, 82(3):217–237.
- Devitt, Michael (1984). *Realism and Truth*. Oxford: Blackwell.
- Devitt, Michael (2008). Realism/anti-realism. In S. Psillos & M. Curd (Eds.), *The Routledge companion to philosophy of science* (pp. 224–235). London: Routledge.
- Dicken, Paul (2016). *A Critical Introduction to Scientific Realism*. London: Bloomsbury Academic.
- Diéguez, Antonio (1998). Los compromisos del realismo científico. *Contrastes. Revista Interdisciplinar de Filosofía*, 3:145–173.
- Doppelt, Gerald (2007). Reconstructing Scientific Realism to Rebut the Pessimistic Metainduction. *Philosophy of Science*, 74(1):96–118.
- Doppelt, Gerald (2014). Best Theory Scientific Realism. *European Journal for Philosophy of Science*, 4(2):271–291.
- Fahrbach, Ludwig (2011). How the Growth of Science Ends Theory Change. *Synthese*, 180(2):139–155.
- Fine, Arthur (1984). The Natural Ontological Attitude. In Jarret Leplin (Ed.), *Scientific Realism* (pp. 61–77). Berkeley: University of California Press.
- Fine, Arthur (1986). Unnatural Attitudes: Realist and Instrumentalist Attachments to Science. *Mind*, 95:149–179.
- Fine, Arthur (1991). Piecemeal Realism. *Philosophical Studies*, 61(12):79–96.

- Frege, Gottlob (1984). *Collected papers on mathematics, logic, and philosophy*. Oxford: Blackwell.
- Hardin, Clyde L. & Rosenberg, Alexander (1982). In Defence of Convergent Realism. *Philosophy of Science*, 49(4):604–615.
- Harman, Gilbert H. (1965). The inference to the best explanation. *Philosophical Review*, 74(1):88–95.
- Henderson, Leah (2017). The no miracles argument and the base rate fallacy. *Synthese*, 194(4):1295–1302.
- Hesse, Mary (1976). Truth and the Growth of Scientific Knowledge. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2:261–280.
- Howson, Colin (2000). *Hume's problem: Induction and the justification of belief*. Oxford: Oxford University Press.
- Howson, Colin (2013). Exhuming the No-Miracles Argument. *Analysis*, 73(2):205–211.
- Iranzo, Valeriano (forthcoming). Realismo, probabilidad y éxito predictivo. *ArtefaCToS. Revista de Estudios sobre la Ciencia y la Tecnología*, 7(2).
- Khalifa, Kareem (2010). Default Privilege and Bad Lots: Underconsideration and Explanatory Inference. *International Studies in the Philosophy of Science*, 24(1):91–105.
- Kitcher, Philip (1993). *The Advancement of Science*. Oxford: Oxford University Press.
- Lange, Marc (2002). Baseball, pessimistic inductions and the turnover fallacy. *Analysis*, 62(276):281–285.
- Laudan, Larry (1977). *Progress and Its Problems: Towards a Theory of Scientific Growth*. California: University of California Press.
- Laudan, Larry (1981). A Confutation of Convergent Realism. *Philosophy of Science*, 48(1):19–49.
- Leplin, Jarrett (1997). *A Novel Defense of Scientific Realism*. New York: Oxford University Press.
- Levin, Michael (1984). What kind of explanation is Truth? In Jarret Leplin (Ed.), *Scientific Realism* (pp. 124–139). Berkeley: University of California Press.
- Lewis, Peter J. (2001). Why The Pessimistic Induction Is A Fallacy. *Synthese*, 129(3):371–380.
- Lipton, Peter (1993). Is the best good enough? *Proceedings of the Aristotelian Society*, 93:89–104.

- Lipton, Peter (1994). Truth, Explanation, and the Best Explanation. In A. A. Derksen (ed.), *The Scientific Realism of Rom Harré* (pp. 89–111). Tilburg: Tilburg University Press.
- Lipton, Peter (2000). Tracking track records. *Proceedings of the Aristotelian Society*, 74:179–205.
- Lipton, Peter (2004). *Inference to the best explanation*. London: Routledge.
- Lyons, Timothy D. & Clarke, Steve (2002). Scientific Realism and Commonsense. In Steve Clarke & Timothy Lyons (Eds.), *Recent Themes in the Philosophy of Science: Scientific Realism and Commonsense* (pp. ix–xxiii). Dordrecht: Kluwer.
- Lyons, Timothy D. (2003). Explaining the Success of a Scientific Theory. *Philosophy of Science*, 70(5):891–901.
- Magnus, P. D. & Callender, Craig (2004). Realist Ennui and the Base Rate Fallacy. *Philosophy of Science*, 71(3):320–338.
- Maxwell, Grover (1962). The Ontological Status of Theoretical Entities. In Herbert Feigl & Grover Maxwell (eds.), *Scientific Explanation, Space, and Time* (pp. 3–27). Minnesota: University of Minnesota Press.
- McMullin, Ernan (1987). Explanatory Success and the Truth of Theory. In Nicholas Rescher (Ed.), *Scientific Inquiry in Philosophical Perspective* (pp. 51–73). Lanham: University Press of America.
- Mizrahi, Moti (2012). Why the ultimate argument for scientific realism ultimately fails. *Studies in History and Philosophy of Science*, 43:132–138.
- Mizrahi, Moti (2013). The Pessimistic Induction: A Bad Argument Gone Too Far. *Synthese*, 190(15):3209–3226.
- Musgrave, Alan (1985). Realism vs. Constructive Empiricism. In Paul M. Churchland & Hooker A. Clifford (eds.), *Images of Science: Essays on Realism and Empiricism* (pp. 1088–1113). Chicago: Chicago University Press.
- Musgrave, Alan (1988). The ultimate argument for scientific realism. *Relativism and realism in science*, 6:229–252.
- Nola, Robert (2008). The optimistic meta-induction and ontological continuity: the case of the electron. In Léna Soler, Howard Sankey & Paul Hoyningen-Huene (Eds.), *Rethinking scientific change and theory comparison: Stabilities, ruptures, incommensurabilities* (pp. 159–202). Dordrecht: Springer.
- Park, Seungbae (2011). A Confutation of the Pessimistic Induction. *Journal for General Philosophy of Science*, 42(1):75–84.
- Poincaré, Henry (1902). *Science and hypothesis*. London: Walter Scott Publishing.
- Popper, Karl (1989). *Objective knowledge: An evolutionary approach*. Oxford: Oxford University Press.

- Psillos Stathis (1996). Scientific Realism and the “Pessimistic Induction”. *Philosophy of Science*, 63:S306–S314.
- Psillos Stathis (1999). *Scientific Realism: How Science Tracks Truth*. New York: Routledge.
- Psillos, Stathis (2007). The Fine Structure of Inference to the Best Explanation. *Philosophy and Phenomenological Research*, 74(2):441–448.
- Putnam, Hilary (1975). *Mathematics, Matter and Method: Philosophical Papers*. London: Cambridge University Press.
- Putnam, Hilary (1976). What is Realism? *Proceedings of the Aristotelian Society*, 76(1):177–194.
- Putnam, Hilary (1978). *Meaning and the Moral Sciences*. London: Hutchinson.
- Ruhmkorff, Samuel (2013). Global and Local Pessimistic Meta-inductions. *International Studies in the Philosophy of Science*, 27(4):409–428.
- Saatsi, Juha (2005). On the pessimistic induction and two fallacies. *Philosophy of Science*, 72(5):1088–1098.
- Saatsi, Juha (2009). Grasping at Realist Straws. *Review Symposium. Metascience*, 18:355–362.
- Smart, John Jamieson C. (1963). *Philosophy and scientific realism*. London: Routledge.
- Sprenger, Jan (2016). The probabilistic no miracles argument. *European Journal for Philosophy of Science*, 6(2):173–189.
- Sterpetti, Fabio & Bertolaso, Marta (2018). The Pursuit of Knowledge and the Problem of Unconceived Alternatives. *Topoi*:1–12.
- Tarski, Alfred (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomonological Research*, 4(3):341–376.
- Teira, David (2011). Frequentist vs. Bayesian clinical trials. In Fred Gifford (Ed.), *Handbook of the Philosophy of Science. Volume 16. Philosophy of Medicine* (pp. 255–297). Amsterdam: Elsevier.
- van Fraassen, Bas C. (1980). *The Scientific Image*. Oxford: Oxford University Press.
- van Fraassen, Bas C. (1989). *Laws and Symmetry*. Oxford: Clarendon Press.
- Vega, Luis (1998). Aristotle’s Endoxa and Plausible Argumentation. *Argumentation*, 12(1):95–113.
- Vega, Luis (2011). Plausible / Plausibilidad. In Luis Vega & Paula Olmos (Eds.), *Compendio de lógica, argumentación y retórica* (pp. 455–458). Madrid: Trotta.
- Worrall, John (1989). Structural Realism: The best of both worlds. *Dialectica*, 43(1-2):99–124.

- Worrall, John (1994). How to Remain (Reasonably) Optimistic: Scientific Realism and the 'Luminiferous Ether'. *PSA*, 1:334–342.
- Wray, K. Brad (2010). Selection and predictive success. *Erkenntnis*, 72(3):365–377.
- Wray, K. Brad (2013). Success and truth in the realism/anti-realism debate. *Synthese*, 190(9):1719–1729.
- Wray, K. Brad (2015). Pessimistic Inductions: Four Varieties. *International Studies in the Philosophy of Science*, 1(29):67–73.

# Probing geno-centric approaches to conservation biology

**Davide Vecchi**

Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

**Elena Casetta**

Department of Philosophy and Educational Sciences, University of Turin, Italy

**Abstract** The literature in biodiversity studies and conservation biology mirrors the same lurking ideological clash between geno-centric approaches and their alternatives pervading the practice of other biological disciplines. On the one hand, some biologists and philosophers argue that genes cannot be chosen as the fundamental unit of conservation. On the other hand, many conservation biologists consider genetic diversity as primary. We suggest that geno-centric approaches to conservation biology rest on an explicit rationale and an implicit assumption. The explicit rationale is that genetic diversity is the main prerequisite for evolutionary adaptation, i.e. the key condition for a species to survive in a rapidly changing environment and also to speciate. The underlying implicit assumption is that phenotypic diversity somehow depends on genomic diversity. In this contribution we propose to scrutinize both the explicit rationale and the implicit assumption. While the tenability of the first is, at least in part, challenged by what might be called the “paradox of evolution and conservation”, the tenability of the second is challenged by the difficulty of characterizing the causal link between genomic and phenotypic diversity.

**Keywords** Genomics; conservation biology; geno-centrism; evolutionary adaptation.

Criticisms of geno-centrism abound in the biological and, above all, in the philosophical literature (see for instance the many contributions in Oyama et al. 2001 or the more recent monograph by Griffiths and Stotz 2013). This view is in sharp contrast with contemporary biological practice in foundational disciplines like developmental genetics (Wagner 2014), molecular population genetics (Lynch 2007) and comparative genomics (Koonin 2009). Geno-centric approaches of the latter kind are certainly not committed to the traditional fallacies of genetic determinism and reductionism, recognize the important causal role of “environmental” (i.e., non-sequence) factors in development as well as the existence of non-genetic forms of inheritance. Nonetheless, geno-centric approaches seem to strongly vindicate the central causal role of DNA, genes and genomes both in development and evolution.

Interestingly, the literature in biodiversity studies and conservation biology mirrors the same lurking ideological clash between geno-centric approaches and their alternatives pervading the practice of other biological disciplines. On the one hand, some biologists and philosophers argue that genes cannot be chosen as the fundamental unit of conservation: by conserving all genomic variation we do not concomitantly preserve all the phenotypic variation that characterizes populations of conservation interest (not to speak of the conservation of endangered biological phenomena or ecosystems). Sarkar (2002) notes that a geno-centric position in conservation biology can be justified only if some form of “global genetic reductionism” is vindicated. Geno-centric views of conservation in fact typically spouse the implicit assumption that phenotypic diversity somehow depends on genomic diversity. Mace et al. (2014), for instance, suggest that there might be a fundamental reason to focus conservation effort on the preservation of some portions of the “genetic library of life”; in fact, genetic diversity might be seen as representing all diversity (i.e., phenotypic) and its potential future utility (e.g., functionally, adaptively). But, in a very clear sense, global genetic reductionism is wrong, fundamentally because phenogenesis at all levels (from transcription, translation and protein folding up to cellular differentiation and morphogenesis) is causally influenced by a variety of environmental inputs; thus, saving all genes would not save all possible phenotypic outcomes unless we also conserved all possible developmental environments. On the other hand, choosing genes seems to be, in the case of conservation biology, motivated by pragmatic rather than ideological reasons. Many conservation biologists consider genetic diversity as the foundational basis of conservation efforts because, insofar as it is needed for evolutionary adaptation, it is what allows the long-term survival of species (Van Dyke 2008, ch. 6). Moritz (1994) for instance suggested a genetic definition of the concept of evolutionarily significant unit. The explicit rationale for this geno-centric view is clearly not a defence of some form of genetic determinism for its own sake, but rather the realisation that conservation effort should be focused on species’ capability to adapt to environmental change over the long term, and maintaining species’ genetic diversity (by guaranteeing that a significant amount of genetic variation is preserved, and by preventing the fixation of deleterious alleles) is the *conditio sine qua non* to maintain their evolutionary potential.

In this contribution we propose to assess merits and limits of geno-centric approaches to biodiversity conservation by situating the analysis in the context of the more general debate concerning the status of geno-centrism in the life sciences. In particular, we shall discuss both the explicit rationale and the implicit assumption at their basis. The explicit rationale is that genetic diversity is the main prerequisite for evolutionary adaptation, i.e. the key condition for a species to survive in a rapidly changing environment and also to speciate. The underlying implicit assumption is that phenotypic diversity somehow depends on genomic diversity. The tenability of the explicit rationale is, at least *prima facie*, challenged by what might be called the “paradox of evolution and conservation”, i.e. the fact that evolution implies change, and conservation implies keeping things as they are, or even bringing them back to their previous, allegedly pristine, state. The tenability of the implicit assumption is challenged by the difficulty of characterizing the causal link between genomic and phenotypic diversity. That there exists a dependence of some kind between genomic and

phenotypic complexity can be reasonably inferred from the large-scale evidence concerning the correlated differences in genomic and phenotypic architecture between prokaryotes and eukaryotes. But such correlations are both difficult to characterize precisely and to assess empirically (Sarkar 2015). For instance, the correlation between genome size and body size (two crude measures of complexity) is positive only for certain taxa (Dufresne and Jeffery 2011). So, how should the concepts of genomic and phenotypic diversities be characterised? As in the case of the correlation between genomic and phenotypic complexity, this turns out to be a profound biophilosophical question.

## Bibliography

- Dufresne, F. and Jeffery, N. 2011. A Guided Tour of Large Genome Size in Animals: What We Know and Where We Are Heading. *Chromosome Research*, 19:925–38.
- Griffiths, P. and Stotz, K. 2013. *Genetics and philosophy: an introduction*. Cambridge: Cambridge University Press.
- Koonin, E. V. 2009. Darwinian evolution in the light of genomics. *Nucleic Acids Research*, 37(4):1011–1034.
- Lynch M. 2007. *The Origins of Genome Architecture*. Sinauer: Sunderland.
- Mace, G. et al. 2014. Approaches to defining a planetary boundary for biodiversity. *Global Environmental Change*, 28:289–297.
- Moritz, C. 1994. Defining “Evolutionarily Significant Units” for conservation. *Trends in Ecology and Evolution*, 9(10):373–5.
- Oyama, S., Griffiths P.E. and Gray, R. (eds). 2001. *Cycles of contingency. Developmental Systems and Evolution*. Cambridge: MIT Press.
- Sarkar, S. 2002. Defining “biodiversity”; Assessing Biodiversity. *The Monist*, 85(1):131–155.
- Sarkar, S. 2015. The genomic challenge to adaptationism. *British Journal for the Philosophy of Science*, 66(3):505–536.
- Van Dyke, F. 2008. *Conservation Biology*. Springer, Dordrecht.
- Wagner, G. P. 2014. *Homology, Genes, and Evolutionary Innovation*. Princeton University Press.



# Kant and Wittgenstein on pictures and proofs in mathematics

Pedro A. Viñuela Villa

In the *Critique of Pure Reason*, Kant argues that mathematics is a rational knowledge gained by construction of concepts in intuition and it is because of it that mathematics can be *a priori* cognition and applied to experience. In giving this philosophical interpretation of method of construction in mathematics, he aims to put mathematical knowledge upon *a priori* grounds. And in order to support this thesis, Kant uses to appeal to the common practice of elementary mathematics. Essentially, this is the reason to assign a key role to figures in geometry and to ostensive representations in arithmetic.

Indeed, Kant insists on the need to resort to pictures and intuitive representations (geometrical figures, strokes, dots) to ground the mathematical propositions. For this reason, Kant asserts that number seeks its standing and sense in the fingers, in the beads of an abacus, or in strokes and points that are placed before the eyes (KrV, A 240/B 299). That suggests that to him the diagrams and figures placed *ante oculos*, are not just heuristic aids to understand better a proof; rather, they are essential and irreplaceable components of a mathematical demonstration (see Shabel, 2006).

Notwithstanding, according to Russell (2007, p. 145), “the whole trend of modern mathematics, with its increased pursuit of rigour, has been against this Kantian theory”. In fact, not only Russell but many others have argued that what is *visually intuitive as such* cannot be a solid and certain guarantee for mathematical truth. Leibniz, for instance, a conspicuous mathematician himself, regarded the “ecthetic” manner whereby the Ancient geometers proved a proposition (by using figures and constructions for imagination) something dispensable (see *New Essays*, IV, 1, §9). Similarly, A. J. Ayer maintained later that “the use of diagrams is not essential to completely rigorous geometry. The diagrams are introduced as an aid to our reason.” (Ayer, 1962, p. 83).

Since the last nineteenth century this view has been confirmed for analysis and geometry, for visual intuition and diagrammatic reasoning were regarded as heuristically fruitful resources, but as unreliable in justification. Therefore, they were discredited in rigorous mathematics and relegated to a second-order position.

Yet nowadays this long-standing prejudice against diagrammatic representation is changing. There are many mathematicians asking “more visual” approaches and willing to reconsider the real role of pictures and diagrams in mathematical reasoning. They are attempting to rescue visual thinking for logic and mathematics to the point we can talk about a renaissance of visualization over recent decades. This revival has emerged because of developments in areas as

different as computer science, mathematics, mathematics education, cognitive psychology, and philosophy (see Mancosu, 2005).

Treating the relevance of visualization in mathematics, Giaquinto (2007) shows, for instance, that visual thinking doesn't work merely as a psychological element but also as an epistemological one. As Shin, Lemon & Mumma (2014, ¶1) remark, "recently, many philosophers, psychologists, logicians, mathematicians, and computer scientists have become increasingly aware of the importance of multi-modal reasoning and, moreover, much research has been undertaken in the area of non-symbolic, especially diagrammatic, representation systems."

With this background in mind, it can be of special interest reconsider with new eyes the later Wittgenstein's philosophy of mathematics, for it is well-known his fondness for pictures. As Brown remarks, it is no exaggeration to say that "there are as many pictures in Wittgenstein's published works as there are in all the other great philosophers combined." (2008, p. 136). No doubt Wittgenstein lays special stress on diagrams and pictures in mathematics. In particular, he emphasizes the image of the mathematician as inventor or fashioner of models, pictures, and concepts (see Floyd, 2005, p. 112). But this raises the interesting problem as to why many diagrams are cogent and able to serve as proof, in spite of its possible flaws. In fact, to Wittgenstein the proof is not a series of sentences meeting purely formal deductive requirements; it involves many other things because it is a complex human activity.

However, to Kant some diagrams are constitutive in elementary mathematics, at least in a foundational level. They are not only heuristic devices or a mere aid to support the proof. In fact, philosophical notion of construction of concepts in intuition is taken by Kant firstly from the tradition and practice of ancient Greek geometry. Moreover, geometrical figures aren't auxiliary resources, rather they are objects. In this sense, his view on diagrams is somehow part of his general theory of the constitution of objective experience.

Therefore, in this paper I will focus on examine the role Wittgenstein attributes to pictures and diagrams in mathematics in comparison with the place Kant concedes to constructions and ostensive representations. To compare Wittgenstein with Kant is a good choice because it is possible to "reconstruct a great deal of the philosophy of mathematics simply in terms of attitudes and reactions to Kant." (Brown, 2008, p. 120). Besides, as mentioned earlier, taking Euclidian diagram as model, the German philosopher has placed the pictures in a central place as a constitutive and indispensable factor in mathematical reasoning and method. Additionally, I compare Wittgenstein and Kant because "Wittgenstein radicalized Kant's idea that the laws of logic and/or mathematics require philosophical analysis because of our tendency to misconstrue them as unconditionally true of an ahumanly conceived domain of necessary truth." (Floyd, 2005, p. 79). In this sense, Wittgenstein conceives mathematics as a practice and an activity, not only a science of truth or a body of knowledge. Indeed, though there seem to be similarities between Kant's critical philosophy of mathematics and Wittgenstein's later philosophy of mathematics, there exist certainly profound differences, which I seek to outline in the paper. Thus, the aim is to raise questions as: Is the function of diagrams in a mathematical proof merely heuristic and tentative or rather constitutive and objective? Can and should we dispense with pictures in a rigorous proof? In the end, I believe that

a confrontation between both views of mathematics could provide as a fruitful outcome a mutual clarification in some interesting aspects concerning Kant's and Wittgenstein's philosophy.

## References

- Avigad, J. (2011). Understanding Proofs. In P. Mancosu (Ed.) (2011). *The Philosophy of Mathematical Practice* (pp. 317–353). New York: Oxford University Press.
- Ayer, A. J. (1962). *Language, Truth and Logic*. Victor Gollancz Ltd., London.
- Bangu, S. (s.f.). Ludwig Wittgenstein: Later Philosophy of Mathematics. *Internet Encyclopedia of Philosophy*. URL = <http://www.iep.utm.edu/wittmath/>
- Brown, J. R. (2008). *Philosophy of Mathematics: A Contemporary Introduction to the World of Proofs and Pictures* (second edition). New York: Routledge.
- Coffa, J. A. (1993). *The Semantic Tradition from Kant to Carnap: To the Vienna Station*. New York: Cambridge University Press.
- Fann, K. T. (2013). *El concepto de filosofía en Wittgenstein* (tercera edición). Traducción de Miguel Ángel Beltrán. Madrid: Tecnos.
- Floyd, J. (2005). Wittgenstein on Philosophy of Logic and Mathematics. In Shapiro, Stewart (Ed.). *The Oxford Handbook of Philosophy of Mathematics and Logic* (pp. 75–128). New York: Oxford University Press.
- Frascolla, P. (1994). *Wittgenstein's Philosophy of Mathematics*. London and New York: Routledge.
- Giaquinto, M. (2007). *Visual Thinking in Mathematics. An Epistemological Study*. New York: Oxford University Press.
- Hacker, P. M. S. (2013). *Wittgenstein: Comparisons and Context*. Oxford/New York: Oxford University Press.
- Hanna, R. (2001). *Kant and the Foundations of Analytic Philosophy*. New York: Oxford University Press.
- Kant, I. (1902-1955). *Gesammelte Schriften*. Herausgegeben von der K. Preussischen (23 vols.), bzw. Deutschen Akademie der Wissenschaften, Berlin.
- Kant, I. (1998). *Critique of Pure Reason*. Translated and edited by Paul Guyer and Allen W. Wood. Cambridge/New York: Cambridge University Press.
- Kripke, S. A. (2006). *Wittgenstein a propósito de reglas y lenguaje privado. Una exposición elemental*. Traducción de Jorge Martínez Marqueze. Madrid: Tecnos.
- Leibniz, G. W. (1981). *New Essays on Human Understanding*. Translated and edited by Peter Remnant & Jonathan Bennett. New York: Cambridge University Press.

- Mancosu, P. (Ed.) (2011). *The Philosophy of Mathematical Practice*. New York: Oxford University Press.
- Mancosu, P. (2005). Visualization in logic and mathematics. In P. Mancosu et al. (Eds.). *Visualization, Explanation and Reasoning Styles in Mathematics* (pp. 13–30). Dordrecht, Springer.
- Monk, R. (1990). *Ludwig Wittgenstein: The Duty of Genius*. New York: The Free Press.
- Rodych, V. (2011). Wittgenstein's Philosophy of Mathematics (Summer 2011 Edition). *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. URL = <http://plato.stanford.edu/archives/sum2011/entries/wittgenstein-mathematics/>
- Russell, B. (2007). *Introduction to Mathematical Philosophy*. New York: Cosimo.
- Shabel, L. (2006). Kant's philosophy of mathematics. In P. Guyer (ed.). *The Cambridge Companion to Kant and Modern Philosophy* (pp. 94–128). New York: Cambridge University Press.
- Shin, S.-J., Lemon, O. and Mumma, J. (2014). Diagrams (Winter 2014 Edition). *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. URL = <http://plato.stanford.edu/archives/win2014/entries/diagrams/>
- Sluga, H. & Stern, D. G. (Eds.) (1996). *The Cambridge Companion to Wittgenstein*. Cambridge: Cambridge University Press.
- Tomasini Bassols, A. (2006). *Filosofía y Matemáticas. Ensayos en torno a Wittgenstein*. México, D.F./Madrid: Plaza y Valdés Editores.
- Wittgenstein, L. (2001). *Philosophical Investigations*. The German Text, with a Revised English Translation. Translated by G. E. M. Anscombe. Oxford, UK: Blackwell Publishing.
- Wittgenstein, L. (1983). *Remarks on the Foundations of Mathematics* (revised edition). G. H. von Wright, R. Rhees and G. E. M. Anscombe (Eds.). Translated by G. E. M. Anscombe. Cambridge, Massachusetts/London, England: The MIT Press.

# Metaabstract explanations in science

**Krzysztof Wójtowicz**  
Institute of Philosophy  
Warsaw University

One of the major problems in the philosophy of mathematics is the applicability of mathematics in natural science. This fact has led to interesting philosophical debates concerning, in particular, the existence of mathematical objects, and the truth value of mathematical statements. Quine's indispensability argument is central to the debate, however it faces some problems, and has been questioned frequently. In recent years, the original argument has been modified, and taken on the form of the Enhanced Indispensability Argument (EIA), where the explanatory virtues of mathematics are of primary importance (see e.g. [Colyvan 2001], [Baker 2005]).

The EIA might be viewed as a kind of “inference to the best explanation” account. It stresses the explanatory role of mathematics in science—and claims, that this very role allows us to argue for the existence of mathematical entities presupposed by our (best) explanations. The crucial assumption is, that mathematics plays a genuine explanatory role in science, i.e. that (at least a part of) the explanatory job is done by mathematics. Many authors have argued in favor of the thesis, that there are genuine, non-causal mathematical explanations: speaking in very broad terms, the truths of mathematics explain (some of) the phenomena by identifying abstract features of the system, not by describing the causal nexus or the detailed mechanisms in question. To mention just three (of many) examples: (1) The Borsuk-Ulam theorem explains, why there are two antipodal points on the surface of the Earth, where two physical parameters (say: temperature and pressure) are equal. (cf. [Baker 2005, 2009]; [Baker, Colyvan 2011]). (2) The problem of the periodical life-cycle of cicadas (13 and 17 years—which are primes), was introduced in [Baker 2005], and has been discussed extensively (here a fact in number theory is claimed to do the explanatory job). (3) The number of equilibria (four) of a double pendulum is explained by referring a theorem concerning continuous functions on a torus, [Lange 2013]. These explanations fall into the broad category of abstract explanations in science (for a discussion see e.g. [Pincock 2015], [Reutlinger, Andersen 2016]).

One of the ways of accounting for the explanatory role of mathematics is to interpret mathematical theorems as of a kind programming properties—which impose some kind of modal constraints on the world (cf. the presentation in [Lyon 2012]). This proposal is based on the idea (analogy; metaphor) of program explanation (the idea stems from [Jackson, Pettit 1990]). Generally speaking, some phenomena take place, because the world is constrained (“programmed”) in a certain way (and these “programs” are given by mathematics). A simple example illustrates this approach: if we run a computer to solve an unsolvable

problem, it will loop forever. The explanation is given here rather by mathematical results (i.e. the unsolvability of the problem in question—e.g. the halting problem), then by the laws of physics or chemistry. So, the programming analogy seems to be promising. The programming properties are a kind of (more general) abstract properties, which explain by a reference to mathematical properties of the system, not by tracing back the causal details or mechanisms. If mathematical theorems are considered to be modal constraints, the status of the background assumptions becomes crucial: if the constraint is based on strong, perhaps controversial mathematical assumptions, then its explanatory status becomes problematic.

The problem is not purely speculative. Most of results needed in applied mathematics can be proved under fairly standard and uncontroversial assumptions (and even reconstructed within a subsystem of second-order arithmetic  $Z_2$ , as results in Reverse Mathematics show). But abstract, set-theoretic principles might emerge in an unexpected way: independence phenomena within “ordinary mathematics” are well known; and the same applies even to some extensions of ZFC, as Friedman’s result’s show (cf. [Friedman 1981, 1986, 1998]). They provide a strong link between the seemingly abstract set-theoretic principles and the body of mathematical practice. Moreover, independence results concerning sentences having a physical interpretation, especially in models for relativity theory ([da Costa, Doria 1994, 1996], [Doria 2000]), but also in other subdisciplines of physics, have been obtained. Models for physics using strong set-theoretic techniques have been studied indeed (however, the question of the physical meaning of the models remains to be discussed)—so abstract set theory might have some bearing on the mathematical explanations. A different example is provided by the Topos Quantum Theory program—where a non-standard account (with some natural philosophical motivations) for quantum physics is proposed [Döring, Isham 2011]. It is not formulated within standard set theory, but within topos theory—but on a very general level, it exemplifies the same phenomenon: the possible role of very abstract principles in mathematical explanations (in this case it is the rejection of classical logic and working in a suitable topos). The problem become especially pressing from the point of view of mathematical realism, as these assumptions are considered not to have a hypothetical (or instrumental) character, but to possess truth values.

Set theory is focused on studying models (and even universes within a multiverse—cf. [Antos, Friedman, Honzik, Ternullo 2015])—so for this kind of explanation properties of models, not even abstract properties of particular systems would matter. So the term “metaabstract explanation”, not “abstract explanation” would better grasp the idea. It turns out, that seemingly abstract metatheoretical principle might affect our modeling and understanding phenomena. So the problem of the role of metaabstract explanations well-posed, and deserves a thorough discussion.

**Acknowledgment:** The preparation of this paper was supported by a NCN grant 2016/21/B/HS1/01955.

## References

- Antos C., Friedman Sy-D., Honzik R., Ternullo C., (2015). “Multiverse conceptions in set theory”, *Synthese*, 192, 2463–2488.
- Baker A. (2005). “Are there Genuine Mathematical Explanations of Physical Phenomena?”, *Mind*, 114(454), 223–238.
- Baker A. (2009). “Mathematical Explanation in Science”, *British Journal for the Philosophy of Science*, 60(3), 611–633.
- Baker A., Colyvan M. (2011). “Indexing and Mathematical Explantation”, *Philosophia Mathematica*, 19, 232–224.
- Colyvan M. (2001). *The indispensability of mathematics*, New York, Oxford University Press.
- Da Costa N.C.A., Doria F.A. (1994). “Suppes predicates an the construction of unsolvable problems in the axiomatized sciences”, w: Humpreys P. (red.) *Patric Suppes: scientific philosopher*. Kluwer Academic Publishers, 151–193.
- Da Costa N.C.A., Doria F.A. (1996). “Structures, Suppes Predicates, and Boolean-Valued Models in Physics”, w: Bystrov P.I., Sadovsky V.N. (red.), *Philosophical Logic and Logical Philosophy*, Kluwer Academic Publishers, 91–118
- Doria F.A. (2000). “Is there a simple, pedestrian arithmetic sentence, which is independent of ZFC?”, *Synthese*, 125, 69–76.
- Döring A., Isham C. (2011). “What is a thing? Topos Theory in the Foundations of Physics”, in: Coecke B. (Ed.) *New structures for physics*, Springer-Verlag, Lecture Notes in Physics (813), 753–937
- Friedman H. (1981). “On the necessary use of abstract set theory”, *Advances in Mathematics*, 41, 209–280.
- Friedman H. (1986). “Necessary uses of abstract set theory in finite mathematics”, *Advances in Mathematics*, 60, 92–122.
- Friedman H. (1998). “Finite functions and the necessary use of large cardinals”, *Annals of Mathematics*, 148, 803–893.
- Jackson F., Pettit P. (1990). “Program Explanation: A General Perspective”, *Analysis*, 50(2), 107–117.
- Lange M. (2013). “What Makes a Scientific Explanation Distinctively Mathematical?”, *British Journal for the Philosophy of Science*, 64(3), 485–511.
- Lyon A. (2012). “Mathematical Explanations of Empirical Facts, And Mathematical Realism”, *Australasian Journal of Philosophy*, 90(3), 559–578.
- Pincock C. “Abstract Explanations in Science”, *British Journal for the Philosophy of Science*, 66, 857–882.
- Reutlinger A., Andersen H., (2016). “Abstract versus Causal Explanations?”, *International Studies in the Philosophy of Science*, 30(2), 129–146.

# CIENCIA, TECNOLOGÍA Y SOCIEDAD



# Predecir *personas*: la opresión de género en los perfiles los *buyers* persona

Cristina Bernabeu

Mi objetivo es ofrecer, desde una perspectiva feminista, una explicación estructural del rol normativo que exhiben los artefactos tecnológicos en sus interacciones con los seres humanos; concretamente, de su rol normativo en tanto que *colaboradores* de relaciones de opresión en contextos androcéntricos. Para ello, presto especial atención al trabajo desarrollado por Sally Haslanger en torno a la estructura social y los procesos de objetivación sexual (Haslanger, 2017; 2016; 2012).

Esta propuesta pretende ofrecer una respuesta a uno de los asuntos más destacados dentro de la filosofía de la tecnología contemporánea: *el problema de la agencia tecnológica* (Illies & Meijers, 2009; Latour, 1994; 2005; Rosenberg, 2014; Verbeek, 2011, 2006). Rosenberg lo ha definido, por un lado, como un problema relacionado con la cuestión acerca de si los artefactos tecnológicos reducen y limitan o, por el contrario, ayudan y refuerzan la toma de decisiones de los usuarios; por otro, como un problema relacionado con el hecho de que los artefactos tecnológicos son al mismo tiempo entidades que construyen pero que también son construidas por *actores colectivos*. Asimismo, algunas propuestas de la filosofía de la tecnología contemporánea interpretan este rol “constructivo” en relación a su potencial para producir relaciones y condiciones imprevisibles. Gran parte de la ética del diseño gira en torno a la relevancia *moral* de este hecho.

Si bien todas las posturas en este debate (Verbeek, 2008, 2009, 2011; Peterson & Spahn, 2011; Illies & Meijers, 2009; Waelbers, 2009; Pols, 2013; Floridi & Sanders, 2004) comparten la hipótesis de los artefactos como mediadores activos que estructuran relaciones entre los agentes y los contextos materiales, existen dos posturas diferenciadas:

- (i) Las posiciones fuertes o Strong Views (SV) (Verbeek, 2011; 2005) suponen una radicalización de la tesis de la moralidad de los artefactos. Defienden que los artefactos tecnológicos tienen responsabilidad moral, como el resto de agentes morales.
- (ii) Las posiciones débiles o Weak Views (WV) (Illies & Meijers, 2009) sostienen que los artefactos no son susceptibles de ser considerados agentes morales, aunque sí tienen relevancia moral.

Sin negar la dimensión moral que pudiera tener dicha relevancia, mi objetivo es examinar su dimensión estructural; específicamente, la dimensión estructural de la “actividad” que exhiben los artefactos tecnológicos en tanto en cuanto que colaboran con la normatividad social de género en contextos androcéntricos.

Esta dimensión permite explicar los casos en los que hay opresión pero no intención o agencia. Al tratarse de contextos de desigualdad social, me centro en los artefactos como colaboradores de las relaciones de *opresión estructural* (Haslanger, 2012) que, a diferencia de las relaciones que se dan en el tipo de *opresión agencial* (definida a partir de las acciones que ejecutan individuos o grupos), son relaciones referidas a nuestras disposiciones colectivas. En este sentido, mi objetivo es abordar el problema de la agencia tecnológica desde una perspectiva estructural de tales disposiciones, tomando en consideración la posición social de los agentes en tanto que ocupan ciertos *nodos* dentro de la estructura social (Haslanger, 2016). En tanto que *nodos*, los individuos se relacionan con su rol social a partir de patrones regulares de compartidos (por los individuos ocupantes de un mismo nodo) de pensamiento y acción. Sostengo que los artefactos tecnológicos tienen un papel activo (si bien no agencial), estructurante de subjetividad, en dicha relación. En otras palabras, que los artefactos colaboran en la activación de disposiciones que son compartidas por los individuos que ocupan un mismo nodo dentro de la estructura social. Parto de que el *rol de colaborador*, expuesto por Haslanger en relación a los procesos de objetivación sexual (2012), es apropiado para explicar la actividad normativa que exhiben los artefactos tecnológicos, al (re)producir esquemas regulares sociales de pensamiento y acción [*Action Schemes* (Illies & Meijers, 2009); *Habitus* (Bourdieu, 1991; 1999)]; tal y como en contextos androcéntricos, en los que se observan patrones regulares de comportamiento masculinos (connotados positivamente) o femeninos (connotados negativamente).

A modo de ejemplo, desarrollo el caso de los *persona* (*buyers persona* o *customer persona*), utilizados como herramientas de predicción (normalmente en marketing, pero no sólo). Los *persona* son representaciones ficticias (masculinas o femeninas) configuradas a partir de personas reales que tienen como objetivo simular, imitar y predecir su comportamiento o disposición a partir de la recolección de cierta información sobre de ellas. En la medida en que son creadas como modelos de personas hiper-generalizadas, (re)producen la desigualdad estructural definitoria de los contextos androcéntricos. En este sentido, explican el papel activo de los artefactos como colaboradores de opresión estructural.

A su vez los *persona*, en cuanto que herramienta de marketing predictiva, se configuran a partir de masas de datos, sobre las que los algoritmos realizan las predicciones. “De manera probabilística, [el algoritmo] sospecha que una persona podría hacer tal o cual cosa que todavía no haya hecho, porque aquellas que se le parecen ya lo han hecho” (Cardon, 2018, pág. 45). A este respecto sostengo que, en la medida en que las regularidades en las que se basan los algoritmos son regularidades de carácter social, a partir de las cuales se predice comportamiento, éstos pueden ser explicados en términos de *esquemas de acción*, es decir, en tanto que esquemas regulares sociales de pensamiento y acción.

## Bibliografía

Bourdieu, P. (1977). *Outline of a theory of practice*. (R. Nice, Trad.) Cambridge, U.K. ; New York: Cambridge University Press.

- Bourdieu, P. (1987). Cosas Dichas. En P. Bourdieu, *Cosas dichas* (págs. 127–142). Barcelona: Gedisa.
- Bourdieu, P. (1991). *El sentido práctico*. Madrid: Taurus.
- Bourdieu, P. (1999). *Meditaciones pascalianas*. Barcelona: Anagrama.
- Bourdieu, P. (2000). *La dominación masculina*. (R. Nice, Trad.) Barcelona: Anagrama.
- Broncano, F. (2009). *La melancolía del ciborg*. Barcelona: Herder.
- Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, 1164–1181.
- Cardon, D. (2018). *Con qué sueñan los algortimos. Nuestras vidas en el tiempo de los big data*. (A. Davila Legerén, Trad.) Madrid: Dado.
- Fielding, L. (02 de 04 de 2016). *www.huffingtonpost.com..* Recuperado el 25 de Agosto de 2018, de How to Hack Your Habits: [https://www.huffingtonpost.com/lara-fielding/how-to-hack-your-habits\\_b\\_9154792.html](https://www.huffingtonpost.com/lara-fielding/how-to-hack-your-habits_b_9154792.html)
- Foucault, M. (2008). *Tecnologías del yo y otros textos afines*. (M. Allendesalazar, Trad.) Barcelona; Buenos Aires; México: Paidós Ibérica.
- Fox Keller, E. (1991). *Reflexiones sobre género y ciencia*. Valencia: Alfons el Magnànim, Institució Valenciana d'Estudis i Investigació.
- Gallagher, M. (2014). Media and the representation of gender. En C. Carter, L. Steiner, & L. McLaughlin, *The Routledge Companion to Media and Gender* (págs. 23–31). London; New York: Routledge.
- Golbeck, J., Koepfler, J., & Emmerling, B. (2011). An Experimental Study of Social Tagging Behavior and Image Content. *Journal of the American Society for Information Science and Technology*, 62(9):1750–1760.
- Gray, K. L. (2017). The Internet: Opression in Digital Spaces. En C. P. Campbell, *The Routledge Companion to Media and Race* (págs. 107–115). New York: Routledge.
- Haraway, D. (2004). *Testigo\_Modesto@Segundo\_Milenio.HombreHembra\_Conoce\_Oncorotón*. UOC.
- Haslanger, S. (2012). *Resisting reality: social construction and social critique*. New York: Oxford University Press.
- Haslanger, S. (2016). What is a (social) structural explanation? *Philosophical Studies*, 173(1):113–130.
- Haslanger, S. (2017). Failures of Individualism: Materialism. *Draft Version*.
- Haslanger, S. (2017). Objectivity, Epistemic Objectification, and Oppression. En j. Kidd, J. Medina, & G. Pohlhaus, *The Routledge handbook of epistemic injustice*. London; New York: Routledge, Taylor & Francis Group.

- Illies, C., & Meijers, A. (July de 2009). Artefacts Without Agency. *The Monist*, 92(3):420–440.
- Latour, B. (1994). On technological mediation: Philosophy, sociology, genealogy. *Common Knowledge*, 29–64.
- Latour, B. (2005). *Reassembling the social : an introduction to actor-network-theory*. Oxford: Oxford University Press.
- Longino, H. (1990). *Science as Social Knowledge : Values and Objectivity in Scientific Inquiry*. Princeton UP: Princeton.
- Neff, G., & Nafus, D. (2016). *Self-Tracking*. Cambridge, Massachusetts: The MIT Press.
- Parisier, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. New York: The Penguin Press.
- Reckwitz, A. (2002). Toward a Theory of Social Practices: A Development in Culturalist Theorizing. *European Journal of Social Theory*, 5(2):243–263.
- Rettberg, J. W. (2014). *Seeing Ourselves through Technology: How We Use Selfies, Blogs, and Wearable Devices to See and Shape Ourselves*. London: Palgrave Macmillan UK.
- Rosenberg, R. (2014). Multistability and the Agency of Mundane Artifacts: from Speed Bumps to Subway Benches. *Human Studies*, 37(3):369–392.
- Rubin, G. (1986). The Traffic of Women: Notes on the Political Economy of Sex. *Nueva Antropologia*, 95–145.
- Sirois, F. M. (2016). Perfectionism and Health Behaviors: A Self Regulation Resource Perspective. En F. M. Sirois & D. S. Molnar, *Perfectionism, Health, and Well-Being* (págs. 45–67). Switzerland: Springer.
- Verbeek, P.-P. (May/Jun de 2005). Beyond Interaction: A Short Introduction to Mediation Theory. *Interactions*, 22(3):26–31.
- Verbeek, P.-P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. University Park, Pa.: Pennsylvania State University Press.
- Verbeek, P.-P. (2006). The Morality of Things: A Postphenomenological Inquiry. En E. Selinger, *Postphenomenology: a critical companion to Ihde* (págs. 117–128). Albany: State University of New York Press.
- Verbeek, P.-P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago, IL: University of Chicago Press.
- Wajcman, J. (2006). *El tecnofeminismo*. Madrid: Cátedra.
- Young, N. (2012). *The virtual self: how our digital live are altering the world*. Toronto: McClelland & Stewart Ltd.
- Zafra, R. (2010). *Un cuarto propio conectado: (Ciber)espacio y (auto)gestión del yo*. Madrid: Fórcola.

Zafra, R. (2015). La censura del exceso. Apuntes sobre Imágenes y Sujeto en la Cultura-Red. *Paradigma*, 1-5.

Zafra, R. (2015). *Ojos y capital*. Bilbao: Consonni.

# Antropoceno y praxeología de la cultura científica

**José Manuel de Cózar Escalante**

Departamento de Historia y Filosofía de la Ciencia, la Educación y el Lenguaje  
Universidad de La Laguna  
jcozar@ull.edu.es

El Antropoceno es una nueva época en la historia geológica y humana. A falta de confirmación oficial por parte de los geólogos, cada vez hay mayor consenso académico y popular en que ya hemos entrado en ella. Esta “época del hombre”, o mejor, “época de la humanidad”, se caracteriza por la capacidad de nuestra especie para transformar el planeta en su totalidad, a la manera de un agente geológico que manifiesta un enorme poder (de Cózar, 2017). Los efectos de la acción humana sobre el sistema Tierra son muy variados: contaminación, destrucción de ecosistemas, extinción, transformación física del terreno a gran escala, alteraciones en los ciclos químicos, etc. Sin embargo, el más destacable –y tal vez el más preocupante– es el cambio climático. La alteración del clima causa un impacto global. Además, intensifica los otros cambios de origen antropogénico y, a su vez, puede ser agravada por ellos.

Está naciendo una interesante cultura científica del Antropoceno, en la que se hallan implicadas variadas formas de presentar (*framing*) y de comunicar a la ciudadanía la información científica necesaria para comprender esta situación sin precedentes en la historia de la humanidad. Ya se están generando nuevas percepciones y actitudes sobre nuestra relación con los sistemas biofísicos planetarios, junto con distintas “narrativas” o “relatos” sobre el significado del Antropoceno (Arias, 2018; Hamilton, Bonneuil, y Gemenne, eds. 2015). En función de estos y otros factores se producirán nuevos comportamientos y cursos de acción con objeto de adaptarse al Antropoceno y, lo que es más importante, a fin de encauzarlo en la medida de lo posible.

La praxeología de la cultura científica consiste en la investigación de los elementos y dimensiones de la cultura científica que tienen que ver con la praxis. Más en concreto, trata de la creación o modulación de las disposiciones comportamentales y de la acción humana en respuesta a la adquisición de información científica. Un ejemplo muy claro serían los cambios que un individuo decide efectuar en su dieta para prevenir ciertas enfermedades (exceso de colesterol, diabetes, cáncer, etc.). En esta comunicación se propone comenzar a elaborar una praxeología de la cultura científica específica del Antropoceno.

Una primera cuestión de índole praxeológica, muy amplia y hasta cierto punto especulativa, es la que queda formulada como sigue: ¿Quién es el (principal) agente del Antropoceno? ¿El ser humano o la Tierra? ¿la especie *Homo sapiens*? ¿las sociedades modernas? ¿algunos sectores de la población? ¿Es más bien un nuevo “híbrido” formado por la combinación de nuestra agencia y la agencia

natural? Esta pregunta por la agencia del Antropoceno, que parece revestir un interés puramente abstracto, circunscrito a la ontología y a la antropología filosófica, puede tener sin embargo un impacto en la praxis cotidiana de los ciudadanos: si, por razones científicas o incluso ideológicas, piensan que no tienen capacidad de acción o responsabilidad en la nueva condición antropocénica del mundo, podrían abstenerse de actuar para mejorar la situación, continuar con sus prácticas habituales. En cambio, si consideran que la agencia humana, al menos en el plano colectivo, es poderosa, y que tienen alguna clase de responsabilidad por el estado del planeta, su acción puede ser decidida y no carecer de efectividad.

Ahora bien, a fin de acotar más el análisis en esta comunicación, destacaremos dos temas que, por su relevancia, deben ser abordados por dicha praxeología. En primer lugar, el impacto de las tecnologías digitales (particularmente internet) en la diseminación entre la ciudadanía de conocimientos científicos y de interpretaciones sobre el Antropoceno (y del cambio de comportamiento relacionado con ello); en segundo lugar, la peculiaridad que supone que los cambios en los hábitos y en los estilos o formas de vida que la existencia en el Antropoceno requiera vean sus efectos positivos solo a largo plazo, superando con mucho la expectativa de vida de los individuos que lleven a cabo tales cambios.

Desde hace bastantes años es notoria la transformación que las tecnologías digitales han operado en los patrones de adquisición y transmisión de información del tipo que sea por parte de los ciudadanos (Trench, 2008). La discusión sobre el Antropoceno y, sobre todo, su “efecto estrella” (el cambio climático) ha saltado desde los contextos científicos especializados hasta los medios de comunicación de masas, desde la esfera técnica a la política; ha pasado de los foros académicos a los sociales, de la cultura científica a la cultura popular. Por todos es conocida la naturaleza enrevesada de las “controversias” sobre los orígenes y efectos del cambio climático, sobre las medidas a tomar para prevenirlo, mitigarlo o adaptarnos a él. Para muchos expertos resulta preocupante la incidencia de los factores ideológicos, políticos y económicos en lo que se supone debería ser un debate puramente racional, basado tan solo en los datos técnicos. Con el Antropoceno este problema de la interpretación de la información técnica con vistas a la toma de decisiones es mayor si cabe. Incluye el cambio climático entre otros muchos efectos, amplía el periodo en el que los actos humanos continuarán ejerciendo su influencia sobre el decurso planetario, acrecienta los riesgos y eleva la incertidumbre. La complejidad de la información que es preciso manejar, junto a la posibilidad de acceso inmediato a la misma gracias a Internet y al resto de tecnologías digitales, nos lleva a la conveniencia de replantear la alfabetización científica ciudadana sobre el Antropoceno (y, más en general, la adquisición de una cultura científica relativa al mismo). Hay que transitar desde un modelo de “almacenamiento”, en el que los individuos almacenan mentalmente gran cantidad de datos científicos durante su formación a otro “just-in-time” (“justo a tiempo” o mejor, “a demanda”), en el que se obtiene la información a medida que se va precisando (Miller, 2014). Ahora bien, para que dicha información cobre sentido a ojos de los ciudadanos se requiere la formación de unos constructos que actúen como herramientas que permitan organizar, interpretar y otorgar significado a la información a lo largo del tiempo (Miller, 2014). La noción de Antropoceno trata primariamente de un proceso de cambio a lo largo de siglos y milenios. Las dificultades que encuentran la mayoría de los ciudada-

nos para comprender o imaginar cambios a escalas temporales y espaciales tan dilatadas deben ser compensadas con una educación que incida en herramientas conceptuales y simbólicas que permitan afrontar tales cambios tanto en el plano intelectual como en el de la praxis, ligando ésta a efectos positivos a más corto plazo.

En tal sentido, se están generando unas narrativas o relatos sobre el Antropoceno: van desde los más catastrofistas hasta los más optimistas (Arias, 2018; Fernández, 2011, Stiegler, 2018). Los primeros (los “misanthropocénicos”) pintan un futuro apocalíptico ante el que nos hallaríamos impotentes; en otras versiones de esta narrativa, la perspectiva de dicho futuro debe asustarnos tanto como para llevarnos a tomar medidas drásticas en el presente que permitan cuando menos atenuar sus aspectos más siniestros. En contraste con el mencionado punto de vista, los escenarios optimistas, propuestos por quienes cabría denominar “antropocenistas”, defienden la posibilidad de un “Antropoceno bueno”, es decir, un mundo controlado hasta en sus más pequeños detalles para el bienestar de nuestra especie. Según esta interpretación, gracias a tecnologías como la geoingeniería, la gestión de los ecosistemas, la manipulación genética o la inteligencia artificial, seremos capaces de aprovechar la oportunidad que nos plantea el Antropoceno para hacer de nuestro planeta un mundo mucho mejor. Cada narrativa propone un conjunto de prácticas distintas y a menudo, incompatibles entre sí.

En todos los casos, sin embargo, se plantea la cuestión de que las soluciones propuestas y los cambios en la forma de vivir deben ser colectivos y tener repercusiones a largo plazo. Habría que poner en marcha las medidas de inmediato, cambiar ya nuestros hábitos. El objetivo es el de disminuir la probabilidad de que en el futuro nuestros descendientes padezcan las peores consecuencias de unas malas prácticas, prácticas que verosímilmente conducirían a hacer realidad la versión más sombría del Antropoceno. Ello supone una convicción ético-política profunda, tanto en el plano individual como en el colectivo. Por si esta exigencia no fuera lo bastante elevada, las narrativas catastrofistas pueden conducir a la parálisis a muchas personas, demasiado atemorizadas por las predicciones como para desear comprender a fondo las consecuencias del Antropoceno. Nada cambiarían en su existencia cotidiana aquellas personas que se sintieran impotentes ante la magnitud del problema, escépticas frente a la posibilidad de que su acción surtiera algún efecto significativo.

En la actualidad se dispone de numerosas evidencias, desde la psicología y la sociología, acerca de los motivos que conducen a ciertos individuos a negar la información científica que reciben sobre temas climáticos o cuando menos a ignorarla. La “paradoja psicológica del clima” (*psychological climate paradox*): a pesar de que en las últimas décadas la ciencia del clima ha proporcionado datos y modelos cada vez más fiables, la preocupación pública por el cambio climático ha ido disminuyendo, y en todo caso, la importancia del tema es mejor que la de otros problemas. Esto se traduce en un menor apoyo a las políticas sobre el clima ambiciosas (Stoknes, 2014). Se han ofrecido diversas explicaciones para las actitudes de escepticismo, negación, indiferencia y apatía en relación a los problemas ambientales globales (y especialmente en los países desarrollados con respecto al clima): la percepción de una gran distancia temporal y espacial, la ausencia de tratados políticos globales (o su insuficiencia e ineffectividad), la



priorización del desarrollo económico y de los asuntos personales inmediatos, las crisis financieras, la complejidad científica de los problemas, que propicia los errores y confusiones por parte de los ciudadanos, los filtros culturales, la disonancia cognitiva, la reacción ante un discurso en ocasiones “catastrofista” o “apocalíptico”, la responsabilidad individual limitada, el sentimiento de impotencia ante la magnitud del problema, las poderosas campañas en contra y las estrategias psicológicas de evitación del miedo y la culpa. (Stoknes, 2014, 2015; Norgaard, 2011). Sea como fuere, no parece que la falta de respuesta pública pueda resolverse tan solo con el incremento de la información, la comunicación de una mayor cantidad de hechos científicos y de datos técnicos (el viejo modelo del déficit). En cambio, sí que puede recurrirse a varias estrategias (psicológicas y sociales) para hacer más efectiva, sobre todo en términos de respuesta conductual, la comunicación de los problemas climáticos (Stoknes, 2014). De ahí la importancia de ahondar en estos planteamientos a fin de revertir la inacción de aquellos grupos de ciudadanos que, conociendo el problema, no se sienten movidos a actuar de manera significativa para encararlo consistentemente (Norgaard, 2011). Si con el cambio climático la falta de respuesta pública viene detectándose desde hace años, la paulatina popularización del concepto de Antropoceno arroja más incertidumbre si cabe sobre la acción ciudadana en un escenario de crisis ecosocial global (Hamilton, Bonneuil, y Gemenne, eds. 2015).

Con todos estos fenómenos de adquisición e interpretación de la información científica, así como de eventual cambio en las conductas de los ciudadanos, el Antropoceno se puede estar transformando ya en un tema a “negociar” culturalmente. Se difuminan los límites entre la producción científica de conocimientos y el público; también se produce una combinación de lo epistémico y de lo normativo. Como con el cambio climático en sentido estricto, sucede que se están cuestionando las clases de evidencia que han de ser investigadas y aceptadas<sup>1</sup>. Los ciudadanos han de poder participar con garantías en el debate sobre estas y otras cuestiones, de modo que se mejore la generación de conocimientos, la comunicación de los mismos, su comprensión pública y, en definitiva, el proceso político de toma de decisiones (López Cerezo, 2017).

## Referencias

- ARIAS, M. (2018): *Antropoceno. La política en la era humana*, Madrid, Taurus.
- DE CÓZAR, J. M. (2017): “¿Qué significa vivir en el Antropoceno?”, Iberoamérica divulga, octubre. Disponible en: <http://www.oei.es/historico/divulgacioncientifica/?que-significa-vivir-en-el-antropoceno>
- FERNÁNDEZ, R. (2011): *El Antropoceno. La expansión del capitalismo global choca con la biosfera*, Barcelona, Virus editorial.
- HAMILTON, C., BONNEUIL, C y GEMENNE, F. (eds.) (2015): *The Anthropocene and the Global Environmental Crisis*, Londres, Routledge.

---

<sup>1</sup>Véase, entre otros, el sitio web del taller ‘Anthropocene and Citizen Science: Evidence Gained through the “Opening-up” of Academic Knowledge Production?’ [http://www.carsoncenter.uni-muenchen.de/events\\_conf\\_seminars/calendar/180719\\_anthropo-n-citizen-sci/index.html](http://www.carsoncenter.uni-muenchen.de/events_conf_seminars/calendar/180719_anthropo-n-citizen-sci/index.html)

- LÓPEZ CEREZO, J. A. (2017): *Comprender y comunicar la ciencia*, Madrid, Libros de la catarata-OEI.
- MILLER, J. (2014): “La importancia de la alfabetización científica cívica en un mundo just-in-time”, en B. Laspra y E. Muñoz, coordinadores: *Culturas científicas e innovación. Progreso social*, Buenos Aires, Eudeba, pp. 73–99.
- NORGAARD, KARI MARIE (2011): *Living in denial. Climate change, emotions, and everyday life*, Cambridge, MA., The MIT Press.
- STIEGLER, B. (2018): *The Neganthropocene*, Londres, Open Humanities Press.
- STOKNES, PER ESPEN (2014): “Rethinking climate communications and the ‘psychological climate paradox’ ”, *Energy Research and Social Science*, 1, pp. 161–170.
- STOKNES, PER ESPEN (2015): *What we think about when we try not to think about global warming. Toward a new psychology of climate action*, White River Junction, Vermont, Chelsea Green Publishing.
- TRENCH, B. (2008): “Internet. Turning Science Communication Inside-Out?”, en Bucchi, M. & Trench, B. (Eds.). *Handbook of public communication of science and technology*. Routledge.

# A new framework for thinking about life extension

Pablo García-Barranquero

In the last two decades, the possibility of extending human lifespan has been a highly debated topic by both biomedical scientists (de Grey and Rae, 2007; Olshansky and Carnes, 2002) and philosophers (Agar, 2010; Overall, 2003). I propose an alternative framework to clarify different categories in the theoretical literature. This model builds upon the distinctions that Juengst and collaborators (2003), and Wareham (2016) make about the levels of human senescence and means to control it. In addition, I incorporate the classic perspective of Callahan (Stock and Callahan, 2004) as a different counterpoint. Besides this, I have to introduce some notions which are basic to understand my proposal.

Firstly, I compare these previous approaches with my own perspective:

| Eric Juengst et al. (2003) | Daniel Callahan (2004)                              | Cristopher Wareham (2016)              | My own model                                                                    |
|----------------------------|-----------------------------------------------------|----------------------------------------|---------------------------------------------------------------------------------|
| Prolonged Senescence       | X                                                   | X                                      | <b>Prolonged Senescence</b>                                                     |
| Compression of Morbidity   | The Natural Progress Model<br>The Normalizing Model | X                                      | <b>Compression of Morbidity</b>                                                 |
| Decelerated Aging          | The Optimizing Model<br>The Maximizing Model        | Slowed Aging                           | <b>Decelerated Aging</b><br><b>First Level of Negligible Senescence or SENS</b> |
| Arrested Aging (A)         | X                                                   | SENS                                   | <b>Arrested Aging (A)</b><br><b>Second or Complete Level of SENS</b>            |
| X                          | X                                                   | Arrested Aging (B)<br>Cryopreservation | <b>X</b>                                                                        |
| X                          | X                                                   | Reverse Aging                          | <b>Rejuvenation</b>                                                             |
| X                          | X                                                   | Escaping Aging<br>Mind Uploading       | <b>X</b>                                                                        |

This model comprises the following mix of conceptual categories:

- Prolonged Senescence:** involves the prolongation of life without attention to age-related diseases and pathologies or the health span.
- Compressed of Morbidity:** involves the prolongation of life without the burden of lifetime illness since it can be confined to a shorter period

before the time of death at the same time that the age of onset of the first chronic infirmity can be postponed a little later. The objective is not to delay aging but rather to extend the health span. It is applied to age-related diseases like Alzheimer or cancer, and age-associated pathologies like muscle-wasting. On Callahan's scheme, this model is represented by two categories. He describes The Natural Progress Model: "trying to understand the aging process in the same way that we are doing in current medicine. In favor of improving the quality of life of the elderly (...)" (Stock and Callahan, 2004, 56). Then, he offers another scenario, which he labeled as The Normalizing Model: "trying to reduce premature death and trying that more people reach the 85 years that Japanese women have maximum life expectancy" (Stock and Callahan, 2004, 556). I consider the two levels of Callahan at the same level as he focuses on improving the quality of life. Nevertheless, if we live healthier, we will extend our live expectancies a couple years (perhaps to the 85 year life expectancy of contemporary Japanese women).

- c) **Decelerated Aging:** is what de Grey has labeled Strategies for Engineered Negligible Senescence or SENS (version 1.0 or beta phase): it involves the prolongation of life with less degradation of the cognitive and physical function of the organism due to the effects of aging. A decelerated aging clock would still move through all the normal phases of senescence, merely at a slower rate, extending life expectancy and health span. This is the same as de Grey (2004) is talking about when he describes the type of life extension that will be required to reach what he calls Longevity Escape Velocity (LEV). On his view, we only need to slow down aging long enough for science to mature enough to take us into the next model. De Grey claims that LEV will provide (SENS 1.0) 20 additional years from the increase of 30 % healthy life. On Callahan's scheme, this model is represented by two categories. He describes The Optimizing Model as: "trying to get more people to reach the 122 years that Jeanne Calment lived. It is a realistic goal in so far as there have been super-centenarians over time" (Stock and Callahan, 2004, 556). Then, he offers a more hypothetical scenario, which he labeled as The Maximizing Model: "trying to extend the double of our life expectancy" (Stock and Callahan, 2004, 556). Obviously, The Optimizing Model seems more feasible. However, we could imagine a life of about 160 years without eliminating aging completely.
- d) **Arrested Aging:** (A in Juengst's model and in my own model) or SENS (version 2.0 or ultimate phase): involves the complete control of the aging process. This is the same as de Grey (2004) postulated: LEV will improve the limits or errors until it is a definitive set of therapies against aging. On this category, a person born after the availability of arrested aging interventions might live 1,000 years without senescence, as long as they avoid other causes of death, such as dehydration, fatal disease, or starvation.
- e) **Rejuvenation:** involves not only the complete elimination of the aging process but it is also capable of turning back our biological clock to early biological stages. We will enjoy a full cellular and molecular health and our aesthetic and physical appearance will be returned to what we look like at whatever age we choose, whether that is as a more robust and active

elder, a mentally mature young adult, or a growing adolescent, or even an immature child. One could then turn on and off” one’s aging process as one wished, or even return one’s body to previous ages (to try some new lifestyle or somatic fashion).

Now, I present four reasons why my own model differs from the others that I have described in this section (Juengst et al., 2003; Wareham, 2016).

First, no one advocates more research in order to perpetuate a state of **Prolonged Senescence** that would merely extend the lives of the old without mitigating the degenerative effects of aging (Glannon, 2009). Thereupon, no one endorses the social consequences that would result from what Fukuyama calls the “national nursing home scenario”, “in which people routinely live to be 150 but spend the last fifty years in a state of childlike dependence on caretakers” (2002, 69).

Second, unlike Wareham (2016), I also think that a comprehensive model of life extension should include **Compression of Morbidity**. The possibility of extending life can also be achieved if we are able to control (and even, eliminate) some or all of the age-related diseases. It is one thing to fight against aging and another is to try to counteract the virulence of Alzheimer or some types of cancer.

Third, unlike Juengst and collaborators (2003), I also think that a comprehensive model of life extension should include *Rejuvenation*. The possibility of eliminating aging and a maximizing health span are in the gerontological literature but a person may ask: if SENS arrives when I am older, who will I see when I look in the mirror?

At the other end of the taxonomy, I also ignore that a comprehensive model of life extension should include Arresting Aging (B in Wareham’s model) and **Escaping Aging** (or Death) (Wareham, 2016) entirely, by achieving immortality or resurrection. For two reasons: (1) it is a utopia with almost no scientific basis that serves more for projecting philosophical speculations than other things; (2) the only way to achieve these goals would be to forego our physical bodies entirely, such as by integration (partial or complete) with the machine; for example, mind uploading in the last level. This differs from the nature of the other levels, and raises issues outside the scope of the debate over biomedical life extension.

Finally, the model that I am proposing in the table above builds on these categories, but improves upon Juengst and collaborators (2003) and Wareham (2016) models in several ways:

**Flexibility:** it is open to add or restructure categories with the evolving of gerontology. Its maxim is to focus on aging and senescence; therefore, there are not all the possibilities to live longer (or forever). **Overlap:** there is a continuity between each of the categories since there is no substantial gap between them. For example, between Compression of Morbidity and Decelerated Aging (we may think that we have delayed senescence but we have only slowed down the appearance of age-related diseases) or if de Grey (2004) develops his entire research project (from SENS 1.0 to SENS 2.0). **Levels of fantasy:** is intermingled with established science in this model. It spans medicine’s current efforts to prevent late-life disease and the craziest ideas of charlatans in this field. Alt-

though categories (a, b, and c) is, understandable in terms of the status quo and ongoing science, it takes a strong tolerance for fantasy to envision what is meant categories (d and e). The realistic aspiration does not consist of eliminating aging or reversing aging (Bostrom, 2005; de Grey and Rae, 2007; Harris, 2007; Kurzweil, 2005).

## References

- Agar, N. 2010. *Humanity's End: Why We Should Reject Radical Enhancement*. Cambridge, MA: MIT University Press.
- Bostrom, N. 2005. The Fable of the Dragon Tyrant. *Journal of Medical Ethics*, 31(5):273–277.
- de Grey, A. 2004. Escape Velocity: Why the Prospect of Extreme Human Life Extension Matters Now. *PLoS Biology*, 2(6):726–730.
- de Grey, A. and M. Rae. 2007. *Ending Aging: The Rejuvenation Breakthroughs that Could Reverse Human Aging in Our Lifetime*. St. Martin's Press: London.
- Fukuyama, F. 2002. *Our Posthuman Future: Consequences of the Biotechnology Revolution*. Baltimore: Johns Hopkins University Press.
- Glannon, W. 2009. Decelerating and Arresting Human Aging. In: *Medical Enhancement and Posthumanity* (pp. 175–190). B. Gordijn and R. Chadwick (eds.). Dordrecht: Springer.
- Harris, J. 2007. *Enhancing Evolution: The Ethical Case for Making Better People*. Princeton: Princeton University Press.
- Juengst, E. T. R. H. Binstock, M. J. Mehlman, S. G. Post, and P. Whitehouse. 2003. Biogerontology, “Anti-aging Medicine,” and the Challenges of Human Enhancement. *Hastings Center Report*, 33(4):21–30.
- Kurzweil, R. 2005. *The Singularity is Near*. New York: Vikings.
- Olshansky, S. J. and B. A. Carnes. 2002. *The Quest for Immortality*. New York: W. W. Norton & Company.
- Overall, C. 2003. *Aging, Death, and Human Longevity: A Philosophical Inquiry*. Berkeley: University of California Press.
- Stock, G. and D. Callahan. 2004. Debates: Point-Counterpoint: Would Doubling the Human Life Span Be a Net Positive or Negative for Us Either as Individuals or as a Society? *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59(6):554–559.
- Wareham, C. 2016. The Transhumanist Prospect: Developing Technology to Extend the Human Lifespan. In *The Palgrave Handbook of the Philosophy of Aging* (pp. 517–538). G. Scarre (ed). London: Palgrave Macmillan.

# El papel de las series de televisión en la democratización de la ciencia

Laura García Díaz  
Abraham Hernández Pérez  
Universidad de La Laguna<sup>1</sup>

La filosofía de la ciencia desarrollada en el siglo XX se caracteriza, sobre todo, por haber introducido en sus análisis un elemento hasta el momento ignorado: lo histórico. Esto abrió un panorama completamente nuevo y dio lugar en los años 70 y 80 a los estudios sobre la ciencia desde un punto de vista sociológico, antropológico, económico, de género o de políticas científicas. A diferencia de lo que había estado haciendo la filosofía de la ciencia tradicional, centrada mayoritariamente en aspectos metodológicos o normativos, estas nuevas tendencias vuelcan sus análisis sobre todo en la práctica científica. Esta dimensión práctica de la ciencia, y todo aquello con lo que se relaciona, es reflexionada, entre otros y otras, por Philip Kitcher, el cual defiende la necesidad de comprender el papel de la ciencia en una sociedad democrática. Kitcher se suma así al debate en torno a la “democratización de la ciencia” abierto por Paul Feyerabend en los años 70. Uno de los elementos centrales dentro de dicho debate es el papel del sujeto. En relación con la desafección pública acerca de las direcciones y consecuencias de ciertas investigaciones científicas comentó Feyerabend en su obra *La ciencia en una sociedad libre* (1978): “Si el asunto es importante, bien para un pequeño grupo o para la sociedad como un todo entonces este criterio debe someterse al escrutinio más minucioso”. Aunque con una perspectiva y actitud distinta, Kitcher también insistirá en la necesidad de que el público, en un sentido amplio, se involucre en la práctica científica.

Siguiendo a Robert Dahl, Kitcher señala la imposibilidad de proporcionar a la ciudadanía de las democracias contemporáneas una medida de control sobre las decisiones que les afectan y que respondan a su esquema de valores. La democracia, que surge como respuesta al problema de la opresión, se encuentra ahora ante la imposibilidad de representar por completo los intereses de su ciudadanía, más aun teniendo en cuenta la persistencia de opresiones difícilmente identificables. Gran parte de las decisiones que afectan a la población son demasiado complejas para que los individuos formen una opinión responsable al respecto. Además, los representantes políticos no tienen posibilidad, según Kitcher, de saber si las políticas que incentivan responderán a las necesidades de la población, una situación que se complica aún más en la comunidad científica. De esta manera, Kitcher nos propone pensar modos en los que instituciones alternativas puedan permitirnos llegar a su ideal de “ciencia bien organizada”, un modelo que depende de deliberaciones basadas en preferencias que responden

---

<sup>1</sup> Este trabajo ha sido realizado en el marco del proyecto *Puntos de vista, disposiciones y tiempo. Perspectivas en un mundo de disposiciones* (FFI2014-57409-R), financiado por el Ministerio de Economía y Competitividad del Gobierno de España).

tanto a la comprensión de las oportunidades disponibles como a las necesidades expresadas.

Para Kitcher es un error dar por hecho el valor intrínseco de la ciencia para comprender nuestro mundo si no se pone antes de manifiesto que aquello que denominamos “conocimiento puro” sólo está disponible para un número reducido de personas y el poco interés por hacer accesible a todos “nuestra herencia científica”. Es por esto que en esta comunicación se propone resaltar el papel democratizador de ciertos medios de producción cultural que pueden permitir al público acercarse a algunos problemas científicos o, en términos de Kitcher, que permitan formar una opinión responsable sobre conocimiento científico y sus posibles aplicaciones tecnológicas. La propuesta de ciencia bien organizada exige, entre otras cosas, que la defensa del conocimiento puro provenga de deliberadores que comprendan las diferentes opciones y que incorporen las necesidades de los otros. Este aspecto permite analizar el valor de algunas series de televisión como un medio para instruir al público, podemos ver en ellas un medio de democratización del conocimiento.

Desde hace algunos años ha surgido un nuevo panorama de fragmentación de audiencia televisiva en el que aparece un lugar específico para un perfil sofisticado de espectador, al que le interesa la innovación genérica, las temáticas actuales y la representación sin prejuicios de ciertos tabúes. Este espectador recibe, a menudo con gran entusiasmo, contenido científico y filosófico que por lo general parece más asequible y atractivo que el contenido ofrecido por las oportunidades de aprendizaje presentadas en el ámbito académico. Las series aparecen como un aparato divulgador del conocimiento al mostrar y cuestionar algunos de los desarrollos científicos y tecnológicos más controvertidos. El déficit de democratización de la ciencia señalado por Kitcher puede ser de cierta forma nivelado a través de las series, éstas son capaces de crear narraciones y transmitir el contenido que debería ser divulgado por la propia comunidad científica o política. Se trata de un medio capaz de proyectar el resultado de ciertas investigaciones científicas, es decir, más allá de llevar al espectador un conocimiento que le resulta desconocido, es capaz de incentivar el cuestionamiento de las prácticas y fines en la ciencia. En un contexto en el que la ciencia se ha privatizado, las series pueden servirnos como un recurso para tomar conciencia, esto es, ejemplificar, problematizar y generar opiniones sobre los desarrollos científicos actuales y sus repercusiones en el futuro –tal y como entiende Kitcher que sería necesario en una sociedad democrática–. Por ejemplo, la inteligencia artificial, que está aún en pleno desarrollo, ya existe en el plano del imaginario audiovisual y contamos con múltiples proyecciones y problematizaciones al respecto.

En definitiva, se propone analizar la posibilidad de encontrar en las series televisivas una forma de democratización o, al menos, de divulgación del conocimiento. Series actuales como *Black Mirror*, *Master of Sex* o *The Knick* ofrecen al espectador la oportunidad de reflexionar sobre temas que por lo general no le resultan cercanos, como el desarrollo tecnológico ligado al transhumanismo o los problemas morales que surgen del desarrollo científico. Mediante este producto cultural vemos cómo se representan controversias, repercusiones positivas y negativas, la relación entre ciencia y política o incluso el propio desarrollo histórico de la ciencia. Se trata de un medio audiovisual capaz de proyectar diferentes puntos de vista que pueden resultar relevantes para que los individuos



se formen una opinión responsable.

## Referencias bibliográficas

- Cascajosa, C. (2016), *La cultura de las series*, Laertes, Barcelona.
- Feyerabend, P. (1978), *Science in a free society*, NLB, Great Britain.
- Kitcher, Ph. (2001), *Science, truth, and democracy*, Oxford University Press.
- Kitcher, Ph. (2010), “Science in a Democratic Society”, en González, W.J. (ed), *Scientific Realism and Democratic Societies: The Philosophy of Philip Kitcher*, Poznan Studies in the Philosophy of Science and the Humanities, Rodopi, Amsterdam, pp. 95–112.
- Newman, M. y Levine, E. (2012), *Legitimizing television: Media convergence and cultural status*, Nueva York, Routledge.

# Socially responsible philosophy of science

Inmaculada Perdomo<sup>1</sup>

University of La Laguna  
mperdomo@ull.es

Science, directed at the achievement of collective goals, fulfils the principal aim of responding to relevant and significant questions (Kitcher 2001; van Fraassen, 2008). But these questions are dependent upon context and decidable within the framework of democratic societies. Traditional efforts to design a notion of epistemic significance independent of contextual restrictions, free from social and moral values, and demanding that the obtaining of epistemically significant truths is valuable in principle, is an undertaking impossible: moral and social values are inherent to scientific practice. Nor is it possible to defend the idea that all sciences can be unified in a hierarchical chain, or the vision that the integration of all theoretical proposals in a unified scheme is the essence of the objective understanding of our world. In fact, the image of a patchwork of theories, laws and disciplines without a hierarchical order or systematic relationship defended by N. Cartwright (1999) seems a more appropriate vision of the structure of science.

Social epistemology has turned its attention away from the question: how do we know? and now focuses instead on the question of the goals and results of science: what do we know and what do we not know, and why? The basic questions now are: who is science for? How does the fact that some people are excluded from science influence our knowledge? Which projects and goals are pursued and which ignored? Which experiences are considered valid or adequate and which are not? It is precisely these types of decisions that, to a large extent, make up the practice of science and which, finally, define the problems deemed relevant and the type of solutions deemed acceptable. Once a problem has been defined, just one response may be the most appropriate solution, but the problem could have been defined in a different way, and perhaps even other problems may have been identified as more relevant. From this point of view, Kitcher's arguments and thoughts regarding the possibility of a *well-ordered* science and the conditions under which such a science would be possible in the heart of democratic societies are extremely suggestive.

In my opinion, the question of *democratic science* is posed transversally in the reflection on science and enables us to identify three areas of analysis:

---

<sup>1</sup>Inmaculada Perdomo, PhD in Philosophy of Science and Professor of the Area of Logic and Philosophy of Science, Faculty of Humanities, ULL. Director of the Research Institute of Gender and Women Studies at the University of La Laguna. She is part of the team that develops the research project: "Justice, Citizenship and Vulnerability: Narratives of precariousness and intersectional approaches", FFI2015-63895-C2-1-R IP: María José Guerra y Eva Darias (2016- 2018)

- a. That related to *theory contents*.
- b. That related to the *practice of science*.
- c. That related to the *goals and directions of science*.

With regard to the first point, many interesting works by feminist science critics, which focus on warning about the biases present in scientific contents, mainly in biomedical theories, share this concern. With regard to the second, as an example, Feminist epistemology had explored the processes and activities of social construction of knowledge concluding that the subject is plural but women are not in the center of that process, their voices are devaluated and not authorized (Helen Longino, 2001). It is an epistemic injustice (Miranda Fricker, 2007), and political action must be considered to increase the number of women in STEAM areas, and the structures of science and academies must be changes in order to create a more democratic or inclusive decision-making processes (Londa Schiebinger, 2011).

The dynamics and gender asymmetries in career paths of women and men, and real evacuation of women from classrooms in relevant areas and professions must be also analyzed. As an example, this year we knew that Tokyo medical school altered test scores to keep women out. The University began to keep the percentage of women admitted each year to around 30 % of the incoming class. This had been done since 2010 when around 40 % of successful applicants were women. This is an explicit exclusion, but some other mechanisms must be observed.

Miranda Fricker's concepts (hermeneutical and testimonial injustices) explain these mechanisms of exclusion well, reaching the area of epistemology and ethics. Epistemic activity evaluated, not in ideal and abstract terms, but considering that it is carried out by socially situated individuals, in the highly relevant context of shaping our current techno-culture such as in education, profession and institutions. Inequalities and power issues associated with market dynamics form a highly competitive world displacing subjects defined as different i.e. women and others.

There are two basic epistemic practices that we need to consider to address Miranda Fricker's proposal. First, transmitting knowledge to others and secondly, to make sense of our own social experiences. Both types of epistemic practices involve social and discursive interactions so these interactions must be situated in the field of policies and ethics of epistemic practices. These mechanisms or types of *epistemic injustice* are at the root of the dynamics of exclusion of women from the center of shaping our current techno-culture, which produces, transmits and reiterates narratives of "authorized" subjects and make those of different subjects invisible. The myths and images that structure our visions of technology continue to transmit the image of men as authorized agents of scientific and technological development, and women as subjects not interested in it, or not as much capable as men.

Furthermore, in a clear reference to Helen Longino criteria for criticism in scientific communities, when members of certain groups are taken less seriously in the classroom, in the lab, at conference venues, and in the grant proposal process, one effect can be that the researcher does not receive the level of rigorous

criticism that might be required in order to strengthen their work or help them identify in what positive direction the work needs to be developed. This is too a participatory injustice, argues Grasswick (2017).

In the case of science, the significance of such participatory epistemic injustices is exacerbated because of the high degree of cognitive authority placed in the institutions of science. When society, as a whole, relies on and privileges the institutions of science to direct and produce knowledge that will have social relevance, the impact of participatory epistemic injustices that prevent or deter access to these communities of knowledge generation is more pronounced than in other areas of knowledge production (Grasswick 2017, 317).

And finally in relation to third point, Kitcher's analysis of democratic decision-making processes applied to science suggests several interesting dilemmas which may be posed in response to the questions: how should we decide democratically which research areas and projects should be funded? What goals should science obtain? What projects would contribute to the common good if put into practice? Firstly, what do we actually mean by "the common good": that which is beneficial to the citizens of modern, rich, western democracies or the ones that will benefit all humanity? This debate is plagued with dilemmas but the question of how we should make decisions about science in today's modern, democratic societies is closely linked to yet another question: how can we democratise decision-making processes in science?

The relevant question here is that the agents involved in this type of discussion are not only scientists; moreover, the decisions made only by scientists do not result in a *well-ordered science*, since the pressures of competitiveness may prompt them not to consider the interests of other individuals, the majority of human beings. In fact, it is precisely this overlooking of the interests of the vast majority of human beings that constitutes the starting point for the need of a *well-ordered science*. In that sense, the question: *what kinds of science should be done?* is crucial (Kitcher, 2011).

Scientific policies play a key role in the establishment of the goals and objectives, which science should strive to achieve. This is now a big problem for all the countries because of the risks associated to the climate change and the advances of Biotechnologies and promises of Transhumanism, and worries about the Posthumanist new era (R. Braidotti, 2013). Funding projects, establishing priority areas of research, making a commitment to applied or basic research and laying out selective research quality evaluation and assessment criteria, etc. are all ways in which science aims to achieve certain goals. This orientation necessarily involves the ignoring of other possible lines of action which may have otherwise been initiated or developed. In democratic contexts, scientific policies should pursue the *common good*. This notion becomes a central one, and as such, it also becomes a critical task of the philosophy of science, in all its plural trends, to warn of any "failures to comply with" or "deviations from" the main objective, namely, that of improving the living conditions of all human beings taking care of the all living beings and of our common house.

A final point must be included: Philosophy of science must deliver also an important social function: making the public more aware of the importance of science, promoting a critical and informed public opinion. Philosophers of science

ce could be considered as public intellectuals who speak up for science, and rectify common misconceptions or uninformed judgments that may feed into political lobbies, agendas and ultimately policy-making. Philosophy of science is an integral part of our public discourse on science. Recent public discourses of P. Kitcher or M. Massimi emphasize our responsibility to communicate the value of science to society at large. These all are crucial tasks for a socially responsible Philosophy of Science.

## References

- Braidotti, R. (2013). *The Posthuman*. Cambridge University Press.
- Cartwright, N. (1999). *The Dappled World. A Study of the Boundaries of Science*. Cambridge University Press.
- Graswick, H. (2017) in Kidd, Medina, Pohlhaus (eds.) *The Routledge Handbook of Epistemic Injustice*, Routledge, pp. 313–323.
- Longino, H. (2001). *The Fate of Knowledge*. Princeton University Press.
- Kitcher, P. (2001). *Science, Truth and Democracy*. Oxford University Press.
- Kitcher, P. (2011). *Science in a Democratic Society*. Prometheus Books.
- Kitcher, P., Fox Keller, E. (2017). *The Seasons Alter. How to save our planet in six acts*. Liveright Publishing Corporation.
- Schiebinger, L. (2011). “Interdisciplinary Approaches to Achieving Gendered Innovations in Science, Medicine, and Engineering”. *Interdisciplinary Science Reviews*, vol.36, n°2, 154–67.
- Van Fraassen, B. (2008). *Scientific Representation. Paradoxes of Perspective*. Oxford University Press.

# Biotecnología y maternidad por sustitución: ¿cosificación, mercantilización o altruismo de género?

Eulalia Pérez Sedeño  
IFS-CSIC

Se suele pensar que la biotecnología es un conjunto de métodos que se pueden desplegar o utilizar, con fines y medios diversos; pero también se puede considerar que es una institución desarrollada gracias a determinadas visiones ideológicas y políticas. En efecto, las investigaciones que se plantean y las innovaciones que se desarrollan en biotecnología están en gran medida vinculadas a los intereses y prioridades de las agencias de financiación, las empresas, los gobiernos e incluso determinados grupos de presión. Y, a menudo, los intereses de las mujeres, los ideales democráticos y feministas se quedan a un lado a la hora de obtener financiación para poder ser desarrollados. Y de ese modo, se aprecia que los objetivos o fines e imaginarios de las biotecnologías afectan de manera desigual a las mujeres (y otros grupos vulnerables) y pueden ayudar a perpetuar las desigualdades existentes.

En este trabajo abordamos la biotecnología de la maternidad por sustitución, una de las tecnologías de reproducción asistida que más controversia ha producido en nuestro país en el último año. Esta técnica, prohibida por ley en España, es ampliamente utilizada *de facto* gracias a la globalización médica. A través de informes y reportajes que aparecen en los medios de comunicación daremos cuenta de las complejas narrativas que hay alrededor de esta práctica y los debates suscitados recientemente. Por lo general, estos debates están muy polarizados y se centran en si la globalización del trabajo reproductivo es una explotación de la madre de sustitución, una oportunidad para ella, una expresión del altruismo de género e incluso de empoderamiento de género.

## Introducción

La *Ley 35/1988, de 22 de noviembre, sobre técnicas de reproducción asistida* en su artículo 1.2 afirmaba que “Las técnicas de Reproducción Asistida tienen como finalidad fundamental la actuación médica ante la esterilidad humana, para facilitar la procreación cuando otras terapéuticas se hayan descartado por inadecuadas o ineficaces”. Primero, no tratan la esterilidad sino la infertilidad y en el caso de la gestación por sustitución<sup>1</sup> ni eso.

---

<sup>1</sup> Siguiendo a Natalia Fernández (2017), utilizamos esa expresión en lugar de las más populares “maternidad subrogada” o “vientre de alquiler” porque es la que utiliza la Ley 14/2006.

Por lo que se refiere a la maternidad, en el artículo Segundo del Preámbulo se afirma: “Desde una perspectiva biológica, la maternidad puede ser plena o no plena, y ello es importante en relación con las técnicas que aquí referimos; en la maternidad biológica plena, la madre ha gestado al hijo con su propio óvulo; en la no plena o parcial, la mujer sólo aporta la gestación (maternidad de gestación), o su óvulo/s (maternidad genética), pero no ambos...Por su parte, la paternidad sólo es genética, por razones obvias de imposibilidad de embarazo en el varón. Finalmente, pueden la maternidad y la paternidad biológicas serlo también legales, educacionales o de deseo [deseos vs. Derechos], y en tal sentido, es importante valorar cuál es la más humanizada, la más profunda en relación con el hijo, pues habida cuenta de las posibilidades y combinaciones que puedan darse, especialmente cuando en la gestación intervienen donantes de gametos u óvulos fecundados, los Códigos han de actualizarse sobre cuestiones determinadas que no contemplan. En cualquier caso, y sin cuestionar el alcance de las otras variantes, se atribuye a la maternidad de gestación el mayor rango, por la estrecha relación psicofísica con el futuro descendiente durante los nueve meses de embarazo”. Aquí, aunque no se habla de gestación por surrogación si se habla de maternidad de gestación, es decir, aquella en que la mujer gesta, pero no aporta sus óvulos, considerándose que la más importante es la de gestación.

Esta Ley, muy adelantada en su época, tuvo que ser varias veces actualizada, algo que, como acabamos de leer, ya preveía ella misma. Es en la *Ley 14/2006, de 26 de mayo, sobre técnicas de reproducción humana asistida* donde aparece por primera vez la maternidad por surrogación bajo la denominación “gestación por sustitución”: *Artículo 10 Gestación por sustitución*: “1. Será nulo de pleno derecho el contrato por el que se convenga la gestación, con o sin precio, a cargo de una mujer que renuncia a la filiación materna a favor del contratante o de un tercero. 2. La filiación de los hijos nacidos por gestación de sustitución será determinada por el parto. 3. Queda a salvo la posible acción de reclamación de la paternidad respecto del padre biológico, conforme a las reglas generales”. Es decir, en nuestra legislación no se contempla para nada este tipo de maternidad, o mejor dicho, la gestante, es decir, la sustituta, será quien ostente la maternidad legal<sup>2</sup>. Es interesante señalar esta cuestión, pues desde ciertas instancias (ciertas asociaciones y partidos políticos) claman por la regulación de esta tecnología, pero esa regulación ya existe: está prohibida.

Muchas de las controversias en biotecnología, sobre todo las que tienen que ver con los cuerpos de las mujeres adquieren un tono moral o político. Todo ello nos hace pensar que es mejor no entender la biotecnología como un conjunto de métodos que se pueden desplegar o utilizar, con fines y medios diversos, sino como una institución que se ha desarrollado gracias a determinadas visiones ideológicas y políticas (Bliss, 2012). En efecto, las investigaciones que se plantean y las innovaciones que se desarrollan en biotecnología están en gran medida vinculadas a los intereses y prioridades de las agencias de financiación, las empresas o los gobiernos (Pérez Sedeño, 2017). Y queda aún más patente en los debates sobre la gestación por sustitución.

Pero, además, las TRA, y la maternidad por sustitución como una de ellas, permiten ampliar nuestra noción de parentesco y familia y qué personas pueden tener hijos. Dicho de otro modo, estas tecnologías están cambiando nuestra

---

<sup>2</sup>Con respecto a la paternidad, podría ser reclamada por el padre biológico.

concepción de la maternidad, de la paternidad, las enfermedades, la naturaleza del ser humano, etc. Y no son sólo un procedimiento para ‘atajar’ temporalmente la infertilidad, pues son procedimientos para que personas fértiles, pero del mismo sexo o sin pareja, puedan tener descendencia, “por lo que son procedimientos *creadores* u *originadores* de maternidades o paternidades” (Pérez Sedeño y Sánchez Torres, 2014: 238).

## Maternidad por sustitución

La gestación por sustitución, maternidad por sustitución o vientres de alquiler consiste en implantar un embrión previamente desarrollado por fertilización *in vitro* (FIV) en una mujer a la que se ha contratado para que lleve a término el embarazo, renunciando a sus derechos de maternidad y entregándose el bebé a quienes han contratado a la mujer. Los actores implicados son muchos, no solamente la madre gestante o el vientre de alquiler, sino también los padres que pueden ser de distinto sexo o del mismo sexo, pueden aportar todo el material genético (esperma y óvulos) o sólo parte y que la otra provenga de alguna otra persona no necesariamente de la pareja, e incluso de la propia madre de sustitución. Los casos que se han dado son múltiples en ese sentido. Además intervienen el personal médico, las propias tecnologías y las clínicas donde se llevan a cabo la implantación de los embriones y también el posterior parto.

La creencia subyacente es que la madre de sustitución no aporta ningún “material genético” al feto y, por tanto, carece de cualquier relación con él<sup>3</sup>. De hecho, en las clínicas en que se practica esta técnica, la maternidad no se presenta como un modelo afectivo, sino simplemente como un proceso biomédico, en el que no interviene la sexualidad ni media intercambio sexual alguno. Sin embargo, como bien saben pediatras, genetistas, etc. el medio en que se desarrolla el feto, el útero de la madre de sustitución a través del cual se alimenta y crece, tiene un influjo determinante. Lo paradójico es que es un hecho científico que las informaciones genéticas que hay en el cigoto no son suficientes para que alcance la capacidad de existir con vida autónoma, pues el desarrollo embrionario requiere, obligatoriamente, informaciones operativas exógenas, que, por ahora, solo la madre puede proporcionar en la especie humana. Y, de hecho, las madres de sustitución son sometidas a estrictos protocolos que van de la alimentación hasta el ejercicio y los periodos de descanso.

La gestación por sustitución es una técnica prohibida en España (al igual que en Francia o China), lo que hace que se recurra a una empresa privada fuera de nuestro país por supuesto determinadas personas de determinada clase social, elevada, debido a los altos costes. Hay varios países en los que la gestación subrogada es legal y que admiten que ciudadanos de otros países la realicen. Lo que varía en cada uno es la cobertura legal que ofrecen y el precio. Éste, por lo general, está en función las garantías sobre registro y filiación de los hijos una vez nacidos: en Estados Unidos, uno de los países más garantistas al respecto, cuesta entre 100.000 y 150.000 €, aunque podría ser mayor si surge alguna complicación médica; en Kazajistán, el precio es de unos 80.000 €, en

---

<sup>3</sup>Excepto en el caso de que la madre de gestación también aportara el óvulo, algo que no es habitual.



Rusia entre 60.000-70.000 € en Ucrania, unos 50.000 €—aunque en estos dos países no se permite para parejas homosexuales— y en Tailandia, entre 35.000 y 40.000 €). En casi ninguno de estos últimos países se ofrecen garantías de que las mujeres no estén siendo explotadas o de que no sean víctimas de mafias. De hecho algunas fuentes (como #NosomosVasijas) denuncian la existencia de “granjas de mujeres en la India” y en Sudáfrica (Pérez Sedeño, 2017).

En cualquier caso y dentro de una misma “agencia” o “clínica”, el precio varía en función de los servicios que se prestan. Así por ejemplo, en *BioTexCom Center for Human Reproduction* (Ucrania, con web en español: [https://biotexcom.es/?gclid=CiwKCAiA\\_c7UBRAjEiwAt%3ECZi8aK0zN\\_ioGVliK](https://biotexcom.es/?gclid=CiwKCAiA_c7UBRAjEiwAt%3ECZi8aK0zN_ioGVliK)) se ofrece un “paquete económico” (29.900 €) que incluye tales servicios como: Número ilimitado de intentos, todos los exámenes médicos necesarios, medicaciones y consumibles médicos, servicios jurídicos, prueba de ADN, alojamiento de 20 m<sup>2</sup>, alimentación, intérprete y transporte aeropuerto-clínica-hotel. En caso de parto prematuro los padres de intención pagan los gastos extra y si se produce un aborto involuntario después de la semana 12, se abonarán 6.000 euros extra. En el caso de que nazcan gemelos los padres de intención tendrán que pagar 3000 euros adicionales.

El paquete VIP (49.900 €) incluye, además de lo anterior: Diagnóstico Genético Preimplantacional (DGP) con modificaciones genéticas, si fuera necesario, y elección del sexo; servicios del pediatra 24h, canastilla del recién nacido, servicios de niñera de 9h a 18h, alojamiento de 100-150 m<sup>2</sup>, transporte extra y Smartphone moderno con número ucraniano. En el caso de parto prematuro, aborto involuntario o del nacimiento de gemelos, los gastos extra corren por cuenta de la clínica.

Los países que han legislado la gestación por sustitución siguen distintos modelos. Uno de los más recientes en aprobar una legislación sobre esta práctica ha sido Portugal, promovido por el bloque de izquierdas, pero con la oposición de comunistas y conservadores. Está restringida a mujeres que no tengan útero o a las que por algún problema médico no les funcione. En el Reino Unido también es legal y, al igual que en Portugal, debe ser altruista, pero en este caso el contrato no es vinculante porque la madre legal es la gestante en un principio. En Estados Unidos depende de cada Estado, pero algunos permiten esta técnica. Por lo general, es el destino de las parejas con mayor poder adquisitivo, puesto que tiene la legislación más garantista para las parejas que buscan un hijo. En Canadá, se realiza de forma altruista y en la sanidad pública, pero se contempla que los gastos del embarazo, pruebas, desplazamientos o días de trabajo perdidos sean cubiertos por las personas que buscan ser padres. En Brasil está permitido, pero sólo entre familiares. Hay otros países en los que existen legislaciones favorables (Holanda, pero sólo sus propios ciudadanos pueden acudir a la técnica (por ejemplo, México) y que además sean parejas heterosexuales (Israel). Y hay una serie de países donde no hay legislación al respecto (Guatemala, Japón, Panamá o Perú<sup>4</sup>), lo que hace sumamente vulnerables a las mujeres en esos países.

---

<sup>4</sup>Irlanda es un caso especial, pues ni siquiera tiene reguladas las TRA.

## Discusión y reflexiones finales

Como ya se ha mencionado, las posturas con respecto a la gestación por sustitución adquieren un tono político y ético, pero, al contrario de lo que sucede en casos como el aborto, los lineamientos no son claros. Así por ejemplo, hay feministas y partidos liberales de derecha (y algunos de izquierdas como en el caso de Portugal) que defienden esta tecnología aludiendo a la autonomía de las mujeres, altruismo o al empoderamiento de género. Y otras feministas y algunos partidos de izquierda, junto con la Iglesia Católica, que consideran que es una explotación más de los cuerpos de las mujeres. Se habla del derecho de las personas a satisfacer sus deseos. Podría pensarse que igual que una mujer “tiene derecho” a aumentar el tamaño de sus senos, si así lo desea, también una mujer (o un hombre) que no puede o no quiere gestar un hijo “tienen derecho” a satisfacer su deseo y no hacerlo, limitaría su autonomía.

En medicina, a veces es necesario limitar la autonomía del paciente, por ejemplo en el caso de enfermedades contagiosas cuando se puede obligar a la población a tratarse. Esa consideración ha llevado a que algunos países, como Francia, obliguen ciertas vacunaciones (no así en España, donde no es obligatorio). Ahora bien, esa limitación de la autonomía debe estar justificada, ser proporcional y racional.

Como señala María Luisa Balaguer (2017), las limitaciones que el Estado puede imponer a sus ciudadanos dependerá del concepto de Estado que se maneje: “mínimo, liberal, abstencionista respecto de las actuaciones y necesidades de las personas que lo conforman, o de un Estado social, redistributivo y atento a los grupos de individuos, en tanto que representativos de diversos intereses sociales” (Balaguer, 2017: 170). En el primer caso, las limitaciones a la autonomía individual serán mínimas; en el segundo, habrá que tener en cuenta que la actuación de un individuo incide en el grupo al que pertenece.

Además, no debemos confundir ni equiparar la autonomía del o la paciente con la soberanía del consumidor. Si bien el consumidor o consumidora accede a los bienes de consumo según sus deseos y capital, los y las pacientes no tienen derecho a recibir cualquier tratamiento que pidan y puedan pagar: una persona puede querer un tratamiento y poder pagarlo, pero si el médico no lo considera adecuado, no se le da. En qué medida la gestación por sustitución está orientada a y depende de la demanda del paciente/consumidor puede ser indicativo del carácter ético o no.

En otra parte (Pérez Sedeño, 2017) he señalado los problemas que presentan estas tecnologías para las mujeres: explotación de mujeres vulnerables, comercialización y cosificación de sus cuerpos, coste físico y psíquico-emocional (además del económico). Pero hay, además, un problema adicional: la comercialización y cosificación del bebé, que se convierte en un producto de consumo, que se puede encargar cuando se quiera... y devolver cuando la necesidad ya no exista, o no la satisfaga.

## Referencias

- Balaguer, María Luisa (2017). *Hij@s del Mercado. La maternidad subrogada en un estado social*. Madrid, Ed. Cátedra.
- Bliss, Catherine (2012). *Race Decoded: The Genomic Fight for Social Justice*. Stanford, CA: Stanford University Press.
- Fernández Jimeno, Natalia (2017). “Nuestros cuerpos, nuestra decisión: Tecnologías reproductivas y participación pública”. Ponencia presentada en el *XII International Workshop on Science, Technology and Gender: Knowledge, Practices and Activisms from the Feminist Epistemologies*. Madrid, June 21-23.
- Ley 35/1988, de 22 de noviembre, sobre técnicas de reproducción asistida*, BOE de 24 de Noviembre de 1988.
- Ley 14/2006, de 26 de mayo, sobre técnicas de reproducción humana asistida*, BOE de 27 de mayo de 2006.
- Pérez Sedeño, Eulalia (2017). “Promesas biotecnológicas. Determinismo genético, cáncer y maternidad por sustitución: un análisis crítico”. *Daimon. Revista Internacional de Filosofía*, Suplemento 6 (2017), 27–40.
- Pérez Sedeño, Eulalia y Sánchez Torres, Ana (2014). “Asimetrías y olvidos en las tecnologías de reproducción asistida”, en Pérez Sedeño, Eulalia y Ortega Arjonilla, Esther (eds.), *Cartografías del cuerpo: Biopolítica de la ciencia y la tecnología*. Madrid: Ed. Cátedra, 195–244.

# Proyecto Isotipo: Hacia un lenguaje visual unificado

Thais Rivero

Otto Neurath (1882-1945) fue un filósofo y economista austriaco perteneciente al Círculo de Viena con quien compartió el ideal de un lenguaje unificado y el activismo político. En el lenguaje visual unificado<sup>1</sup>, en el que se centra esta comunicación, se conjugan ambos ideales. Preocupado por la educación de la sociedad que le rodeaba, Neurath crea el LVU con el fin de promover la emancipación social y la democratización del conocimiento. Esto se materializó en el proyecto Isotipo. El lenguaje del Isotipo<sup>2</sup> es una forma visual de transmisión de información relevante y comprensible para cualquier ciudadano. A partir de este lenguaje el conocimiento podría ser de fácil acceso e intuitivo por ciudadanos que no compartieran el mismo lenguaje natural. Neurath creía, además, que tras la asunción de este lenguaje visual por la población, esta sería capaz de cuestionarse las bases educativas del sistema y preguntarse por las imposiciones dadas. El propósito perseguido por Neurath, tanto en la forma teórica como en la forma más práctica de su filosofía, era la humanización de la vida y la democratización del conocimiento<sup>3</sup>, ideales rescatados del proyecto ilustrado, lo que daría como resultado una sociedad más comprometida y justa.

Neurath instituyó la Fundación Internacional para la Educación Visual en La Haya en 1934, aunque la idea del Isotipo era anterior. A partir de 1940, en Oxford, la formación pasó a denominarse Instituto del Isotipo. Neurath prosiguió con el proyecto hasta su muerte (1945). Tras su fallecimiento, su mujer Marie Neurath continuó difundiendo su trabajo y manteniendo vivo el espíritu de transformación social arraigado al proyecto desde sus inicios.

Para el desarrollo del Isotipo, Neurath colaboró con Gerd Arntz (1900-1988), diseñador gráfico alemán. El Isotipo traducía los datos estadísticos de orden económico y social en pictogramas numéricos-fácticos que se hacían públicos mediante exhibiciones en museos y medios impresos<sup>4</sup>. A partir de estas muestras pretendían formar y educar a la población en cuestiones de economía y sociología. Las pautas del Isotipo eran muy simples: 1) usaban símbolos del mismo tamaño y forma (no variaban la escala sino que se multiplicaban los símbolos); 2) los objetos eran reproducidos de forma isométrica (es un método de representación gráfica en el que se reduce un objeto tridimensional a dos dimensiones); 3) la nitidez a la hora de la presentación. En la figura 1, podemos visualizar estas tres características. En este ejemplo, se compara la cantidad de automóviles, teléfonos y radios utilizados en cuatro países diferentes en 1937.

---

<sup>1</sup>A partir de ahora, LVU.

<sup>2</sup>*Isotype* sería el acrónimo de International System of Typographic Picture Education.

<sup>3</sup>Prono y Amino (2015).

<sup>4</sup>Guzmán (2015).

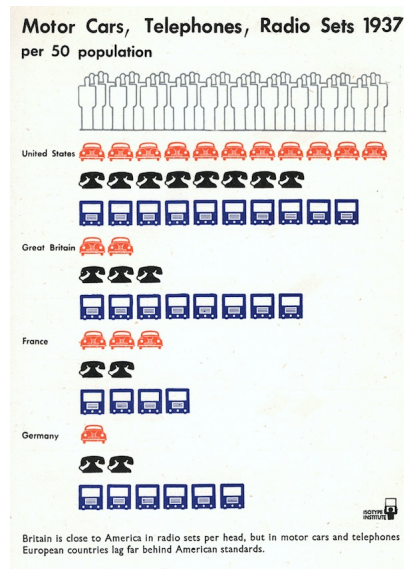


Figura 1. Representación basada en los principios de *Isotype*. Imagen extraída de la web EagerEyes (URL: <https://eagereyes.org/techniques/isotype>; Fecha de consulta: 15 mayo de 2018).

Como podemos observar en la Figura 1, la representación es clara y sencilla, no posee ornamentos, por lo que representa la información necesaria para el público en general, una de las ideas básicas que se perseguía a través de la educación visual.

La idea subyacente al LVU, es que Neurath consideraba que, debido a la complejidad del conocimiento, muchas imágenes no son auto-evidentes y necesitan completarse con un lenguaje de símbolos normalizados. Por eso desarrolló un vocabulario visual que sirviera para este propósito. El impulso de Neurath hacia el universalismo, que acompaña a cada una de las parcelas del Isotipo, se basa en nociones de transparencia, libertad de valor y una confianza básica en las instituciones históricas del conocimiento<sup>5</sup>. Añadía además a esta idea de universalismo una célebre frase: «Las palabras crean divisiones, las imágenes crean conexiones»<sup>6</sup>. Neurath enfrenta los lenguajes naturales a un lenguaje visual que es capaz de unificar sociedades, completamente divergentes entre sí, mediante imágenes.

Uno de los ejemplos más claros que tenemos de la aplicación del proyecto Isotipo son las señales de tráfico. Una muestra clara es la figura 2.

Vemos que, el LVU de Neurath posibilita y abre las puertas a un entendimiento entre diferentes sociedades a través de signos. Hace ya casi un siglo, que Neurath quería incitar a reflexionar a los ciudadanos austriacos mediante las exposiciones públicas de pictogramas. Además, se pueden encontrar sus huellas en algunos proyectos de filosofía visual. Por ejemplo, *Wonder Ponder* es un proyecto de filosofía visual dirigido a la infancia donde, mediante imágenes y la mínima expresión de texto, se induce a pensar acerca de la crueldad, la identidad personal, la realidad, entre otros temas. A través de esta iniciativa se puede ver como la imágenes son una herramienta muy potente para el desarrollo y la reflexión. En

<sup>5</sup>Hartle (2017).

<sup>6</sup>Neurath (1936) p. 18.

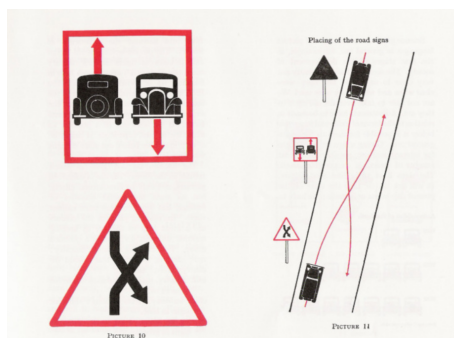


Figura 2. Señalización vial. Imagen extraída de la obra *International Picture Language* de Otto Neurath.

conclusión, encontramos en Neurath y el proyecto Isotipo el desencadenante de nuevos modelos educativos que intentan arraigar el pensamiento crítico a través de una filosofía visual.

## Bibliografía

- Guzmán, Valeria (2015). Aprender a ver. La *Autobiografía visual* de Otto Neurath, *Anales del Instituto de Investigaciones Estéticas*, vol. XXXVII, 107, 39–67.
- Hartle, Johan F. (2017). Otto Neurath's Visual Politics. An Introduction to "Pictorial Statistic following the Vienna Method", *ArtMargins and the Massachusetts Institute of Technology*, 98–107.
- Neurath, Otto (1936). *International Picture Language*, London, Kegan Paul, Trench, Trubner & Co.
- Prono, María I., Aimino, Matías (2015). *Otto Neurath: Lenguaje pictórico, educación visual y humanización del conocimiento (Seminario)*, Universidad de La Plata.

# El desequilibrio de género en filosofía

**Obdulia Torres González**  
Universidad de Salamanca

En los estudios de género, la preocupación exclusiva por las disciplinas STEM hace que se haya pasado por alto campos pertenecientes a las humanidades o las ciencias sociales, como la filosofía, la historia o la economía, donde la mujer se encuentra igualmente infrarrepresentada. En lo que respecta a España y con datos del curso 2016-2017 del total del PDI en filosofía un 25,8 % son mujeres, un 11,1 % catedráticas, un 25,2 % titulares, un 28,3 % ayudantes doctoras y un 44 % contratadas doctoras. Estos datos están muy alejados del correspondiente a su agregado estadístico, esto es humanidades, donde el porcentaje de mujeres sobre el total de PDI es del 44,4 %. Por categorías tenemos un 28,3 % de catedráticas, un 48 % de titulares, 53.2 % de contratado doctor y 53.8 % de ayudante doctor. Si la comparación se realiza con el porcentaje femenino del PDI de la Universidad tenemos la siguiente tabla.

| Categoría     | % mujeres         |                |
|---------------|-------------------|----------------|
|               | Total Universidad | Filosofía      |
| Total PDI     | 40.5              | 25.8           |
| Catedráticas  | 21.3              | 11.1           |
| Titulares     | 40.3              | 25.2           |
| Contratadas D | 48.6              | 44             |
| Ayte. Doctora | 50.4              | 28.3           |
| Ayudante      | 48.7              | 0 <sup>1</sup> |
| Asociada      | 40.3              | 25.7           |
| Emérita       | 25.8              | 17.6           |

Tabla 1: Porcentaje de mujeres en filosofía comparada con porcentaje de mujeres en el total de la Universidad.

Si observamos la evolución de los datos de filosofía vemos que la situación no es coyuntural y que responde a un patrón persistente dado que la tasa de crecimiento en dieciocho años ha sido mínima.

| Categoría            | Curso académico |       |       |       |       |       |
|----------------------|-----------------|-------|-------|-------|-------|-------|
|                      | 99-00           | 02-03 | 05-06 | 09-10 | 14-15 | 16-17 |
| Total PDI            | 19,3            | 20,7  | 21,7  | 22,4  | 26,3  | 25,8  |
| Catedráticas         | 7               | 9,2   | 11,3  | 11,5  | 12,2  | 11,1  |
| Titulares            | 19,5            | 21,2  | 19,1  | 21,1  | 24,3  | 25,2  |
| Contratadas doctoras |                 |       | 45,7  | 43,9  | 41,8  | 44    |
| Ayudantes doctoras   |                 | 60    | 56,3  | 44,4  | 26,3  | 28,3  |

<sup>1</sup>El porcentaje no es relevante dado que el número total de ayudantes asciende a 4.

|           |      |      |      |      |      |      |
|-----------|------|------|------|------|------|------|
| Ayudantes | 36,4 | 50   | 40   | 33,3 | 30   | 0    |
| Asociadas | 27,1 | 26,5 | 30,3 | 16,8 | 28,8 | 25,7 |
| Eméritas  | 0    | 0    | 0    | 0    | 9,5  | 17,6 |

Tabla 2: Evolución del porcentaje de mujeres en filosofía por categoría.

En cuanto al alumnado hemos pasado de un 42 % de mujeres matriculadas en el curso 1999-2000 a un 36,7 % en el 2016-2017, lo que denota que la infrarrepresentación, aunque menor también se encuentra en el alumnado.

La situación es común a otros países del ámbito anglosajón. En Australia, según la Asociación Australiana de Filosofía las mujeres detentan sólo un 28 % de las plazas fijas en el 2009. En Estados Unidos son un 21 % de las académicas en filosofía, en Canadá sólo un 21 % de las catedráticas son mujeres. (Jenkins y Hutchison, 2013: 5-6)

Aunque la situación no varía sustancialmente en las distintas áreas de filosofía, los peores datos los ostenta el área de lógica y filosofía de la ciencia con un 21,8 % de PDI total, un 9,3 % de catedráticas y un 21,2 % de titulares.

El 16 de julio de 2012 la discriminación de la mujer en la filosofía se hace visible, a nivel internacional, en un *call for paper* que apareció en la revista *History of Philosophy of Science*, (HOPOS) para la participación en un congreso sobre la figura de Feyerabend. De los quince conferenciantes invitados ninguno era mujer. A continuación, se produjo una acalorada discusión que incluyó llamadas al boicot del congreso, numerosos post en distintos blogs y acabó en un foro en HOPOS donde participaron las filósofas más importantes del ámbito anglosajón. (Andersen, 2013; Koertge, 2013; Lloyd, 2013; Longino, 2013; Sargent, 2013; Wuest, 2013) Esta llamada de atención ha supuesto un crecimiento exponencial de la bibliografía sobre el tema. Los estudios estadísticos muestran que la situación de España es igual a la de diversos países del ámbito anglosajón.

Aunque a nivel internacional la bibliografía acerca de la infrarrepresentación de la mujer es abundante existen pocos datos que la acrediten y los que existen están basados mayoritariamente en informes realizados por distintas asociaciones filosóficas, a través de solicitudes no oficiales de datos a las distintas universidades. En un trabajo clásico Haslanger (2008) señala que, en el año 2006, de un total de 412 contratados permanentes en filosofía, en los 20 departamentos más importantes de EEUU, un 19 % eran mujeres. En España para ese año hay un 22 % y para el 2016 asciende a un 26 %. El acceso a los datos de alumnado es considerablemente mejor. Según *American Academy of Arts and Sciences* en el año 2014 el porcentaje de mujeres que han obtenido el bachelors en filosofía es del 31 %, en master un 28 % y en doctorado un 31 %, mientras en España obtienen el grado en 2014 un 42 % y en el 2017 un 40 %<sup>2</sup>. El porcentaje de egresadas de máster en España para el año 2017 es de 50,2 %<sup>3</sup>.

<sup>2</sup>El porcentaje de estudiantes matriculadas parece ser mejor indicador que el de egresadas dado que el mejor aprovechamiento de las mujeres respecto a los hombres hace que exista una diferencia de 4 puntos porcentuales entre alumnas matriculadas y egresadas. Para el año 2015 y 2017 el porcentaje de alumnas matriculadas fue de 36,5 y 36,7 respectivamente.

<sup>3</sup>Nuevamente el porcentaje estudiantes matriculadas parece ser mejor indicador ya que para el mismo año baja a un 44 %.



No disponemos del número de tesis defendidas, pero si del porcentaje de estudiantes matriculadas en estudios doctorales que es un 36 %. Vemos que el porcentaje de mujeres en los estudios de filosofía es sensiblemente mayor especialmente en los estudios de máster.

Un dato relevante, dado que tiene que ver con alguna de las hipótesis propuestas para explicar la infrarrepresentación de la mujer en la filosofía es el porcentaje de mujeres por área de conocimiento. Según APA (2009) las 5 áreas donde hay más mujeres son, por este orden, filosofía feminista, ética aplicada, ética normativa, filosofía social y filosofía política. Mientras las cinco áreas ocupadas por varones son metafísica, epistemología, filosofía de la mente, ética normativa y metaética. Hay 38 puntos de diferencia a favor de las mujeres en filosofía feminista mientras que hay 15 y 12 a favor de los hombres en metafísica y epistemología respectivamente. Aunque no es directamente comparable, en España el porcentaje de mujeres por área es 27 % para filosofía moral, 23 % para filosofía y 24 % para lógica y filosofía de la ciencia.

En un estudio realizado por la Asociación Australiana de Filosofía entre 2003 y 2009 se contabilizaban un 28 % de mujeres entre el personal docente e investigador en filosofía, en las universidades australianas. En cuanto a la distribución por categorías sólo disponemos de datos del 2005 (Australasian Association of Philosophy, 2008) que son los que se muestran en la tabla 3. Los datos españoles para ese año son ligeramente inferiores en las categorías superiores de catedrática y titular y superiores en las categorías intermedias.

|                   | % Mujeres<br>PDI Filosofía |
|-------------------|----------------------------|
| A Assoc. Lecturer | 5,8                        |
| B Lecturer        | 28,5                       |
| C Senior Lecturer | 31                         |
| D Assoc Prof      | 21,8                       |
| E Professor       | 12,9                       |

Tabla 3: Porcentaje de mujeres entre el PDI en las Universidades australianas por categoría.

En Reino Unido, los datos del personal docente e investigador para el periodo 2008-2011 (British Philosophical Association, 2011) son similares a los españoles excepto en el porcentaje de catedráticas que sobrepasa en 7 puntos la proporción española. El porcentaje de estudiantes de grado asciende a un 44 %, las de máster un 33 % y las doctorado un 35 %. En lo que se refiere al personal docente e investigador los datos son un 26 % de lecturer, un 28 % de senior lecturer, un 22 % de reader y un 19 % de professor.

Aunque existen problemas para comparar los datos debido a la discrepancia de años, figuras y características de los distintos sistemas universitarios, las cifras permiten disponer de una panorámica general que avala la hipótesis de que la infrarrepresentación de la mujer en filosofía es un fenómeno generalizado.

Es importante señalar que cuando hablamos de infrarrepresentación no nos estamos refiriendo a un único fenómeno. Es decir, cuando explicamos la escasa presencia de la mujer en filosofía debemos explicar tres fenómenos distintos y que posiblemente se expliquen por distintas causas. Debemos explicar en primer

lugar, la baja proporción de alumnas en el grado de filosofía cuando lo comparamos con el porcentaje de alumnas que eligen el bachillerato de humanidades y que asciende a un 62 % para el curso 16-17. En segundo lugar, debemos explicar la baja proporción del PDI en comparación con las mujeres que obtienen el doctorado, que para el año 2011<sup>4</sup> estaba en un 38 %. Finalmente, debemos explicar el techo de cristal reforzado que sitúa a las catedráticas en filosofía en un 11 %, frente a un 21 % de promedio de catedráticas en la universidad y a las titulares con 15 puntos porcentuales menos respecto al promedio general.

No sólo no es un fenómeno único sino que es un fenómeno multicausal. Louise Antony ha sugerido un modelo que hace visible esa pluralidad de causas, lo denomina *la tormenta perfecta*. Con ello alude “a una situación especialmente mala causada por una combinación de circunstancias desfavorables” (Antony, 2012:231) Se trata de la interacción de los efectos de distintas formas de discriminación bien conocidas, que convergen en una determinada disciplina académica debido a sus características. Un primer factor puede ser las características definitorias de la disciplina. La filosofía es definida como racional, objetiva, mental, penetrante, seminal y rigurosa y su práctica debe basarse en el ataque y la demolición del oponente. Según las autoras, se produce un conflicto entre las normas esquemáticas (conjunto de creencias implícitas acerca de los géneros) de feminidad y las características que se consideran necesarias para el éxito en la disciplina como asertividad, persistencia, ingenio, tenacidad y un cierto desprecio por las normas que rigen el uso de la palabra. (Antony, 2012:238, Haslanger, 2008) Estas características chocan con aquellas consideradas femeninas como ser deferente, agradable o comprensiva. La cuestión es que la mujer es penalizada en cualquiera de los casos, si es asertiva y persistente se la penaliza porque no cumple las expectativas de género, si es atenta y deferente porque no cumple las expectativas de la academia.

Otro de los factores propuestos para explicar la infrarrepresentación de la mujer es el sesgo implícito, Está bien documentado que evaluadores (independientemente de su sexo o posición) califican inferiormente el mismo currículum cuando este lleva nombre de mujer que cuando lo lleva de hombre (Haslanger, 2008:213).

Fiona Jenkins (2013) argumenta que la meritocracia, en tanto sistema en el que se depositan las esperanzas para acabar con el sesgo, puede estar de hecho reforzando la discriminación. Esto es así porque los privilegios arraigados se identifican con los estándares de excelencia y en función de ellos se configuran los requisitos de entrada y promoción dentro de la disciplina. Todo ello no hace sino reproducir el dominio masculino.

Un tercer factor que ha sido propuesto en la literatura es la amenaza del estereotipo. Esta explicación se refiere a como la conciencia de pertenecer a un grupo, al que se considera inferior en algún sentido, afecta al rendimiento de los miembros de ese grupo. Las investigaciones se han centrado en como la amenaza del estereotipo, que afirma que las niñas son peores en matemáticas, influye en el rendimiento de las niñas en esta disciplina. ¿Es aplicable esta explicación a filosofía? Según Jennifer Saul (2013), el hecho de que el alumnado de filosofía sea predominantemente masculino, que la mayoría de los profesores sean varones

---

<sup>4</sup>A partir del año 2011 no existen datos desagregados dentro de las humanidades para las tesis de filosofía, aunque como ya se señaló hay un 36 % de estudiantes de doctorado en filosofía.

y que los autores que se tratan en los planes de estudio sean casi exclusivamente hombres, contribuye al estereotipo de que la filosofía es una disciplina masculina y que las mujeres tienen un rendimiento inferior (Saul, 2013:44). La conciencia de ese estereotipo hace que las mujeres abandonen filosofía tras los cursos introductorios.

Recientemente se ha propuesto una hipótesis denominada *specific-field ability belief* (Ver Leslie et al., 2015) Dicha hipótesis explica los datos de infrarrepresentación de la mujer, en determinadas áreas, como resultado por un lado de las creencias, que mantienen sus profesionales, en la necesidad de un talento específico, excepcional e innato para la práctica de la disciplina y por otro, de la creencia en que la mujer carece de esas habilidades. La investigación mostró que los profesionales de la filosofía eran los que mantenían esas creencias en su grado más extremo (Ver Torres, 2018b).

Este conjunto de factores, que no son exclusivos de filosofía, confluyen en el campo y configuran la *tormenta perfecta*.

Pero esta, la hipótesis de la discriminación es solo una de las tres hipótesis<sup>5</sup> que han sido propuestas para explicar la baja presencia de la mujer en la filosofía. La segunda de las hipótesis es la diferencia en habilidades cognitivas.

Sesardic y de Clerck argumentan que son las diferencias estadísticas entre los sexos en habilidades mentales, claves para el desarrollo de determinadas disciplinas, lo que explica la infrarrepresentación de la mujer. Concretamente citan “tareas que requieren transformaciones en la memoria de trabajo visual-espacial... y razonamiento fluido, especialmente en dominios matemáticos y científicos abstractos” y añaden que esto es relevante para la situación en filosofía y lo fundamentan en la menor participación de la mujer en áreas como lógica, teoría de la decisión y filosofía de las matemáticas (2014:465). Hay tres cuestiones que tienen que ser señaladas aquí: en primer lugar, los estudios señalan que no hay diferencias en éxito académico en las habilidades que se postulan diferentes a nivel de bachillerato. El informe de la Comisión Europea (basado en los últimos informes PISA), señala que en algunos países de la UE las chicas puntúan igual o superior a sus compañeros varones. (European Commision, 2012:73). En segundo lugar, hay que demostrar que la pretendida diferencia en habilidades cognitivas (recabadas a través de las pruebas SAT que son las que invocan Sesardic y de Clerk) es un buen indicador de las elecciones vocacionales de hombres y mujeres y, según los informes, esto no es así. Como se señala en Torres (2018a:240) en Estados Unidos no son los que obtienen puntuaciones más altas en el SAT-M los que después eligen ciencia e ingeniería. La estadística señala que, menos de un tercio de los hombres que se dedican a ciencia e ingeniería obtuvo una puntuación por encima de los 650 puntos, que se considera el umbral mínimo para tener éxito en estos campos. Finalmente, los tres puntos porcentuales de diferencia entre el área de lógica y el de filosofía moral (pretendidamente preferida por las mujeres) no apoyan una hipótesis de diferencias en habilidades cognitivas.

Se ha argumentado que hombres y mujeres difieren en sus intuiciones filosóficas, entendiendo intuición como una respuesta espontánea de la gente a los experimentos mentales filosóficos (Buckwalter y Stich, 2014). Las diferencias se

---

<sup>5</sup>Una tercera hipótesis de la que no vamos a hablar aquí por motivos de espacio es la de la diferencia en intereses.

encontraron en los casos de Gettier, el cuarto de Mary, cerebros en una cubeta, las tierras gemelas o la habitación china. Cómo pasamos de esas diferencias en intuiciones filosóficas a la infrarrepresentación de la mujer en filosofía. La segunda premisa del argumento es que algunas mujeres pueden pensar que no son buenas en filosofía dado que no tienen las intuiciones que sus compañeros (mayoritarios en el aula) o sus profesores (también mayoritarios) insisten en que son las correctas. Lo que la hipótesis plantea es que parte de la infrarrepresentación de la mujer en filosofía se explica como un efecto de autoselección. La hipótesis es problemática por varias razones. Adleberg et al (2014) señalan tres: en primer lugar, los resultados no han podido ser replicados, en segundo lugar no hay una intuición filosófica aceptada o estándar de los experimentos mentales y cuando la hay la evidencia sugiere que la mujer mantiene la intuición estándar frente a los hombres. A estos argumentos podemos añadir, que la hipótesis podría explicar la infrarrepresentación de la mujer entre el alumnado de filosofía, pero no las dificultades para desarrollar la carrera académica o el techo de cristal reforzado. Finalmente, habría que ahondar en las diferencias de intuición intragéneros y relacionarlas con otras variables demográficas antes de afirmar que hay diferencias en intuición debido al género.

Sean cuales sean las razones, la discriminación de género en filosofía no afecta sólo a las mujeres, afecta a la disciplina y la percepción social que se tiene de ella, a la escasez de alumnos y las dificultades para atraer mujeres a un campo que se percibe como poco hospitalario, a la calidad de una reflexión sobre el mundo que refleja sólo a la mitad de la población. En resumen, afecta a la supervivencia institucional de la propia disciplina.

## Referencias

- Adleberg, Tony, Thomson, Morgan y Nahmias, Eddy (2014) "Do men and women have different philosophical intuitions? Further data." *Philosophical Psychology*, 28(5):1–27.
- American Academy of Arts and Sciences (2018) *Humanities Indicators*. <https://www.humanitiesindicators.org/content/indicatordoc.aspx?i=267> consultado el 04/09/2018.
- Andersen, Hanne (2013) "Women in the History of Philosophy of Science; What We Do and Do Not Know." *History of Philosophy of Science*, 3(1):136–139.
- Antony, Lousie (2012) "Different Voices or Perfect Storm: Why Are There So Few Women in Philosophy?" *Journal of Social Philosophy*, 43(3):227–255.
- APA (2009) *Preliminary Report of the Survey on Publishing in Philosophy*. [https://www.apaonline.org/general/recommended\\_links.asp?cc=43686](https://www.apaonline.org/general/recommended_links.asp?cc=43686) consultado el 04/09/2018.
- Australasian Association of Philosophy (2008) *Improving the Participation of Women in the Philosophy Profession*. [https://aap.org.au/Resources/Documents/publications/IPWPP/IPWPP\\_ExecutiveSummary.pdf](https://aap.org.au/Resources/Documents/publications/IPWPP/IPWPP_ExecutiveSummary.pdf) Consultado el 17/09/2018.

- British Philosophical Association (2011) *Women in Philosophy in the UK A report by the British Philosophical Association and the Society for Women in Philosophy UK*. <https://bpa.ac.uk/resources/women-in-philosophy/journals> Consultado el 05/09/2018.
- Buckwalter, W y Stich, S. (2014) "Gender and philosophical intuitions" en Knobe, J y Nichols, S., *Experimental Philosophy*, Oxford, Oxford University Press, pp. 307–340.
- European Commisión (2012) *Meta-analysis of gender and science research*. Luxembourg: Publications Office of the European Union.
- Haslanger, Sally (2008) "Changing the Ideology and Culture of Philosophy: Not by Reason (Alone)" *Hypatia*, 23(2):210–223.
- Jenkins, Fiona (2013) "Singing the Post-discrimination Blues: Notes for a Critique of Academic Meritocracy," en Jenkins, Fiona y Hutchison, Katrina (2013)
- Koertge, Noretta (2013) "Feyerabend, Feminism and Philosophy." *History of Philosophy of Science*, 3(1):139–141.
- Leslie, Sarah, et al. (2015) "Expectation of brilliance underlie gender distributions across academic disciplines." *Science*, 347(6219):262–265. [doi: 10.1126/science.1261375]
- Lloyd, Elizabeth (2013) "Constitutional Failures of Meritocracy and their consequences." *History of Philosophy of Science*, 3(1):142–144.
- Longino, Helen (2013) "Data Please." *History of Philosophy of Science*, 3(1):144–146.
- Sargent, Rose-Mary (2013) "Forum: Pluralism? Reflections on Women in the History of Philosophy of Science." *History of Philosophy of Science*, 3(1):135–136.
- Saul, Jennifer (2013) "Implicit Bias, Stereotype Threat, and Women in Philosophy."
- Sesardic, Neven y de Clerck, Rafael. "Women in Philosophy: Problems with the Discrimination Hypothesis." *Academic Questions*, 27(4):461–473.
- Torres González, Obdulia (2018a) "La segregación horizontal el riesgo de los agregados estadísticos." *Feminismos*, 31:231–250.
- Torres González, Obdulia (2018b) "La situación de la mujer en los estudios de filosofía. Un análisis basado en indicadores." *Investigaciones Feministas*, en prensa.
- Wuest, Amy (2013) "Yes, There Is a Problem: What is to Be done About the Climate for Women in Philosophy?." *History of Philosophy of Science*, 3(1):146–150.

# Puntos de vista y referencias temporales en una serie de televisión. El caso de “The affair”

Margarita Vázquez Campos  
Laura García Díaz  
Universidad de La Laguna<sup>1</sup>

En este trabajo, relacionaremos la noción filosófica de puntos de vista con estructuras temporales ramificadas y lo ejemplificaremos en una serie de televisión actual, *The Affair*.

Siguiendo a Vázquez (2015), podemos decir que nuestros puntos de vista están llenos de contenido indéxical. Uno de estos índices es el emplazamiento temporal. Esto es especialmente importante cuando el sujeto del punto de vista es una “persona”. Ser una persona “como nosotros” lleva implícito tener una perspectiva temporal, al menos, mínimamente correcta. Aunque esta corrección podría ser solamente interna al propio punto de vista personal.

Tenemos puntos de vista temporales (PdVT) porque nuestros pensamientos y lenguajes tienen estructura (gramatical) temporal<sup>2</sup>. En la definición de Liz y Vázquez (2015) los PdVT son puntos de vista (PdV) en los que algunas diferencias en el contenido son identificadas como cambios en el tiempo. El problema sería cómo entender el tiempo para representar esos cambios en el tiempo.

Arthur Prior, el creador de la mayor parte de los sistemas de lógica temporal, pensaba que nuestro lenguaje y nuestros pensamientos tienen “una perspectiva interna del tiempo”, en la cual el tiempo es representado como estableciendo un pasado y un futuro para un cambiante ahora. Esto está profundamente relacionado con la idea de que ser una persona “como nosotros” lleva implícito tener una perspectiva temporal, al menos, mínimamente correcta. Esta perspectiva temporal podría ser sólo, como hemos dicho más arriba, interna al punto de vista. Sin al menos esta perspectiva interna no podemos entender lo que el tiempo es. En los sistemas desarrollados por Prior, los aspectos internos y externos del tiempo están íntimamente relacionados, puesto que en la semántica se utiliza una relación antes/después y en la sintaxis aparecen operadores temporales para el pasado y para el futuro.

Para representar el emplazamiento en el tiempo de los puntos de vista parece que necesitaríamos, entonces, aspectos internos y externos. Pero para represen-

---

<sup>1</sup>Este trabajo ha sido realizado en el marco del proyecto *Puntos de vista, disposiciones y tiempo. Perspectivas en un mundo de disposiciones* (FFI2014-57409-R), financiado por el Ministerio de Economía y Competitividad del Gobierno de España).

<sup>2</sup>En inglés hay una diferencia entre *temporal* y *tensed* que, a veces, resulta difícil de señalar en castellano. En este caso, nuestros pensamientos y lenguajes estarían *tensed*.

tar el cambio en el tiempo puede que necesitemos algo más. Especialmente si no queremos restringirnos a una perspectiva determinista. Si el cambio involucra indeterminismo, nos vemos abocados a introducir la noción de posibilidad (y, con ella, la de necesidad). Para explicar este cambio necesitamos dos tipos de elementos indécicos, uno para el emplazamiento temporal y otro para las posibilidades de cambio. Esto también es así en los sistemas de lógica temporal que dan cuenta del tiempo indeterminista. En estos sistemas, necesitamos operadores temporales para el emplazamiento temporal y operadores modales para representar las posibilidades.

Es habitual utilizar la metáfora de un árbol para representar el tiempo indeterminista. Paul Horwich (1987) y Sebastian Álvarez (2013) dicen que se puede representar la concepción del tiempo con un único pasado y un futuro lleno de posibilidades mediante la imagen de un árbol, cuyo tronco, único y sólido, sería símbolo del pasado, y cuyas ramas lo serían de las diversas posibilidades que constituyen su futuro; la sección del tronco en que este comienza a ramificarse representaría el presente. En Vázquez (2013, p. 249) se discrepa con esta imagen y se presenta la imagen de un cable, por ejemplo como los de las líneas telefónicas. Hacia la mitad del cable, se le quita la cubierta exterior y quedan multitud de cablecitos más finos, que podemos separar unos de otros. La mitad del cable que tiene la cobertura exterior representa el pasado y la otra mitad sin cobertura, abierta, representa el futuro en un momento determinado (un presente determinado), que sería aquel en el que abrimos el cable.

Así, esta manera de representar los futuros posibles es a través de historias independientes con pasados paralelos. Para explicar esta representación también podemos acudir al concepto de simulación de un modelo por ordenador y a la idea de escenario. Un escenario de una simulación es una serie de resultados que conseguimos partiendo de unas condiciones iniciales. Si cambiamos las condiciones iniciales, partiendo del mismo modelo, obtendremos otro escenario. Estos escenarios pueden ser muy diferentes, incluso opuestos. Imaginemos un modelo sobre poblaciones de zorros y conejos que interaccionan. En un escenario las dos poblaciones pueden mantenerse más o menos estables, en otro desaparecer los conejos y en un tercero desaparecer los zorros.

De esta forma la idea de puntos de vista temporales, tiempo indeterminista y escenarios de una simulación parecen compartir una estructura con muchas similitudes. En esta vemos historias paralelas que comparten hechos similares. A partir de un cierto momento (el presente, el momento actual, un punto marcado en una simulación...) divergen, permaneciendo en el espacio de lo posible.

Estas ideas encuentran su ejemplificación en la serie de televisión *The Affair*.

La serie presenta una ruptura con los géneros. No se ajusta a los parámetros de las series policiales ni de las dramáticas. Sabemos que hay un crimen, en el futuro, pero no sabemos quién es el asesinado ni quien el asesino. Ese punto del asesinato nos sirve como anclaje a partir del cual se reconstruye la historia. Una historia de amor, celos y dramas, pero también de intriga. Ese crimen funciona como presente/futuro a partir del cual se estudia el pasado. El pasado de cada personaje es el mismo y es diferente. Coinciden los tiempos, los espacios, pero no las percepciones ni los recuerdos. Por ejemplo, dos personajes se encuentran en una casa y discuten, pero cada uno recuerda esa conversación con un tono

diferente, la discusión puede haber empezado por una causa distinta y ni siquiera recuerdan igual la ropa con la que iban vestidos. Para expresar estas distintas historias, los guionistas de la serie recurren a la narración en distintas voces. Cada episodio se divide en dos partes, cada una contada por un protagonista. A veces es el mismo periodo temporal y lugar narrado por dos personajes. Otras veces hay variaciones de tiempo y de espacio. Con las narraciones de los distintos personajes, vamos llegando al momento del crimen en la primera temporada. También hay escenas posteriores a ese crimen. Es decir, nos adentramos en el futuro de ese hecho. Con esto, el crimen y sus momentos posteriores también pasan a ser pasado de ese otro futuro en el que se celebra un juicio y nos encontramos a nuestros protagonistas muy cambiados. Es más, sabemos quién será el acusado, pero todavía, hasta el final de la primera temporada, no sabemos de qué crimen. Lo que nos interesa de esta serie es su relación con el símil del cable del tiempo y, también, con la noción de escenario de una simulación, y su evidente idoneidad para explicar los puntos de vista de sujetos diferentes acerca de un mismo hecho. Esto se puede ejemplificar con distintas escenas y situaciones de todos los episodios de la primera temporada.

## Referencias bibliográficas

- Álvarez Toledo, S. (2013). Imágenes del tiempo. En Liz, M. (Ed.) (2013), pp. 211–248.
- Horwich, P (1987). *Asymmetries in Time*. Cambridge, Massachusetts: MIT Press.
- Liz, M. (Ed.) (2013). *Puntos de vista. Una investigación filosófica*. Barcelona: Laertes.
- Liz, M y M. Vázquez (2015). Temporal Points of View. En Vázquez, M y M. Liz (Eds.) (2015).
- McTaggart, J. (1927). *The Nature of Existence, vol.II* (compilado por C.D. Broad). Cambridge: Cambridge University Press.
- Prior, A. (1967). *Past, Present and Future*. Oxford: Clarendon Press.
- Vázquez, M. (2013). El cable del tiempo. En Liz, M. (Ed.) (2013), pp. 249–262.
- Vázquez, M. y M. Liz (Eds.) (2015). *Temporal Points of View. Subjective and objective aspects*. Heidelberg: Springer.



# Enhancement technologies and inequality

Walter Veit  
University of Bristol

## 1.Introduction

Recognizing the variety of dystopian science-fiction novels and movies, from *Brave New World* to *Gattaca* and more recently *Star Trek*, on the future of humanity in which eugenic policies are implemented, genetic engineering has been getting a bad reputation for valid but arguably, mostly historical reasons. In this paper, I critically examine the claim from Mehlman & Botkin (1998: ch. 6) that human enhancement will inevitably accentuate existing inequality in a free market and analyze whether prohibition is the optimal public policy for this objection as egalitarians might advise (Lamont and Favor, 2014). Of course, various moral theories will evaluate this claim, which I shall call the *inequality-objection*, differently. Yet, for the purposes of the present paper, I shall only consider whether the *inequality-objection* is a problem to the dominant moral framework in policy-making, i.e. utilitarianism. Whether enhancements it poses a problem for agents' well-being or not. In what follows I argue that the *inequality-objection* has several problems. My criticism is structured as follows: In **Section 2** I will attempt to bring the argument at stake into formal argument structure and clarify terms that are unclear. In **Section 4** I attack the claim that Human Enhancement leads to inequality. In **Section 5** I analyze why inequality decreases well-being and if it is the case that the benefits from human enhancements do not actually outweigh the costs, which is what utilitarians should be concerned with. In **Section 6** I attack the conclusion and analyze whether prohibition is actually the best possible policy in maximizing well-being. Lastly, in **Section 7** I will summarize my attacks on the argument and conclude that the *inequality-objection* does not sufficiently support the conclusion that human enhancement should be prohibited by policy-makers.

## 2.What is at stake?

Though perhaps having raised the largest concerns in the public perception of human enhancement, the *inequality-objection* is neither a very sophisticated argument nor a very new one, against the introduction of new technologies that could enhance humans beyond what might be considered 'normal'. In fact, the whole argument is old wine in new bottles. Concerning inequality, Bostrom and Roache claim that one could hold the justified worry that:

...people with radically enhanced cognitive capacities might gain vast advantages in terms of income, strategic planning, and the ability to influence others; in other words, an enhanced cognitive elite may gain socially significant amounts of power. (Bostrom & Roache; 2007: p. 15)

Of course, it seems undesirable if an enhanced elite accumulates a lot of power, but whether this is even a plausible scenario is an open question. Rather than considering these extreme cases I focus instead on inequality in general, so those scenarios where inequality has the potential to decrease well-being.

From a utilitarian perspective the inequality-objection can be fleshed out as follows:

P1: Human enhancement leads to inequality.

P2: Inequality leads to a decrease in well-being.

P3: What leads to a decrease in well-being should be prohibited by policy-makers.

C: Therefore human enhancement should be prohibited by policy-makers.

In order to attack this argument in a precise and clear manner, let me begin by clarifying three crucial terms, i.e. human enhancement, inequality and utilitarianism. First, I adopt the definition of Bostrom and Roache who define human enhancement as interventions that “...aim to improve the state of an organism beyond its normal healthy state.” (2007: p. 1). By inequality, I mean inequality in human capacities and focus on how this could decrease overall well-being. How does this connect to utilitarianism? Utilitarianism has this underlying consequentialist principle: *Good is the maximization of well-being or utility*. As the conclusion is a prohibition, we shall be concerned with whether a prohibition is what utilitarianism would actually advise. With this in mind let me now turn to premise (1) and examine whether human enhancement leads to inequality.

### 3.1.Human Enhancement leads to Inequality?

Each of the following sections will be focused on one of the premises. Let me begin with the premise that human enhancement will increase inequality. This premise is empirical in nature and as such requires supporting evidence. While it is unclear what is evidence for the adequacy of this premise there is at least some evidence that seems to speak against it. More specifically, there are reasons to believe that human enhancement will not affect inequality at all, or even decrease it as Bostrom and Roache (2007: p. 16) suggest by making people “more equal” like it is the case with Modafinil (Randall et. al.; 2005). Bostrom suggests that we know far more about genes responsible for inheritable diseases than genes responsible for talents, intelligence and longevity, hence human enhancement might especially help the genetically worst-off in society (Bostrom; 2003: p. 18). Furthermore, as Veit (2018, p. 12) argues, at least when it comes to genetic enhancements, we can expect them to get rid of a major source of inequality, i.e. the elimination of the natural lottery. What we have to be concerned with

is, therefore, the plausibility of the *inequality-objection*. Erik Parens even claims that:

Those who already have economic resources will readily gain access to new technologies and those new technologies will make them stronger competitors for more resources. (Erik Parens; 1998: S8)

If Parens claim is true than an implicit premise needs to be true, that is: Enhancing yourself e.g. cognitive capacities are good investments as they help you in the market. Otherwise, those who enhance themselves would not be better competitors. This means that some enhancements like an increase of cognitive capacities might enable agents to become better competitors. Considering this it seems unlikely that people wouldn't consider taking loans for enhancing themselves, in order to increase their expected wage. Banks might even give loans for human enhancement in a similar fashion to education loans. Instead of paying for advanced training courses, employers might consider paying for cognitive enhancements of their employees. Whether they will remain an open question, as the profitability and risks are currently unknown.

Even so, let us grant for the moment that the above worries are well-motivated. In what follows I consider the premise (2) and question whether inequality will really lead to a decrease in well-being.

### 3.2. Inequality leads to a Decrease in Well-being?

According to egalitarians, everyone should be equal. The conception of this equality might differ, e.g. economic equality, equality in well-being, equality of opportunity, etc. (Arneson, Richard; Summer 2013). Utilitarians, of course, are not concerned with these kinds of consequentialist theories, but rather the maximization of well-being. What matters then is how inequality has the potential to decrease the well-being in a population. Several reasons could come to mind: potential oppression, discrimination, unfairness, exploitation, envy or simply a preference for equality (Mehlman, Maxwell J. & Botkin, Jeffrey R.; 1998). Even if we grant that these are worrisome consequences of human enhancement, in the utilitarian framework we would still have to weigh the benefits against these costs. Proponents of prohibition often referred to as *Bioconservatives* (Bostrom, Nick & Roach, Rebecca; 2007) seem to either ignore the benefits of human enhancement or implicitly think that they do not outweigh the costs. To make their claim as strong as possible, I will give them the benefit of the doubt and tackle the latter. While we may grant that human enhancement might decrease well-being by introducing inequality, there are other areas where human enhancement can potentially increase well-being enormously. Bostrom and Roache (2007) discuss several of these areas e.g. health, intelligence, life extension. Whether the costs outweigh the benefits is an open question. However, as argued in Veit (2018), there could be potential solutions to keep the benefits of human enhancement while limiting the costs of inequality, most obviously state-funding. Therefore what I analyze next is the possible solution of prohibition and how it compares to other alternatives.

### 3.3. Prohibition of Human Enhancement, the Best Solution?

Even if we accept that human enhancement has the potential to increase inequality in the society and the costs outweigh the potential benefits, the question arises whether a prohibition is actually the best way to maximize well-being. What we shall be concerned with are the consequences of this prohibition. I argue that it is wrong to think that we have the choice between status quo and a world where human enhancement takes place. As Vellemann (2014) outlines, one additional option alone can change the whole situation. Applied to human enhancement this means that a world where it is illegal does not equal a world where this technology is not available. This is because of several problems suggested in the literature, which need to be considered: First, Bostrom and Roache highlight that a legal prohibition of human enhancement requires distinguishing current practices of medicine from enhancements, which creates several problems (2007: p. 1–3). Second, the prohibition would have to be globally enforced, otherwise one could still access these technologies legally (Jacob Heller & Christine Peterson; 2006), but would only make them more expensive (Patrick Lin & Fritz Allhoff; 2008). Third, Anders Sandberg and Nick Bostrom claim that a prohibition will be an incentive for the creation of black markets, which as with drugs may tend to make these technologies even more expensive, while legal enhancements would become less risky and less expensive as time goes on (2009: p. 333). Fourth, Maxwell J. Mehlman and Jeffrey R. Botkin (1998: chapter 7) raise the question of how those breaking the law should be punished, for instance, will someone with genetic enhancements be forced to alter his DNA back to the point it was before or have to be sterilized? Will people who would have died without enhancements have to be sentenced to death by revoking their enhancements? These issues are not easily avoidable for legislators when effective prohibition is being implemented.

So even if the claim is correct that human enhancement will lead to inequality, the proposed solution of prohibition seems to face severe problems that proponents of prohibition are invited to address. In short, there are several reasons speaking against prohibition and some that speak in favour of other solutions. Let me shortly expand on them.

For instance, Nick Bostrom (2003: p. 17) proposes subsidies or free access for children of poor families. I agree with Lamont, Julian and Favor, Christi (Fall 2014: section 6) that the middle ground between open market access and prohibition needs to be considered. As such it seems that prohibition is not the best policy against this risk, it might even turn out to be worse than other kinds of regulation that address at least some of the concerns that proponents of the *inequality-objection* have brought forward.

## 4. Conclusion

In conclusion, there seem to be at least three major problems with the *inequality-objection*. First, there is no empirical evidence that human enhancement will lead

to inequality. Second, even if we accept that human enhancement will lead to inequality, it is not clear whether this will outweigh the potential benefits of human enhancement. Third, even if it does, prohibition is not necessarily the best policy for maximizing well-being. This is because it is, for instance, plausible to assume that, given that human enhancement is possible and distributed via black markets, there are several plausible ways in which this will affect overall well-being negatively. In light of these concerns, I conclude that even if human enhancement leads to inequality, this is not sufficient to conclude that human enhancement should be prohibited by policy-makers.

## References

- Bostrom, N. & Roach, R. (2007). "Ethical Issues in Human Enhancement." *New Waves in Applied Ethics*, ed. Jesper Ryberg (Palgrave Macmillan).
- Bostrom, N. (2003). "Human Genetic Enhancements: A Transhumanist Perspective." *Journal of Value Enquiry*, 37(4):493–506.
- Bostrom, N. & Sandberg, A. (2009). "Cognitive Enhancement: Methods, Ethics, Regulatory Challenges." *Sci Eng Ethics*, 15:311–341.
- Parens, E. (1998). "Is Better Always Good? The Enhancement Project." *The Hastings Center Report*, 28(1):S1–S17.
- Randall, D. C., Shneerson, J. M. & File S. E. (2005). "Cognitive effects of modafinil in student volunteers may depend on IQ." *Pharmacology Biochemistry and Behavior*, 82(1):133–139.
- Mehlman, M. J. & Botkin, J. R. (1998). "Access to the Genome. The Challenge to Equality."
- Lamont, J. & Favor, C. (2014). "Distributive Justice." *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), Accessed on June 19, 2015: <http://plato.stanford.edu/archives/fall2014/entries/justice-distributive>
- Arneson, R. (2013). "Egalitarianism." *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), Edward N. Zalta (ed.).
- Heller, J. & Peterson, C. (2006). "Human Enhancement and Nanotechnology: A Foresight Nanotech Institute Policy Issues Brief." Foresight Nanotech Institute. Accessed on June 19, 2015: <http://www.foresight.org/policy/brief2.html>
- Lin, P. & Allhoff, F. (2008). "Against Unrestricted Human Enhancement." *Journal of Evolution and Technology*, 18(1):35–41.
- Veit, W. (2018). "Procreative Beneficence and Genetic Enhancement." *KRITERION – Journal of Philosophy*, 32(1):75–92. URL: <http://www.kriterion-journal-of-philosophy.org/kriterion/issues/Permanent/Kriterion-veit-01.pdf>

Velleman, J. D. (1992). "Against the Right to Die." *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, 17(6):665–681.

# HISTORIA DE LA CIENCIA

# Telos y especie biológica: la historia de dos conceptos inseparables

**Daniel Labrador Montero**

Universidad de Salamanca

danilabra@usal.es

## Resumen

¿Son las especies categorías reales o meras convenciones? ¿Son los enunciados teleológicos un rasgo distintivo de la biología como ciencia? Estas preguntas, comunes en filosofía de la biología, comparten un núcleo común. A primera vista, parece que la noción de finalidad y la de especie tienen escasa relación. El propósito de esta comunicación es mostrar que no es así y que no se puede entender la evolución histórica del concepto de especie sin, a su vez, investigar las diferentes posturas teleológicas.

Para llevar a cabo esta labor, se hará una breve reconstrucción de las principales teorías desde Aristóteles hasta el siglo XIX con el objetivo de mostrar cómo el finalismo ha acompañado a las distintas concepciones de especie y, con ello, las cuestiones fundamentales de la reflexión biológica. Se argumentará que la predilección por una teleología indeterminada implica la apertura a la posibilidad del transformismo. Finalmente, la comunicación se centrará en la escurridiza postura de Darwin en lo relativo a estos dos conceptos.

## Introducción

Antes de adentrarse en los diferentes autores, resulta oportuno detallar qué aspectos se van a resaltar de la noción de finalidad y cuáles son los diferentes sentidos en los que aquí va a ser tratada. Se van a manejar dos distinciones fundamentales:

- 1) Teleología (pre)determinada e indeterminada. En el primero de los casos el estado final está determinado, mientras que en el segundo el estado final no está prefijado (Cf. Diéguez, 2012).
- 2) Teleología intra-orgánica y teleología adaptativa. La primera es propia de cada organismo en sí mismo “causalmente dirigida a la consecución de un objetivo predeterminado”, mientras que la teleología adaptativa es la de un individuo para con su entorno, es decir, surge de las relaciones entre medio y organismo (Caponi, 2002, p. 59).

El paso continuo a una visión teleológica adaptativa a partir del siglo XVIII allanará el camino al transformismo y al evolucionismo, y con ello a una modifi-



cación radical del concepto de especie. Lo mismo pasará con el paso progresivo a una finalidad indeterminada.

## El esencialismo aristotélico

La discusión acerca del concepto de especie en Aristóteles ha sido causa de grandes debates. Fue a partir del artículo de David Balme (1962) que se llegó a cierto consenso acerca de la flexibilidad de los conceptos de género y especie y de la ausencia en la obra biológica aristotélica de una taxonomía sistemática. De acuerdo con Marjorie Grene (1974), el uso de los términos '*genos*' y '*eidos*' es, en cierto sentido, caótico, ligado al lenguaje común y cotidiano. No había en Aristóteles un proyecto complejo de ordenación taxonómica universal (Pellegrin, 1987), sino que sus diferentes clasificaciones dependían más bien del objeto de estudio (morfología, fisiología, etc.).

Aristóteles utiliza el concepto de *eidos* en sus tratados biológicos fundamentalmente como forma (Balme, 1962), para designar un conjunto de rasgos característicos y esenciales. El *eidos* no es más que la posibilidad de eternidad que le es arrebatada a los individuos con la imposición de la vejez y la muerte (Papavero et al., 1995, p. 128). Esta esencia transgeneracional es definición de la especie y determinante del individuo. Como señala John Wilkins (2010), es en la Grecia clásica donde surge esta "concepción generativa de la especie" que se mantuvo hasta la Edad Moderna.

La esencia aparentemente puede identificarse con el alma, como aquello que es propio y germen de la vida. El alma es forma, acto, entelequia (De Anima II, I, 412a10). El alma es esencia y definición de los seres naturales; es su *telos*. Sin embargo, de acuerdo con Etienne Gilson (1980), "la noción de alma depende de la de finalidad y no a la inversa" (p. 45). Como se puede apreciar, el fijismo es inherente a tal doctrina, pues si el alma define a un ser vivo y esta es traspasada por el padre a través del semen; no cabe más que las especies sean inmutables. No se puede entender el concepto de especie sin percatarse de que ese *eidos* definitorio, que forma parte del poderoso *telos* de la naturaleza, estipula un orden y atribuye un sentido a todos los movimientos naturales.

A modo de resumen de lo expuesto, hay dos principales características en el concepto de especie aristotélico: 1) el fijismo y 2) la sumisión de la causa eficiente sobre la final. Sirva de ejemplo el siguiente párrafo en cuanto a su conjugación:

[...] el germen constituyente debe existir ya con tal potencialidad; luego, que lo que produce existe con anterioridad no sólo lógicamente, sino también temporalmente: así el hombre engendra un hombre, de modo que, al tener tales características aquél, el proceso de formación de este otro se produce de tal manera (PA I, 640a20–27).

Este párrafo expresa con fuerza cómo cualquier proceso animal tiene su razón de ser en su *telos*, que define a la especie. Dicho de otra manera, los procesos biológicos se dan para realizar la esencia prefijada e innata. Esto quiere decir que, si la esencia pasa de padre a hijo, también el propio *telos* y, por lo tanto, cualquier proceso natural ha de ser el mismo en cada generación de una misma especie.

Se puede apreciar, por ende, que finalismo y fijismo son interdependientes en la biofilosofía aristotélica.

## El auge de la taxonomía y las semillas del anti-realismo

Tras una Edad Media en la que la ciencia biológica quedó relegada a un segundo plano en pro de la medicina, el ámbito zoológico y botánico adquirió un nuevo interés en la Edad Moderna. Sí que hubo una gran discusión medieval en torno al tema de los universales, que el propio Aristóteles dejó abierto debido a la ambigüedad de su realismo moderado. Sin embargo, los nuevos naturalistas renacentistas, mantenían la concepción generativa y, con ello, la creencia firme en la existencia de especies bien definidas en la naturaleza.

Ahora bien, si alguien puso su confianza y esfuerzo en encontrar esas delimitaciones fijas que dividen en partes el mundo de lo vivo ese fue Carl von Linneo. El naturalista sueco, identificó la naturaleza con la obra de Dios, y al hacerlo le dio una finalidad –ser admirada– y un sentido al ser humano: adorarla y desentrañar los mecanismos de esta (Gilson, 1980, p. 87). Así, parte de la asunción de que todos los seres naturales son “obra de Dios” y defendió una postura fijista con claro eco aristotélico:

As there are no new species (1); as like always gives birth to like (2); as one in each species was at the beginning of the progeny (3), it is necessary to attribute this progenitorial unity to some Omnipotent and Omniscient Being, namely *God*, whose work is called *Creation*. This is confirmed by the mechanism, the laws, principles, constitutions and sensations in every living individual (Linneo, 1735/1964, p. 18).

En el párrafo anterior no solo se afirma que las especies se mantienen debido a que cada ser produce un ser similar, sino que también se alude a la especie como unidad que gobierna el orden. Esto último parece adherirle al realismo y a la concepción generativa de la especie. Para Linneo, las especies son inamovibles, al menos para el primer Linneo. Las clases naturales tienen como principal característica la perpetuidad de sus propiedades (Ruse, 1987, p. 229). Por ello, la especie será también la unidad taxonómica de Linneo, y su inmutabilidad es un axioma necesario para cualquier clasificación.

La suposición de un creador intencional daba irremediablemente un carácter (pre)determinado a la teleología de Linneo. Las especies, generadas por obra divina, tienen como fin el que Dios les haya impuesto. La teleología intra-orgánica es evidente también: cada ser es perfecto y ha de desarrollarse bajo el plan divino.

Sin embargo, el afán taxonómico y sistemático y las ideas realistas del primer Linneo tenían su enemigo natural en Buffon, para el cual no existen más que los individuos y las clasificaciones no son más que categorizaciones humanas que introducen a la fuerza el mundo vivo en cajones demasiado estrechos.

Buffon se vio incapaz de afirmar la discontinuidad entre especies, pues estas no están delimitadas de forma precisa, existen cantidad de eslabones intermedios que no permiten dividir clara, concisa y tajantemente a las diferentes especies:

Mais la nature marche par des gradations inconnues, et par conséquent elle ne peut pas se prêter totalement à ces divisions, puisqu'elle passe d'une espèce à une autre espèce, et souvent d'un genre à un autre genre, par des nuances imperceptibles; de sorte qu'il se trouve un grand nombre d'espèces moyennes et d'objets mi-partis qu'on ne sait où placer, et qui dérangent nécessairement le projet du système général (Buffon, 1884, p. 370).

Por otro lado, Buffon asumió la variabilidad provocada generalmente por las circunstancias climáticas y alimenticias. Abre, así, el paso a la teleología adaptativa. Sin embargo, este cambio cualitativo no interrumpía el prototipo permanente de la especie. Dicho de otra manera, existe una especie de transformismo, pero interno a las propias especies, es decir, los individuos de las propias especies varían sin que por esto haya una mutación de especie.

## El transformismo finalista de Lamarck

En el siglo XIX, fijismo y transformismo tuvieron en Cuvier, Saint-Hilaire y Lamarck una de sus mayores disputas. El transformismo evolucionista de este último fue tan rompedor que fue condenado a la desatención de sus coetáneos. Lamarck contaba con un gran número de predecesores evolucionistas que se remontaban a la antigua Grecia, pero solo él fue capaz de conformar esas ideas en una teoría estructurada.

Lamarck asevera que existe un mismo orden, indivisible, continuo, sin saltos. La escala natural dejó de tener forma lineal para adquirir forma de árbol con numerosas ramificaciones (Lamarck, 1809/1986, p. 54) que culmina con el hombre. Esta continuidad tiene como núcleo central: *lo simple es el origen de lo complejo*. Sobre la génesis de lo simple, Lamarck acude a una respuesta típica, a saber, la generación espontánea.

Ahora bien, ¿cómo es posible o bajo qué medios se produce ese progreso de lo simple a lo complejo? Su respuesta revela su finalismo (pre)determinado: existe una tendencia en lo vivo a la complejidad, una propensión de la naturaleza al perfeccionamiento. Toda tendencia supone no solo un fin al que tender, sino un fin activo o causal que impele su búsqueda. Las fuerzas naturales y los impulsos de la materia viva son los principales promotores del cambio, pero, si estos fueran los únicos motores, la escala natural sería lineal. La razón por la que esta es ramificada se debe a las causas externas, es decir, circunstanciales, medioambientales. De esta forma, la modificación del entorno supone, en muchas ocasiones, un cambio en las necesidades de los seres vivos que viven en él. Este cambio en las necesidades lleva irremediabilmente a que los seres vivos varíen sus hábitos. Si los nuevos hábitos implican, de forma prolongada, la utilización de un órgano o, por el contrario, una disminución del uso de otro, estos se verán alterados por tal situación. El uso de un órgano lo fortifica, lo impulsa para su desarrollo; el desuso todo lo contrario.

En Lamarck hay una mezcla clara y contundente entre dos finalismos: uno necesario o determinado, *dirigido a un fin*, proveniente de una naturaleza que tiende a lo complejo desde sus orígenes; pero también uno más de corte adaptativo e indeterminado, que proviene de la interacción entre individuo y entorno y su preservación por la transmisión hereditaria. Este último sería, por tanto, no dirigido e inherente a una postura transformista. Sin embargo, la existencia del humano como cumbre de la escala natural parece indicar que hay una cierta subordinación del segundo respecto al primero. De ahí que, aunque no sea linealmente, sino de forma ramificada, el plan de la naturaleza hacia lo complejo termina por cumplirse.

## Darwin: teleología adaptativa y lucha por la existencia

La verdadera revolución de Darwin consistió en desvincular la adaptación de cualquier diseño, orientación o plan dirigido intencionalmente y dar un verdadero carácter útil a las adaptaciones. No hay una preordenación de la evolución, tampoco hay un sentido o dirección en ella, no va de lo simple a lo complejo, no cambia lo imperfecto por algo más perfecto.

Hasta el darwinismo, la relación entre el ambiente y los seres no iba mucho más allá de la propuesta buffoniana de los efectos del clima y la alimentación (Caponi, 2006). Más que adaptación, esto parecía revelar una *aclimatación*. El gran valor de Darwin fue su capacidad para explicar el transformismo a través de un mecanismo terrenal y coherente con sus observaciones. La selección natural es el principal motor del cambio (aunque no el único) dentro de las especies –al ser el mecanismo gradual a través del cual unos rasgos triunfan sobre otros– y la herencia es la forma en la que se mantienen. Es fácil apreciar, después de esto, que mientras Lamarck explicaba a través de la interacción entre el individuo y su entorno el surgimiento de los rasgos adaptativos, Darwin, a través de esa interacción, explica el triunfo y extensión de dichos rasgos, pero no el origen de las variedades como tal. Darwin afirmaba que las variaciones eran aleatorias, con lo cual quiere decir dos cosas: que su causa es desconocida y que no hay un finalismo detrás del surgimiento de las diferentes variedades, es decir, que no hay relación entre la génesis de las variaciones y las posibles utilidades que pueden tener en el entorno.

La selección se produce sin seleccionador, las variaciones no responden a un fin. ¿Dónde está entonces la teleología? Pues bien, las variaciones no responden a un fin, pero su mantenimiento sí. Y es que muchas veces Darwin trata a la selección natural como el resultado del proceso, pero otras muchas (metafóricamente o no) la supone como una especie de causa o fuerza activa (Diéguez, 2012, pp. 60–62). Ante un entorno hostil, donde hay problemas de escasez de alimentos, depredadores, enfermedades, etc., surgen soluciones, o, mejor dicho, se mantienen las mejores soluciones *para* los diferentes problemas. No se explica la *causa* de las variaciones, pero sí el *porqué* de su mantenimiento y extensión; y resulta que ese *porqué* es un *para qué*. La forma que tiene el biólogo de averiguar el *porqué* de un determinado rasgo es indagar en el *para qué* de este, qué función

cumplía, para qué era útil. En toda adaptación no solo hay un *qué* o *quién* se adapta, sino también un *respecto a qué*. La selección natural hace tender hacia la perfección local o relativa a las especies (Darwin, 1872/1988, p. 253).

Respecto a la noción de especie, Darwin asumió la imposibilidad de diferenciar lo que son las variedades de lo que son las especies. Este pasaje descubre su antirrealismo:

Por estas observaciones se verá que considero la palabra especie como dada arbitrariamente, por razón de conveniencia, a un grupo de individuos muy semejantes y que no difiere esencialmente de la palabra variedad, que se da a formas menos precisas y más fluctuantes (Darwin, 1872/1988, p. 104).

Darwin, no podía definir o delimitar lo que era una especie, pero tampoco podía prescindir de ellas; su teorizar dependía de tal concepto. Especie y finalidad no se pueden desligar fácilmente. Había herido de muerte a la teleología predeterminada y a la intra-orgánica. Sin embargo, aún dejó ciertos principios disposicionales metafísicos en los seres vivos que imposibilitaban la reducción mecanicista en el estudio de la evolución. Darwin no sabía cuáles eran las razones por las que los individuos cambiaban, pero sí sabía por qué algunos de esos cambios se mantenían y provocaban modificaciones graduales en las especies. Pero todo esto supone un principio de corte claramente metafísico: la lucha por la existencia como “lucha universal”. Lo vivo lucha por vivir. ¿Acaso es tan obvio que todo ser tienda a preservarse? La variación y la selección serían inocuas sin este afán de supervivencia, sin ese intento de imposición de lo individual. Esto quiere decir que la vida no es resultado de la evolución, sino que es la causa (Hernández-Pacheco, 2014: 75), es decir, no es la evolución la causa que permite la vida a través de la adaptación, sino que es la propia vida, esto es, el propio afán por vivir lo que pone en marcha a la evolución.

La supervivencia es fundamentalmente reproductiva, que “incluye no solo la vida del individuo, sino también el éxito al dejar descendencia”, pues, al fin y al cabo, “todos y cada uno de los seres orgánicos puede decirse que están esforzándose hasta el extremo por aumentar en número” (Darwin, 1872/1988: 113 y 117). ¿Cómo seguir afirmando solo a los individuos cuando parece que todo depende de su afán de supervivencia a través de la reproducción, o sea, a través de la especie? El finalismo de supervivencia darwiniano requiere, en definitiva, de la noción de especie como algo efectivo que tiene consecuencias en el mundo.

## Conclusiones

Tras este breve repaso histórico se puede concluir que existe una dependencia mutua entre el concepto de especie que la biología maneja y la perspectiva teleológica imperante. De esta forma, un concepto de especie realista y basado en el fijismo requería de un finalismo esencialista, determinado e intra-orgánico. En cambio, la apertura al *transformismo* y a una postura continuista entre las especies va de la mano de una teleología adaptativa e indeterminada. Como se aprecia en la teoría de Darwin, la teleología se vuelve cada vez más sutil y

oculta, pero sigue siendo condición necesaria para entender el funcionamiento de nociones esenciales de la biología evolucionista.

## Agradecimientos

Esta publicación es fruto de la investigación realizada gracias a la beca predoctoral cofinanciada por la Consejería de Educación de la Junta de Castilla y León y el Fondo Social Europeo.

## Bibliografía

- Aristóteles (2000). *Partes de los animales*. Madrid: Gredos.
- Aristóteles (2000). *De anima*. Madrid: Gredos.
- Balme, D. M. (1962). ΓΕΝΟΣ and ΕΙΔΟΣ in Aristotle's Biology. *The Classical Quarterly*, 12(1):81–98.
- Buffon (1884). *Ouvres complètes de Buffon*. Nouvelle édition. París: Librairie Abel Pilon, A. le Vasseur, Succ. Disponible en <http://gallica.bnf.fr/ark:/12148/bpt6k62187908/>
- Caponi, G. (2002). Explicación selectcional y explicación funcional: la teleología en la Biología contemporánea. *Episteme: Filosofia E Historia Das Ciencias Em Revista*, 14:57–88.
- Darwin, C. (1988). *El origen de las especies* (6ª edición). Madrid: Espasa-Calpe.
- Diéguez, A. (2012). *La vida bajo escrutinio. Una introducción a la filosofía de la biología*. Barcelona: Intervención Cultural.
- Gilson, E. (1980). *De Aristóteles a Darwin (y vuelta)*. Ensayo sobre algunas constantes de la biofilosofía. Pamplona: Eunsa.
- Grene, M. (1974). Is genus to species as matter to form? Aristotle and taxonomy. *Synthese*, 28:51–69.
- Hernández-Pacheco, J. (2014). Evolución, erotismo y origen de las especies. De vuelta desde Darwin a Platón y Aristóteles. *Naturaleza y Libertad. Revista de estudios interdisciplinarios*, 4:63–95.
- Lamarck, J. B. (1986). *Filosofía zoológica*. Barcelona: Editorial Alta Fulla.
- Linneo, von C. (1964). *Systema Naturae. Facsimile of the first edition*. Trad. Engel-Ledeboer y Engel. Nieuwkoop: B de Graaf.
- Papavero, N., Pujol, J. R., y Llorente, J. (1995). *Historia de la biología comparada. De Descartes a Leibniz (1628-1716)*. México D. F.: Las Prensas de Ciencias.

- Pellegrin, P. (1987). *Aristotle's classifications of animals: Biology and conceptual unity of Aristotelian corpus*. Berkeley y Los Ángeles: University of California Press.
- Ruse, M. (1987). Biological Species: Natural Kinds, Individuals, or What? *The British Journal for the Philosophy of Science*, 38(2):225–242.
- Wilkins, J. S. (2010). What is a species? Essences and generation. *Theory in Biosciences*, 129(2-3):141–148.

TALLER SOBRE NANCY  
CARTWRIGHT



# Causal pluralism and the boundaries of manipulation

**Renata Arruda**

Universidade Federal de Goiás

The human organism, object of study of medicine, imposes great challenges for its practice. Despite the biological variability of human beings and the environmental peculiarities, medical procedures are predominantly successful. The unlimited connection that takes place in the whole allows, therefore, intended changes in practice. But on what to intervene in this complexity? One of the ways medicine establishes the delimitation of the causal factors set is by the real possibility of manipulating them. This focus reveals new philosophical perspectives on causality. One of the most important approaches to this question is offered by Nancy Cartwright (1999, 2007) from her analysis of the theory of manipulability, such as that developed by James Woodward (2003) and Daniel Hausman and James Woodward (2004). But their theory imposes severe limitations to the practice, which is fundamental to the sciences. My purpose here is to show, in agreement with Cartwright, that the plurality of distinct scientific activities fully justifies the plurality of conceptions of causality. I also accept Cartwright's challenge for the unification between the hunting and the use of causes (Cartwright, 2007, p. 1), arguing, finally, that manipulation is capable of promoting the connection between these notions in the activity of medicine.

In the more general concept of manipulability, the causal relation occurs as a result of a change in the cause that, consequently, produces the effect. This general conception meets quite satisfactorily the purposes of medicine, especially because this area of knowledge is dedicated to the promotion of changes by means of human action, in its various possibilities. However, a more specific formulation of manipulation, such as that proposed by Woodward and analyzed by Cartwright, creates a series of barriers to its application in practice.

In more specific terms, the idea of manipulation is developed by Woodward (2003, p. 249) based on the concept of invariance under interventions. For him, intervention is an exogenous action, an interference that is independent of the processes or internal occurrences of a given relation between causally related factors, so that the changes that follow the intervention can be judged only as a result of it.

These restricted results are consequences that come from the conception defined by Hausman and Woodward (2004, p. 848–849) as modularity. In this condition, each of the minimum causal relations is defined as being isolated as a “module”, which can be manipulated without interfering in the operation of the others.

Traditionally, the investigation of causality in philosophy focuses on very specific and delimited factors, assuming absolutely controlled conditions, as prescribed

in the so-called “Galilean experiments.” This is how Cartwright describes such proposals as Woodward’s for causality. For her, on the one hand, it serves perfectly for the methodology with which causes are “hunted”. The isolation of causal factors is one of the means through which one can, according to Cartwright, come to know the “contribution” that a factor exerts on a phenomenon (Cartwright, 2007, p. 223). On the other hand, she considers the model of the Galilean experiment quite artificial for the analysis of practical contexts, since it does not covers satisfactorily the complexity and the various influences of reality. Thus, the Galilean experiment does not serve the use of the cause, it does not dialogue with the conditions of its application outside the idealized environment (Cartwright, 2007, p. 51).

In agreement with Cartwright’s criticism, I believe that Woodward’s ideas reflect a very narrow perspective for the definition of causality. Its specification that the peripheral variables of a given system must be fixed seems to be an impediment to its application in several fields of several sciences and imposes a very artificial restriction. It goes in the opposite direction to what is expected of the practice of medicine, where any intervention must be thought in order to be incorporated into the context of a series of factors that are acting simultaneously. The functioning of the organism is characterized by the articulation between complex, dynamic systems, which are not always controllable because of their sensitivity to the initial conditions inherent to these systems.

I insist, therefore, that the use and the characterization of causes in medicine depends, essentially, on the possibility of manipulation in its more general conception. Although a more specific formulation such as Woodward’s has led to incompatible restrictions with medical practice, the concept of manipulability cannot be completely abandoned. I argue that a causal factor that can be treated in the field of health sciences is the one that is characterized by being manipulable. Manipulability acts as a methodology for hunting causes in medicine, pointing out what they are and how they can be used.

But why, after all, defend a specific characterization of cause for medicine? Because the objectives pursued in each branch of knowledge require that the concepts that form part of its theoretical framework can be employed in function of its practices. One of Cartwright’s main criticisms is that the theories that propose an absolute conception of cause fail precisely because, in the end, they are only suitable for specific areas. But that is not a bad thing. Basic scientific notions, as they occur in physics, are elaborated so as to dispense with specific assumptions of practice to be considered legitimate. However, practice, undeniably, imposes demands that theoretical formulations are not able to solve or even predict. Theoretical proposals developed to be rendered universal end up functioning effectively only in specific domains, indicating that method, cause and use are multiple, are variable and relative to the different contexts to which they apply.

Understanding the notion of cause through manipulation concerns a vast and important scope: the scope of everything we can touch or influence, directly or indirectly—which is no small matter. Even if it is not possible to explain how a galaxy attracts the other gravitationally through manipulability, as Woodward points out (2013, p. 8), a series of other phenomena, which are not few, can be explained—economical, physical, chemical, psychological, biological phenomena.

One of the results of the definition of the concept of causality from such different perspectives is that the concept itself is not immune to the multiplicity of uses to which it applies. The adoption of a causal pluralism permits the transference of the diversity of uses to the very conception of cause. The acknowledgement of the importance of use, as championed by Cartwright, is once again pressing for the philosophical investigation of science. This consideration is indispensable to narrow the gap between the practice and the philosophy about the practice. In this sense, assuming the need to manipulate causes, as in the practice of medicine, does not simply imply a kind of explanation of occurrences of facts or events, but a condition of possibility of the existence of the cause itself. If a doctor diagnoses a patient with a disease that is intractable, that is where the activity of medicine ends: the impossibility of changing physiological states makes the very existence of medical practice impossible. Without the manipulation of an individual's state of health there would be biomedical sciences, but not medicine.

In response to Cartwright's challenge, I believe that medicine is one of the fields in which questions that seek to characterize methodology, metaphysics and use, can be answered by manipulation. The philosophical elaboration of the concept of use through the understanding of methodology is fundamental to the connection of the very notion of causality with the sciences, allowing causality to inhabit not only the realm of the impalpable metaphysics but also that of concrete reality. By applying the boundaries of manipulation to the boundaries of the causal set I offer, finally, an answer to the author's proposal that the methodology should be able to establish both metaphysics and the use of causality in the practice of medicine.

*This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.*

## References

- Cartwright, N. 1999. *The Dappled World – A Study of the Boundaries of Science*.
- Cartwright, N. 2007. *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge UP, Cambridge.
- Hausman, Daniel; Woodward, James. 2004. Manipulation and the Causal Markov Condition. *Philosophy of Science*, 71 (December), p. 846–856.
- Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press. Oxford.
- Woodward, James. 2013. Causation and Manipulability. *The Stanford Encyclopedia of Philosophy* (Winter Edition), Edward N. Zalta (ed.). <<http://plato.stanford.edu/archives/win2013/entries/causation-mani/>>.

# In the beginning there were no laws of nature and now what is out there? On Nancy Cartwright's "dappled" causal explanations

Laura Bujalance

When in 1983 Nancy Cartwright's wrote *How the Laws of Physics Lie*, it was only just foreseen how much of an impact her claim that the covering-law model of explanation is a mistaken one would have. At that time, the claim was just made against both Hempel's models of explanation (DN and IS) and statistical relevance models of explanation (Suppe and Salmon's models). Contrary to those covering-laws theorists—upon whom the idea that there is a law that covers every case in nature was established—Cartwright stated: "I imagine that natural objects are much like people in societies. Their behavior is constrained by some specific laws and by a handful of general principles, but it is not determined in detail, even statistically. What happens on most occasions is dictated by no law" (1989: 49).

Which was the bottom line of such a claim? Cartwright declared in a report in 2003 that the dominant topics in general philosophy of science during the second half of the 20<sup>th</sup> century were 'the nature of scientific theory', 'realism' and 'explanation'. Each of those topics underwent shifts by the end of the century: theories about 'the nature of scientific theory' moved from a Deductive System's framework to a Semantic View's one; realism began to be a real deal as a theory when the prohibitions of positivism were forsaken; and talking about truth or models was, not only allowed but encouraged; and the hunt for 'explanations in science' began to focus on 'causation'. Cartwright's claim about Laws of Nature originated within this later debate, so the bottom line is that we need new accounts for what is an explanation in science.

But her claim has also had an impact in all the rest of the topics she mentions. First of all, because to position her new claim that laws of nature are not good enough to explain phenomena because they don't state facts—and so they are not entitle to truth—she had to deal with issues of both the realism/anti-realism debate—and so, she addressed a variety of theories, such as van Fraassen's instrumentalism, Putnam's internal realism or theories involving Inference to the Best Explanation—and the nature of scientific theories debate. Nevertheless, her main concern was about 'explanation': since laws of physics did not provide description of facts, there is a need to account for them being relevant to scientific explanations; for Cartwright (2003) that account comes from causation, hence she indentifies explanatory relevance with causal relevance.

Why do laws of nature don't state facts? Cartwright always resorts to the exam-

ple of the law of gravity to explain so: if there is a body which is charged and it's influenced by another charged body, it is true that those bodies are attracted to each other because of gravity, but the law of gravity wouldn't explain those bodies' behaviors since Coulomb's law is at stake in this situation as well. Thus, Cartwright (1983) argues in favor of causal explanations against covering-law explanations—which she later (1989) renamed 'fundamentalist covering-laws explanations'.

Since Cartwright's *rationale* for supporting this account is widely known, it is enough here to highlight some of the main features involved with her theory. First of all, it is important to bear in mind Cartwright's "*affiliation*" to the Stanford School of Philosophy of Science, which focuses primarily on Science in Practice. Because of it, arguing from case studies of scientific practice is a frequent and powerful resource in Cartwright's writing. She studies not only cases from physics—mainly from several theories of radioactivity and exponential decay in quantum mechanics—but also from social sciences, such as Mill's composition of causes principle. And, in doing so, she makes her cases about what is a real explanation of phenomena. Her insight on the cases she studies is so sharp that it's difficult not to be drawn to the conclusions she draws: an explanation *tout court* has to meet the requirement of non-redundancy and only causation does so. Similarly, she maintains that there are redundant theoretical treatments for explaining a phenomena that compete with each other, but there are not redundant casual accounts; why?, cause "causes make their effects happen" (1983: 76).

Second, causation in Cartwright's sense it's not instrumental because it cannot be reduced to probabilistic laws (and here is where Mill's principle comes in handy). For Cartwright, asking for what makes an effect to take place means asking for its cause; and accepting that as an explanation implies accepting the existential component of the cause. As Suarez puts it "Cartwright takes the reality of the cause to be an intrinsic characteristic of the causal explanation: pointing to a nonexistent cause cannot explain anything. A cause can only constitute a genuine explanation if it actually exists" (2007: 142).

Lastly, there is no clear-cut definition of 'cause' in Cartwright's *corpus*. We do certainly have some hints about what the concept 'cause' is related to: a) well-tested claims, i.e. via manipulation, of hypothetical causes and effects do warrant the causes they pose; b) causes carry the content of scientific knowledge as opposed to what laws do, which is making predictions and precise calculations; c) the general metaphysical picture behind Cartwright's concept of 'cause' is the Aristotelian belief in the variety of the concrete and, in this sense, for Cartwright (1989), causal interactions are interactions of causal capacities that, in turn, are predicated from properties or 'causal powers' but not from individuals; d) causal hypothesis can be tested as any other scientific hypothesis (Cartwright, 1989); and, e) causal relations are dependent on causal-systems (Cartwright 2003), so a relevant causal-relation in one system might not be relevant in another causal system.

In this paper I will address some of the consequences of this "dappled" account of explanatory causation, pointing out what may prevent us from fully embracing Cartwright's philosophy of science.

- To a) The use of the term ‘test’ within Cartwright’s *corpus* is misleading. She identifies it with experiments (1983 and 1989), and with some other methodologies in more recent works (2009). Plus, she gets the most of the warrant for her concept of ‘explanatory cause’ by using quantum mechanics experiments, although several physicist have pointed out that there are cases, such as the ‘double-split’ experiment, where causation is not a relevant feature of quantum physics unless causation is singular—which Cartwright herself (1989 and 2017c) denies.
- To b) Cartwright (2007) tackles the issue of the content of science claiming that, ultimately, causes are what can fill the variable terms of laws. She constantly uses the case of  $F = ma$ , where the variable  $F$  is filling in by a capacity, i.e. the “power of attracting bodies”. But that power, in turn, has to be filled in some other physical sense, cause “attracting” doesn’t mean the same in the context of classic mechanics and relativity. So, when do we stop filling in?
- To c). Lately, Cartwright (2017a) has made some clarifications about her metaphysics, stating three distinct categories involved somehow in causal explanations: ‘powers’, their ‘contributions’, and the result that happens when they exercise. But, although she has advocated for the measurement of powers and overall effects, there is no an answer yet as to the status of this so called “contributions”.
- To d). ‘Testing’ is one of the key terms for Cartwright’s defense of causal explanations. As she states “[a]s a rule of thumb it is fair to say that a controlled experiment is the best way to test a causal claim” (2002: 9). But she grounds her non-acceptance of laws of nature as explanation for phenomena on the fact that those laws do never happen at *face value*. So, if phenomena are singular, why not tests, especially the ones within the field of quantum physics? Can tests be taken at *face value*?
- To e). The only feature that Cartwright’s appears to link to the concept of cause is “invariance-under-intervention” (2017b: 138). But, in a plurality of causal systems, is that enough to tell apart causes from mere associations? According to Cartwright (2017b), it is so when “regularities can be derived from the facts about capacities” (151), but a complete answer is far from being given.

In summary, Cartwright took laws of nature as explanations of scientific phenomena away from us; and left us with a set of “causal systems” whose explanatory nature is “entangled” to terms such as ‘tests’, ‘capacities’, ‘contributions’ and so on. Maybe, we all should get loose of laws of nature, but there are still some issues to be addressed for us to be able to hold on Cartwright’s alternative explanatory model.

## References

- Cartwright, N. (2017a). Causal Powers: Why Humeans Can’t Even Be Instrumentalists. In *Causal Powers*. Jacobs, J. New York: Oxford University Press.

- Cartwright, N. (2017b). Can Structural Equations Explain How Mechanisms Explain? In *Making a Difference: Essays on the Philosophy of Causation*. Beebe, H., Hitchcock, C. & Price, H. Oxford: Oxford University Press. 132–152.
- Cartwright, N. (2017c). Single Case Causes: What is Evidence and Why? In *Philosophy of Science in Practice*. Chao, H-K & Reiss, J.: Springer. 11–24.
- Cartwright, N. (2009). If No Capacities then No Credible Worlds. But Can Models Reveal Capacities? *Erkenntnis* 70(1): 45–48.
- Cartwright, N. (2007). What Makes a Capacity a Disposition? In *Disposition and Causal Power*. Kistler, M. & Gnessounou, B.: Ashgate. 195–205.
- Cartwright, N. (2003). “From Causation to Explanation and Back”, Technical Report CTR 09–03. CPNSS, LSE.
- Cartwright, N. (1989). *Nature’s Capacities and Their Measurement*. Oxford University Press.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press.
- Suarez, M. (2007). Experimental Realism Reconsidered. How Inference to the Most Likely Cause Might Be Sound. In *Nancy Cartwright’s Philosophy of Science*. Hartmann, S., Hofer, C. & Bovens L.: Routledge. 137–163.

# Reflecting on Nancy Cartwright's conception of external validity of Randomised Controlled Trials<sup>1</sup>

Mikkel Holding Vembye  
Danish School of Education  
Aarhus University, Denmark  
mihv@edu.au.dk

## Introduction

This symposium paper aims to discuss Nancy Cartwright's criticism against randomised controlled trials (hereafter RCTs), and especially *the use* of the knowledge extrapolated from RCTs. The primary focus of this paper will be on Cartwright's perception of the *external validity* of RCTs. The paper aims in particular to discuss the prospects and boundaries of the generalisability of RCTs. In this regard, *external validity*, sometimes described as *extrapolation* (Reiss, 2018, p. 5), refers to the notion that "*random* selection... permits you to generalize your findings valid back to that population" (Murnane & Willett, 2010, p. 44) or in other words "whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables" (Cook, Campbell, & Shadish, 2002, p. 38). Furthermore, I will try to extend and relate these views and critiques to meta-analysis<sup>2</sup> and systematic reviews<sup>3</sup> applied in evidence-based research.

Throughout the presentation, I will show the significance of Cartwright's theorisation in relation to the evidence-based movement (Cartwright & Hardie, 2012), in particular within the field of Nordic educational research and policy-making. In that regard, I will demonstrate how the theory injects shining nuances into fields, which predominately focuses on hunting/testing causes and less of the use of that hunting. However, I will try philosophically to challenge some of the theoretical conception of the theory, especially the practical use of Cartwright's conception of the INUS conditions.<sup>4</sup>

---

<sup>1</sup>Many thanks to Hans Siggaard Jensen, Christian Kjeldsen, Henrik Kasch and Magnus Birkemose Nordam for valuable comments to the extended version of the abstract.

<sup>2</sup>Meta-analysis "refer to the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the finding" (Glass, 1976, p. 3). Meta-analysis "refer specifically to statistical analysis in research synthesis and not to the entire enterprise of research synthesis" (Harris, Hedges, & Valentine, 2009, pp. 6–7).

<sup>3</sup>"[S]ystematic reviews is a review of existing research using explicit, accountable rigorous research methods" (Gough, Oliver, & Thomas, 2017, p. 4).

<sup>4</sup>"An INUS condition is an *I*nsufficient but *N*ecessary part of an *U*nnecessary but *S*ufficient condition for producing a contribution to the effect" (Nancy Cartwright & Hardie, 2012, p. 25; Mackie, 1965, p. 24).



I want to direct my attention towards meta-analysis and systematic reviews for at least two reasons. First, because it seems to be the case that, “[s]ystematic reviews are the primary method for managing knowledge in the evidence movement approach” (Oakley, 2003, p. 23), and I consider RCTs to be a key part of this movement. Second, many statisticians recognise that meta-analysis of RCTs possibly transcend the deficiencies of individual RCT studies. John Ioannidis argues, for instance, that “research findings are more likely true in confirmatory designs, such as large (...) randomized controlled trials, or meta-analyses thereof” (2005, p. 698), and, “[b]etter powered evidence, e.g., large studies or low-bias meta-analyses, may help, as it comes closer to the unknown “gold” standard” (2005, p. 700). On related matters, Rex Kline argues that: “our data routinely lie, they lie through multiple types of error, and it is only when results are averaged over studies, such as in the technique of meta-analysis, that some of these errors begin to cancel out” (2015, p. 56). It seems to me that many researchers actually reflect on the shortcomings of single RCTs and single studies, but believe that the generalizability/external validity increases when research findings are pooled together. As an example, the Danish Clearinghouse for Educational Research (DCER) states in one of their systematic reviews that “a big part of the included reviews is based on meta-analysis, which implies generalizability in an *absolute sense*. The findings of these reviews reach (...) beyond the context, because in a meta-analysis other conditions and factors have been controlled for” [my italics and translation] (Nielsen, Tiftikci, & Larsen, 2013, p. 85). Consequently, DCER conclude that meta-analytical findings are “*directly* applicable for practice” [italics mine] (Dyssegaard, Larsen, & Tiftikci, 2013, p. 12). In a Danish context, the trust in the enumerations of meta-analysis is often based on the belief in Austin Bradford Hill’s (1965, p. 296) criteria/notion of *consistency* saying; if a relation is “repeatedly observed by different persons, in different places, circumstances and times,” then we have a generalizable causal claim. My claim is that this might be one of the reasons why we find an almost religious<sup>5</sup> trust in (meta-)meta-analytical works such as John Hattie’s Visible Learning (2009).

Although I believe that well-conducted meta-analysis and systematic reviews are able to cast light on the diverse efficacy and effectiveness of given interventions and practices, I will argue that the pooled result does not bring external validity simply because the pooled results can at best show us what has worked in a past states. It cannot tell us anything about the current state of the target population. Unless we believe in simple inductive reasoning, we need additional information and background knowledge about the target population besides the findings of meta-studies when we want to intervene on a policy-level or local level. Therefore, I will argue that (and explore if) the lesson Cartwright teaches us goes beyond the critique of single RCTs. It is my impression that many researchers wrongly ignore Cartwright’s theory due to the rigorous trust in meta-analysis and systematic reviews.

Finally, I intend to question Cartwright’s statement saying that, “we do not have good explicit methodologies for how to establish *tendency claims*” [my italics] (2011, p. 767). In that regard, I will probe the question whether well-

---

<sup>5</sup>For examples of the religious interpretation of the book, see Mansell (2008) stating that the book “reveals teaching’s *Hole grail*” and Egelund & Qvortrup (2014, p. 13) saying that, “the book should be a *bible* for teachers” [my translation].

conducted meta-analysis and systematic reviews are actually able to illuminate the concept of *stable tendencies* in the given causal semantics/model across different contexts, or, on the contrary, indicate whether stable tendencies are absent when we try to apply causal models widely? For my argument, I will use an extended version of Cartwright’s causal principles (CP).

## The significance of Cartwright’s theorisation

In the field of (Nordic) educational science—which is my main scientific theoretical field of interest—we are currently experiencing an extensive and increased use of RCTs. Since 2011, at least 39 new large-scale RCTs have been initiated in the Nordic countries within economics, political science and educational science (Pontoppidan et al., 2018, p. 325). Most RCTs are publicly funded—simply because policy-makers believe and argue that RCTs *by design* yield the highest degree of generalisability on whether given interventions should be disseminated (for these arguments, see the Danish Ministry of Education, 2014, p. 5). Without further hesitation, policy-makers of the Nordic countries in most cases—if the findings verify the political agenda—use the findings of the above-mentioned RCTs as key evidence for amendments of the Nordic educational legislation. I, therefore, think it is reasonable to argue that in the Nordic countries we experience a quite rigorous belief in what Julian Reiss characterises as *a foundationalist methodology*, which “maintains that certain methods are *intrinsically* more reliable than others” [my italics] (2018, p. 10). This belief is intimately connected to the *liberal form of experimentalism*, where RCTs are recognised as the *golden standard* and all other methods should “be ranked with respect to reliability according to how closely they mimic experiments” (Reiss, 2018, p. 9). Cartwright et al. have tirelessly contested and opposed the widely accepted conviction that RCTs (should) work as *the* “golden standard” and are the best method to ensure *internal* and *external validity*. Therefore, Cartwright is a significant voice against this kind of *methodological foundationalism*. Furthermore, Cartwright’s theories have turned out to be very powerful for illuminating the problems at stake in the current movement of evidence-based education in the Nordic countries (Kvernbeek, 2016; Vembye & Jensen, 2018).

Arguing against the external validity of RCTs, Cartwright and Hardie (C&H) show that all study populations (SP) of RCTs are governed by causal principles (CP) (1):

$$(CP) \text{ SP: } y(i)c = a_1 + a_2y_0(i) + a_3b(i)x(i) + a_4z(i) \quad (1)$$

“*i*’s range over the individuals in the population to which the causal principle applies,  $y(i)$  is the outcome,  $x$  is the policy variable,  $a$ ’s are constant across all individuals,  $a_{2...n}$  represent “boost factors”,  $y_0$  is a “base level” of  $y$  for  $i$ ,  $b(i)$  represents all the different factors in all the support teams that work with the  $x$ ,  $z(i)$  represents all other factors and their support teams that contribute additively with  $x$  but do not include  $x$ ” (2012, pp. 26–27).

According to C&H, the anticipation of what knowledge RCTs provide often excludes information of the above-mentioned other causally structurally relevant

factors (1) (2012, p. 27). This is often due to researchers expecting that *randomisation* assures that “ $x$  is probabilistically independent of  $y_0$ ,  $b$ , and  $z$ ” (2) (2012, p. 35), and therefore that the given cause/intervention,  $x$ , has worked with in many different structural factors contained in the given study population, which in turn makes the result highly generalizable. This often results in the widespread belief that “*ideal*” *RCTs* simply reflect the difference between receiving the treatment versus not receiving the treatment (2).

$$T = a_3 \text{Exp}(b(i)(X - X')) \quad (2)$$

$T$  signifies the treatment effect, “ $X$  is the value of the treatment in the treatment group and  $X'$  is its value in the control group” (2012, p. 34).

When one wants to apply the knowledge from an RCT, C&H argue that one cannot simply expect all these factors and especially not the support factor,  $b(i)$ , to be the same or to be equally distributed ( $\text{Prob}_{SP}(b(i) = B) = \text{Prob}_{TP}(b(i) = B)$ )<sup>6</sup> across different populations or even between the study population and the defined population. C&H argue that wanting to apply the knowledge from an RCT to a target population (TP) requires external knowledge about the structural factors that govern TP. Equation (3) reflects this condition.

$$TP: y(i)c = a'_1 + a'_2 y'_0 + a'_3 b'(i)x(i) + a'_4 z'(i) \quad (3)$$

$a'$  denotes that one does not know the relation between  $a$  and  $a'$ . More decisive, we are rarely in a position in the social domain to expect that TP contains the same support factors as in SP. Cartwright thereby shows us that it is much more complicated to expect that the claims *what-works-there = it-will-work-here*. What Cartwright concisely shows is that “[d]ifferent underlying structures yield different causal and probabilistic relations” (Cartwright & Munro, 2010, p. 210). Therefore, Cartwright actually thinks that the idea of external validity is the wrong idea, because for it to be valid it requires that the causal principles of SP is equal to the causal principles governing TP. According to Cartwright, this is a rare case in (the social) reality. Another way that Cartwright (Cartwright & Efstathiou, 2009) expresses her point is to show that CP of the population possibly changes every time one crosses different space/context ( $\Sigma$ ) and times (T). This means that one needs to draw new/different conclusions when one crosses these spatiotemporal points (see table 1). Many think of external validity as the concept of generalising the results from the study population valid back to the target population across settings ( $\Sigma$ ) and times (T). This). But that is where Cartwright show us that it is all-important to take *time* into account as well. As Cartwright (2011, p. 751) shows:

Even if the entire target population were enrolled in the study, predictions will be about future effectiveness where there may be no guarantee that this population stays the same over time with respect to the causally relevant factors.

A different way to put this is to say that every time we move from  $\Sigma_1, T_1$  to  $\Sigma_1, T_2$  (see table 1), we have to be aware of structural variances that might affect the result of the intervention, although one often intuitively thinks that the context is similar, for instance when doing research at fourth grade students across different year groups in the same country.

---

<sup>6</sup>SP=study population, TP=target population

Table 1: *Context and time*

| Space/context<br>( $\Sigma$ ) | Time (T)       |                |
|-------------------------------|----------------|----------------|
|                               | $\Sigma 1, T1$ | $\Sigma 2, T1$ |
|                               | $\Sigma 1, T2$ | $\Sigma 2, T2$ |

The last point I will draw attention to in this paragraph is Cartwright’s argument regarding the use of the findings of the average treatment effect<sup>7</sup> (ATE) deduced from RCTs. To know that an intervention has produced a positive ATE is not a deterministic concept, Cartwright argues (Cartwright, 2007a; Deaton & Cartwright, 2018, p. 15). This means that ‘X causes Y in  $\phi$ ’ is consistent with ‘X causes  $\neg Y$  in  $\phi$ ’ (Cartwright, 2007a, p. 15). I think, this statement has a significant practical impact, because it make practitioners like teachers and doctors aware of that they cannot inductively expect ATE to equal the concrete case. Raising the mean can be an important matter for policy-makers who work with larger systems/populations, but not knowing the mechanism behind given ATEs can actually ‘harm’ people or have no effect, when, for instance, teachers apply knowledge of ATE on the concrete class. Another important point that Cartwright stresses is that the “most difficult thing to predict, even if you are prepared to admit a reasonable degree of error, is the actual value of the outcome, individual by individual” (Cartwright & Hardie, 2012, p. 30). This makes it rather complicated to travel the way from research finding to the concrete practice.

## Critiques and discussions

One of the critiques against (Deaton and) Cartwright’s (D&C) (2018) notion of external validity is that it does not deliver a model for how one can predict the outcome from one (study) population to another (Pearl, 2018). Critics argue that D&C’s approach does not deliver any solution of how to compare and apply the knowledge of the structure of one population with another. In relation to this critique, I find that it would be interesting more deeply to discuss:

- i. What implications Cartwright’s approach on external validity could have on the way researchers and policy-makers extrapolate and use the knowledge from SP to TP?

If Cartwright is right, it implies that we can never build our causal models about the target population *beforehand* (Cartwright, 2011, p. 76). We always need to engage with local knowledge about TP before any prediction of what is going to happen when we intervene. Consequently, this will mean that policy-makers, researchers, and practitioners need to think much more carefully about their intervening than I think they do at the current state. As mentioned before, this is due to the (blind) trust in the return of meta-analysis and systematic reviews, which draw on inductive reasoning. But if we cannot simply use the knowledge

---

<sup>7</sup>ATE equals  $E[\theta] = E[Y^1 - Y^0]$  (Morgan & Winship, 2015), which basically is the same as equation (2).

<sup>8</sup> $\phi$  signifies a subpopulation of the given (study) population.

extrapolated from RCTs the question often becomes:

- ii. How do we determine when a well-conducted RCT is relevant for TP? And what is the application of the knowledge of well-conducted RCTs?

I basically agree with Cartwright's assumption that RCTs are only relevant for TP, when we have knowledge about the causal principles governing both SP and TP (Cartwright & Hardie, 2012, p. 23). To know that a salary bonus for teacher in India decreases days of absent does not require that the same will happen if the intervention is introduced in, for instance, the Nordic countries: simply because teachers in the Nordic countries already have a low number of days of absent and they have a decent wage. This example tries to illustrate how different support factors are in place in different places. Although it occasionally appears to me that Cartwright's advice of the use of support factors ends up in the same trouble as the advice of the US Department of Education saying that RCTs are only relevant when conducted "in school settings similar to yours". It seems to me that RCTs only become relevant when the support factors of SP equal the support factors of TP. This is a strong assumption and difficult to achieve, and solely applying RCTs when the context of TP equals SP will severely limit the use of RCTs. Nevertheless I think Cartwright teaches us a good lesson showing exactly what RCTs can tell us—if there exist a positive difference between the treatment and non-treatment group—"that X causes Y in some causally homogeneous subpopulation  $\phi$  of the study population" (Cartwright & Munro, 2010, p. 261). Therefore, RCTs can show us that X has had the ability to work in some settings. This is a good start when wanting to transfer an effective intervention from one place to another, but it still does not tell us if the study population shares the same causal principles as the target population, or whether the intervention can be expected to work in co-operation with other support factors. Therefore, one could, ask:

- iii. How is it possible to obtain knowledge about the *causal principles* (structures) that govern the SP as well as the TP? Can any methods deliver this knowledge?

To the above question, I will argue in line with Cartwright (2007b, p. 39) that there exists no universal methods for obtaining both the causal principles of SP and TP. To find out whether SP and TP share the same (or have different) causal principles requires additional knowledge about TP, which scientific studies cannot deliver as such. However, I find it difficult to find out how we should obtain knowledge about TP. Should politicians apply register data to a greater extent to recognise the current causal structure of TP before they intervene, for instance, in entire school systems? Or should teachers be consulted when new interventions are introduced etc. This is still an unsettled matter for me. I try to elaborate on these matters in my presentation.

Although, I agree with Cartwright that no methods *as such* can tell whether SP = TP, because as Cartwright among other things says, "time passes, [most] things change," I will contest Cartwright's argument that "we do not have good explicit methodologies for how to establish *tendency claims*" (2011, p. 767). Cartwright argues that we should use the conception of *stable tendencies* instead of external validity. Stable tendencies draw on the idea of whether a factor has the stable capacity to promote the effect across a range of situations under consideration

(2010, p. 262). Contrary to Cartwright's stance that we do not have any methods to find stable tendencies, I will argue that well-conducted meta-analysis and systematic reviews can indicate if given interventions contain stable capacities across different structures or not. I think meta-analysis and systematic reviews allow us to get an overview of the effect of an intervention in various situations. I try to show this fact by using Cartwright's causal principles in an extended version in equation (4).

$$\begin{aligned} y(i_1)c &= a_{1,1} + a_{2,1}y_{0,1}(i_1) + a_{3,1}b(i_1)x(i_1) + a_{4,1}z(i_1) \\ y(i_2)c &= a_{1,2} + a_{2,2}y_{0,2}(i_2) + a_{3,2}b(i_2)x(i_2) + a_{4,2}z(i_2) \\ y(i_3)c &= a_{1,3} + a_{2,3}y_{0,3}(i_3) + a_{3,3}b(i_3)x(i_3) + a_{4,3}z(i_3) \\ y(i_4)c &= a_{1,4} + a_{2,4}y_{0,4}(i_4) + a_{3,4}b(i_4)x(i_4) + a_{4,4}z(i_4) \\ y(i_5)c &= a_{1,5} + a_{2,5}y_{0,5}(i_5) + a_{3,5}b(i_5)x(i_5) + a_{4,5}z(i_5) \end{aligned}$$

The second submerged numbers 1-5 and submerged numbers around  $i$  represent different contexts ( $\Sigma$ ) and times (T). My argument will be that if we use this conception wisely it allows us to know when we have an intervention with stable tendencies or if the intervention just has momentary effects. I actually think that this way of thinking allows us to know when and when not it can be reasonable to draw on Pearl's graphical theory of causation (Pearl, 2009) which mostly presupposes tendencies. I will not argue that this gives us a general licence not to seek any further knowledge about given TPs, but I think this could represent a wise use of how research can inform decision-making. It can indicate whether the intervention contains a stable tendency, and thereby, whether the intervention can be expected to have a wider use. And on the contrary a well-conducted meta-analyses or systematic review might be able to show how/if the intervention has different effects across different settings. This could elucidate when it would be an unwise idea to apply the intervention widely.

However, I think to realise the above-mentioned potential of meta-analysis and systematic reviews requires a renewed idea of the conception of synthesising results across different study populations, which often forces simplified conclusions (Vembye & Jensen, 2018). Furthermore, I will argue that this way of thinking allows us to understand how interventions previously have worked in between various support factors (moderator variables) and how causes are mediated to their effect (mediator variables) in different ways.<sup>9</sup>

For another philosophical point of discussion, Cartwright describes CP as an attempt to represent the notion of INUS conditions,<sup>10</sup> but one of the problems about the INUS condition is that it lacks a clear distinction between causal mechanisms and circumstantial conditions (Pearl, 2009, p. 314). This means that it is difficult to distinguish between relevant causal factors from irrelevant other circumstantial present factors. Even though, C&H (2012, p. 187) argue: "that causes are INUS conditions. But not (...) all INUS conditions are causes," it is often difficult from a perspective of the INUS condition to see:

- iv. How to determine which factors represent conditions and which represent cause? How to avoid over-determination, when obtaining knowledge about

---

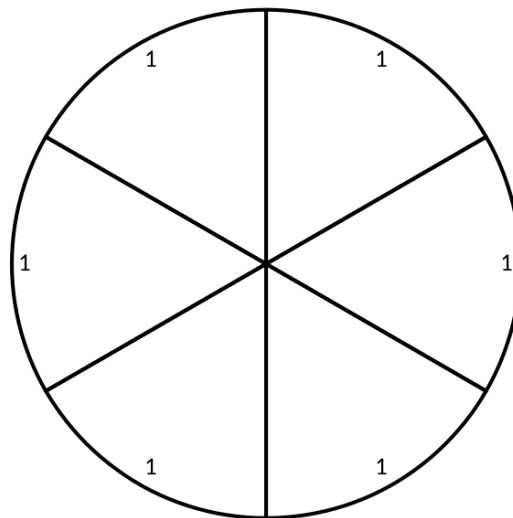
<sup>9</sup>For an excellent example of this kind of systematic review, see Kavanagh et al. (2011).

<sup>10</sup>"An INUS condition is an *I*nsufficient but *N*ecessary part of an *U*nnecessary but *S*ufficient condition for producing a contribution to the effect" (Nancy Cartwright & Hardie, 2012, p. 25; Mackie, 1965, p. 24).

structures?

I always have difficulties to understand when we know that we do not over- or underestimate the important factors that produce a contribution to the effect. Cartwright uses causal cakes (see figure 1) partly to illustrate that causes never work in isolation and partly to make the point that other causal cakes can contribute to an effect independently of the given intervention. Cartwright thereby escapes the argument of similarity and makes a wide space for practitioners to think themselves about their own practice, but Cartwright's use of the factor "other" in the cakes makes it difficult to find out if the cakes contain the *exact* relevant factors to make their contribution to the effect. When do we know that our cakes are satisfied with the right factors? As Cartwright states: "To partition too finely is as bad as not to partition finely enough. Partitioning on an irrelevancy can make a genuine cause look irrelevant, or make an irrelevant factor look like a cause" (1984, p. 36). I will try to discuss these issues further in the presentation.

Figure 1: *Example of Cartwright's causal cakes*



## References

- Bradford-Hill, A. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58:295–300.
- Cartwright, N. (1984). How the laws of physics lie. AAPT.
- Cartwright, N. (2007a). Are RCTs the gold standard? *BioSocieties*, 2(1):11–20. <https://doi.org/10.1017/s1745855207005029>
- Cartwright, N. (2007b). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge University Press.

- Cartwright, N. (2011). Predicting 'it will work for us': (way) beyond statistics. In Illari, P. M., Russo, F., and Williamson, J., editors, *Causality in the Sciences* (pp. 750–768). Oxford University Press.
- Cartwright, N. and Efstathiou, S. (2009). Hunting causes and using them: Is there no bridge from here to there? *The London School of Economics and Political Science Houghton*, 1–29.
- Cartwright, N. and Hardie, J. (2012). *Evidence-based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.
- Cartwright, N. and Munro, E. (2010). The limitations of randomized controlled trials in predicting effectiveness. *J Eval Clin Pract*, 16(2):260–266. <https://doi.org/10.1111/j.1365-2753.2010.01382.x>
- Cook, T. D., Campbell, D. T., and Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Boston.
- Deaton, A. and Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*. <https://doi.org/10.1016/j.socscimed.2017.12.005>
- Dyssegaard, C. B., Larsen, M. S., and Tiftikçi, N. (2013). Effekt og pædagogisk indsats ved inklusion af børn med særlige behov i grundskolen. *Systematisk review*. København: IUP, Aarhus Universitet.
- Egelund, N. and Qvortrup, L. (2014). Forord til bogen. In Hattie, J. and Yates, G., editors, *Synlig læring og læringens anatomi*. Frederikshavn: Dafolo.
- Glass, V. G. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, 5(10):3–8.
- Gough, D., Oliver, S., and Thomas, J. (2017). *An introduction to systematic reviews*. Sage.
- Harris, C., Hedges, L., and Valentine, J. (2009). *Handbook of research synthesis and meta-analysis*. Russell Sage Foundation, New York.
- Hattie, J. (2009). *Visible Learning – A Synthesis of Over 800 Meta-analysis Relating to Achievement*. New York: Routledge.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS medicine*, 2(8):e124.
- Kavanagh, J., Oliver, K., Oliver, S., Mara, A. O., Stansfield, C., Thomas, J., Caird, J., and Eppi-centre, J. T. (2011). Childhood obesity and educational attainment a systematic review. (1901).
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
- Kvernbekk, T. (2016). *Evidence-based Practice in Education. Functions of evidence and causal presuppositions*. New York: Routledge.



- Mackie, J. L. (1965). Causes and conditions. *American philosophical quarterly*, 2(4):245–264.
- Mansell, W. (2008). Research reveals teaching’s holy grail. Retrieved from <https://www.tes.com/news/research-reveals-teachings-holy-grail>
- Morgan, S. L. and Winship, C. (2015). *Counterfactuals and causal inference*. Cambridge University Press.
- Murnane, R. J. and Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.
- Nielsen, T. K., Tiftikci, N., and Larsen, M. S. (2013). *Virkningsfulde tiltag i dagtilbud: Et systematisk review af reviews*. Technical report, Danish Clearinghouse for Educational Research.
- Oakley, A. (2003). Research evidence, knowledge management and educational practice: early lessons from a systematic approach. *London Review of Education*, 1(1):21–33.
- Pearl, J. (2009). *Causality : models, reasoning, and inference* (2nd ed.). Cambridge: Cambridge University Press.
- Pearl, J. (2018). Challenging the hegemony of randomized controlled trials: A commentary on deaton and cartwright. *Social Science & Medicine*. <https://doi.org/https://doi.org/10.1016/j.socscimed.2018.04.024>
- Pontoppidan, M., Keilow, M., Dietrichson, J., Solheim, O. J., Opheim, V., Gustafson, S., and Andersen, S. C. (2018). Randomised controlled trials in scandinavian educational research. *Educational Research*, 60(3):311–335. <https://doi.org/10.1080/00131881.2018.1493351>
- Reiss, J. (2018). Against external validity. *CHESS Working Papers*, (2017-03).
- Vembye, M. H. and Jensen, H. S. (2018). How the laws of education lie. Retrieved from [https://pure.au.dk/portal/da/activities/how-the-laws-of-education-lie\(265be54c-88c7-4e46-894d-4c256fc14006\).html](https://pure.au.dk/portal/da/activities/how-the-laws-of-education-lie(265be54c-88c7-4e46-894d-4c256fc14006).html)

**SIMPOSIO:  
ISSUES AT THE INTERSECTION  
BETWEEN METAPHYSICS AND  
BIOLOGY**

# Introduction

## Short abstract of the symposium

Recent work in Metaphysics and in Philosophy of Science, and in particular in Philosophy of Biology, shows a revival of interest in issues that might be considered to be either metaphysical issues that can be further elucidated by recourse to biological cases or metaphysical consequences that some advancements in Biology have. In some cases, the application of some metaphysical notions to classical debates in Philosophy of Biology helps to clarify what is at stake, and to solve some misunderstandings in the discussion. The interactions that can take place between Metaphysics and Biology are therefore of different kinds. The aim of this symposium is to present four examples of such interaction and both to discuss the issue itself and, through such discussion, to explore the way in which such interaction takes place. In the symposium we will centre our attention on Philosophy of Developmental Biology, Evolutionary Biology and Evo-Devo Biology. We have chosen four classical metaphysical notions which have interacted with Biology in many ways: teleology, essence, dispositions and emergence.

## Justification of interest of the symposium

Recent work in Metaphysics and in Philosophy of Science, and in particular in Philosophy of Biology, shows a renewal of interest in issues that might be considered to be either metaphysical issues that can be further elucidated by recourse to biological cases (as illustrations or as counterexamples that help to redefine and develop further metaphysical theories), or metaphysical consequences that some advancements in Biology have. In some cases, the application of some metaphysical notions to classical debates in Philosophy of Biology helps to clarify what is at stake, and to solve some misunderstandings in the discussion. The interactions that can take place in the between Metaphysics and Biology are therefore of different kinds. The aim of this symposium is to present four examples of such interaction and both to discuss the issue itself and, through such discussion, to explore the way in which such interaction takes place.

In the symposium we will centre our attention on Philosophy of Developmental Biology, Evolutionary Biology and Evo-Devo Biology. We have chosen four classical metaphysical notions which have interacted with Biology in many ways: teleology, essence, dispositions and emergence. In the first talk, Marta Bertolaso and María Cerezo will explore the consequences that the application of some conceptual tools concerning processes and teleology have for the comparison between development and cancer. In the second talk, Christopher Austin argues that the scientific paradigm shift of evo-devo has important philosophical implications for our conception of “natural kinds” and the reality of “formal

---

causation". The interest of this talk is justified by (i) the light it provides on the limits of traditional philosophy of evolutionary biology, by enriching it with the evo-devo perspective, and (ii) the benefits for a theory of causation that such perspective has, when one adopts a neo-Aristotelian perspective to include formal causation in addition to efficient causation. In the third talk, Cristina Villegas and Laura Nuño de la Rosa intend to show how the causal and probabilistic notions implied in the research agenda of evo-devo lead to understanding EC in a broader, more inclusive way than that implied in the traditional, Modern Synthesis frame. In their view, chance is understood as a property of the manifestation process of the dispositions involved, and not as a property of the statistical distributions of a set of their effects. The interest of this talk is justified by (i) the philosophical reflection on chance in evolution and on probability that results from a serious evo-devo approach and (ii) the consequences that it has for the mechanisms/dispositions metaphysical debate (this is also present in the second talk). In the last talk, Vanessa Triviño intends to explore whether fitness is an ontological emergent property and if so, in what sense, i.e., strong or weak, and argues that it is actually strongly emergent. The interest of this talk is justified by (i) the clarification on the notion of fitness that is offered by a dispositional account of fitness assumed by Triviño, and (ii) the fact that, if the arguments are sound, we would have a case of strong emergence, whose exemplification is controversial, in the natural world.

# Form, cause, and explanation in biology: A neo-Aristotelian perspective

Christopher J. Austin  
Oxford

In the aftermath of the Scientific Revolution of the 17th century and the rise of the new mechanistic philosophy, Aristotle's concept of "formal causation" was banished from theoretical biology: its intrinsically teleological activity was explained away as nothing more than the phenomenal residue of the extrinsic forces of selection operating upon the passive participants of evolutionary processes. However, with the advent of the new science of evolutionary developmental biology (evo-devo), organism-centred evolutionary explanations—and with them, non-mechanistic models of development—are again becoming prominent. In this talk, I argue that the scientific paradigm shift of evo-devo has important philosophical implications for our conception of "natural kinds" and the reality of "formal causation".

This talk has three parts. In the first part, I provide an overview of the recent history of evolutionary biology with respect to modelling organismal development, comparing and contrasting the population genetics of the "Modern Synthesis" framework and the new systematics of evo-devo in order to illustrate their respective philosophical implications on the content of our concept of "natural kind". In my discussion of the former, I emphasise two main points. Firstly that, from the perspective of the Modern Synthesis, stability of organismal morphology is an essentially *accidental* phenomena: if groups of organisms share phenotypic features in common over developmental or evolutionary time-scales, they do so as the contingent result of past selective pressures. Secondly that, from the perspective of the Modern Synthesis, the ontogenetic construction of organismal morphology is a fundamentally *extrinsic* phenomena: the developmental specificities involved in the "building of organisms" must be ultimately traced back to the causal influences of broadly environmental factors. These two characterisations of the Modern Synthesis perspective, I will argue, flow naturally from some of its core theoretical commitments: its myopic focus on "population thinking" and its accompanying reliance on statistical analyses of gene-flow as an explanatory stratagem (Amundson 2005). In contrast to this perspective, I show how the conceptual framework of evo-devo provides a novel characterisation of the causal origins and ontological ordering of organismal morphology: on this framework, phenotypic variation is conceptualised as a by-product of the intrinsic developmental architecture of organisms, and the evolutionary trajectories of organism populations as reflecting a non-accidental exploration of the causal structure of their ontogenetic systems. With all of this in mind, I argue that the choice between the framework of the Modern Synthesis and that of evo-devo has important implications for one's conception of "natural kinds".

In the second part of the talk, I discuss in detail the causal models of development employed in the explanatory framework of evo-devo. Here I examine the higher-order representations of “dynamical systems theory”—an abstract modelling technique whose conceptual resources are uniquely capable of capturing the causal-*cum*-structural constraints of organismal ontogenesis (Wang *et al.* 2011). After showing how these models are constructed and discussing their explanatory utility, I draw three important conclusions. Firstly, in showing how their topological features represent both causality and modality, I argue that these models provide a comprehensive picture of the generative capacities of organismal systems. Secondly I argue that, in virtue of these models representing the dynamical privileging of morphological development, the picture of ontogenesis they provide is one rich with teleological structure, as they exhibit both “pleonasy” and “persistence”—the key characteristics of “directedness”. Thirdly, I argue that this structure represents the “formal” features of those systems, one which play an important explanatory role with respect to the system’s possession of the dynamical properties which are of evolutionary import in the framework of evo-devo—especially “generative robustness”, “canalisation”, and “evolvability” (Müller 2007).

In the final part of the talk, I defend this conception of “formal causation” from a few common objections by considering its relation to the explanatory models of the now popular “New Mechanism” movement (Bechtel and Abrahamsen 2005). To do so, I compare the frameworks of the so-called “ontic”, “entities and activities”-based models of “efficient causation” and the higher-order, structural models of dynamical systems theory. In examining the conceptual connection between “causation” and “explanation”, I argue that the claim that the former, mechanistic models are *exclusively* causal is mistaken: when given close attention, especially through the lens of counterfactual dependence, this claim cannot sufficiently stand up to scrutiny. I end the talk with some general, meta-metaphysical reflections on the intersection between the conceptual strategies of “formal” and “efficient” causation in modelling organismal ontogenesis, and the viability and liabilities of metaphysical theorising about ontology and explanation in evolutionary biology.

## References

- Amundson, R. (2005). *The Changing Role of the Embryo in Evolutionary Thought: Roots of Evo-Devo*. Cambridge University Press.
- Bechtel, W., & Abrahamsen, A. (2005). Explanation: A Mechanistic Alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 421–441.
- Müller, G. (2007). Evo-Devo: Extending the Evolutionary Synthesis. *Nature Reviews Genetics*, 8(12), 943–949.
- Wang, J., Zhang, K., Xu, L., & Wang, E. (2011). Quantifying the Waddington Landscape and Biological Paths for Development and Differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 108(20), 8257–8262.

# Development and cancer: Insights from teleology

**Marta Bertolaso**

Rome

**María Cerezo**

Murcia

Classical work on philosophy of cancer has described it as a failure to follow the ‘correct’ developmental trajectory (Bertolaso and Dupré 2018, Bertolaso 2016) or as development gone awry (Soto et al. 2007). However, (some of) the processes of cancer have also been assimilated in some sense to those of development since in cancer the fate of cells seems to depend on the context on which they are, as it happens in development, so that it can be affirmed that “elements of pathological tumor plasticity mirror the developmental history of organs (...) cancer cells can acquire fates associated with neighboring organs” (Tata et al. 2018). In this contribution, we aim at solving this apparent tension by showing the sense in which cancer can be described as the pathology of development.

Development and cancer are here seen as processes: the developmental trajectory extended in time—the dynamical sequence of states that constitute the organism ontogeny—and the complex disease extended in time—the neoplastic pathological process. Recent work on processes in philosophy has focused on the teleological dimension of processes - their tendency towards an end-state or their directedness (Birch 2012, McShea 2012). This tendency must not be understood as a “directing force” or the fact that the process is directed by something, but rather as the fact that the process has a direction, that it in fact ends at a particular final state in normal conditions. Since development and cancer are processes, the question whether they are teleological in such a sense arises.

Our hypothesis is that recourse to such use of teleological discourse can shed light on the tension above indicated, and our purpose in our contribution is to examine whether this is the case. We can anticipate that it does shed light, and that there is a sense in which development can be said to be teleological while cancer cannot, even if the latter is described in processual terms. But the elucidation of this sense in which development and cancer are different due to their teleology requires an analysis of the processes we are talking about and of their subprocesses, and an elucidation of the uses of teleology we have in mind.

We will describe development as the process by which organisms (typically multicellular organisms) grow and transform into adult forms, and will focus on three particular cell subprocesses that are crucial in development: proliferation, migration and differentiation. Cancer in turn can be described as an unrestricted or unregulated cell proliferation, an abnormal cell growth, which has the potential to invade other parts and organs of the body. In cancer, three cellular subprocesses correspond to the three ones in development: proliferation,

invasion and heterogeneization.

Two kinds of teleological discourse can be applied to development and cancer as processes: functional discourse (what contribution a process makes to a further process or system) and end-state discourse (tendency towards a particular end-state). Different uses and degrees of these two kinds of teleological discourse can be acknowledged and applied to development and cancer and their subprocesses, but one of them is particularly useful to distinguish both processes. A process is *end-state teleological* if there is a parameter X that is being maximized/optimized through the process (see Birch 2012). In our contribution we apply this and other notions of end-state teleology to the development and cancer and their subprocesses.

Section 1 is an introduction of development and cancer and their subprocesses. In Section 2, we offer an analysis of the concept of teleology when applied to processes. In Section 3 we show that development is a complex process and that it is teleological in the sense that lineages of cells evolve to maximize (optimize) integration of subprocesses. In Section 4, we address the question whether teleological discourse in the senses relevant in this paper apply to cancer's subprocesses and to cancer itself. We intend to show that it is here where the difference (or at least one of them) between cancer and development lies. Section 5 considers some objections to our analysis. We close with some concluding remarks.

In our contribution we are neutral with respect to two issues. Firstly, whatever we are trying to capture in this work by appealing to teleological discourse is natural. Such naturalness can be considered to be ontological if there is a property of the relevant processes that is referred to by the expression "teleological", or epistemological if there is not such a property but there is a use of the teleological discourse that is epistemologically indispensable. We remain neutral to this question. Secondly, we assume a processual perspective on development and cancer, but there are other perspectives in which they are considered as substances evolving: as an organism developing or as a tumor growing and changing. In our perspective, we are not concerned with the persistence, identity or sameness of what evolves, but with the one of the process as process. This perspective is useful to understand the sense in which cancer is the pathology of development, but there might be purposes for which the substance perspective is more useful.

## References

- Bertolaso, M. 2016. *Philosophy of Cancer – A Dynamic and Relational View*. Springer Series in History, Philosophy & Theory of the Life Sciences.
- Bertolaso, M. and Dupré, J. 2018. A Processual Perspective on Cancer, in Nicholson, D. and Dupré, J. *Everything flows*, OUP, 321–336.
- Birch, J. 2012. Robust processes and teleological language. *European Journal of Philosophy of Science*, 2, 299–312.
- McShea, D. W. 2012. Upper-directed Systems: A New Approach to Teleology in Biology. *Biology and Philosophy*, 27, 663–684.



- Soto, A. M., Maffini, M. V., and Sonnenschein, C. 2007. Neoplasia as Development Gone Awry: The Role of Endocrine Disruptors. *International Journal of Andrology*, 30, 1–5.
- Tata et al. 2018. Developmental History provides a Road Map for the Emergence of Tumor Plasticity. *Developmental Cell*, 44, 679–693.

# Emergentism in a biological framework: The case of fitness

Vanesa Triviño  
Madrid

Emergentism was conceptually developed as a middle position between mechanicism and vitalism (during the late 19th and the early 20th century). Despite of disappearing of the mainstream philosophy between the 1930s and 1970s (Wilson 2013), since the last decades of the 20th century, emergentism has reappeared in scientific and philosophical contexts. Currently, debates are oriented to clarify the basic idea of emergentism: dependence and autonomy (Clayton 2006; Van Gulick 2001).

Recently, Wilson has made a classification of the different forms of conceiving dependence and autonomy. For the author, ontological emergent properties can be characterized as being weak or strong depending on how dependence and autonomy is given. Weak ontological emergent properties have a non-empty proper subset of the token powers had by the base upon which they synchronically depend. Strong ontological emergent properties, are those that instantiate a new token causal power that is different from the token causal powers possessed by the base upon which they synchronically depend (Wilson 2016).

In this presentation, I apply Wilson's schema to the biological property of fitness understood as a causal dispositional property (Triviño and Nuño de la Rosa 2016). This exploration is given within the broader framework of the interactions between metaphysics and biology. In particular, it is an example of the so-called 'Metaphysics *for* Biology', wherein a metaphysical theory or concept is used to clarify the reality to which biological concepts refer. Here, the metaphysical concept is that of emergence, and the biological concept whose reality I want to clarify is fitness.

The general aim of the presentation is to explore whether fitness is an ontological emergent property and if so, in what sense, i.e., strong or weak. To develop this task, I pay attention to the different forms in which the relation of dependence can be given between an emergent property and its basis. In this respect, I argue that (1) the relation between fitness and its basis is not of fusion: the functional dispositions that characterize the organism do not fuse in order to give rise to fitness. (2) There is not a relation of modal covariation: changes in the functional dispositions upon which fitness depends do not imply changes in fitness. (3) There is not a non-reductive realization: fitness cannot possess a mere 'subset' of the causal powers that characterize its basis insofar as its causal power is not given at the base level. Finally (4), the relation between fitness and its basis is a causal one: the functional dispositions that characterize the organism are the combined causes that give rise to an effect they cannot produce in isolation: fitness.

Causal dependence is related to strong emergentism. The effects are different from their causes insofar as they have *powers* different from those having by the base (Wilson 2016, 368). It seems that if relation of dependence between fitness and its basis is a causal one then, fitness can be characterized as a strong emergent property. However, it also needs to be argued that the causal power of fitness is different from that of its basis, such that fitness is a causally autonomous property and not a mere epiphenomenon.

I offer three arguments for the causal autonomy of fitness: non-overdetermination, non-identity and downward causation. *Non-overdetermination* shows that the causal power of fitness is different from the causal power of the combination of functional dispositions that gives rise to it. Both fitness and the combination of functional dispositions are, in fact, elements of two different, although overlapping, causal processes (Mumford and Anjum 2011, 125–126). In one causal process the combination of functional dispositions is the cause and fitness is the effect; in the other fitness is the cause and the survival and reproduction of the organism is the effect. *Non-identity* illustrates that the causal power of fitness is not played by the combination of functional dispositions of the organism since fit organisms interact with the environment and population in a form that the lower-level functional dispositions that characterize it are not able to do. Finally, *Downward causation* appeals to cases of *phenotypic reorganization* and *implantation delay* to argue that it is due to their fitness that organisms can downwardly affect the functional dispositions that characterize them in order to guarantee their survival and reproduction in a given environment and population.

Since fitness causally depends on its basis and is causally autonomous, I conclude that it can be characterized as a strong ontological emergent property.

## References

- Clayton, P. (2006). Conceptual Foundations of Emergence Theory. In: *The Re-Emergence of Emergence*. Oxford: Oxford University Press.
- Mumford, S. and Anjum, R.L. (2011). *Getting Causes from Powers*. Oxford: Oxford University Press.
- Triviño, V. and Nuño de la Rosa, L. (2016). A dispositional account of fitness. *History and Philosophy of the Life Sciences*, 1–18.
- Van Gulick, R. (2001). Reduction, Emergence and Other Recent Options on the Mind/Body Problem: A Philosophic Overview. *Synthese*, 8:1–34.
- Wilson, J. (2013). Nonlinearity and Metaphysical Emergence. In: S. Mumford and M. Tugby (eds.), *Metaphysics and Science*. Oxford: Oxford University Press, pp. 201–235.
- Wilson, J. (2016). Metaphysical emergence: weak and strong. In: T. Bigaj and C. Wüthrich (eds.), *Metaphysics in Contemporary Physics*. Boston: Brill Rodopy, pp. 345–402.

# Chances and propensities in evo-devo

**Cristina Villegas**

Madrid

**Laura Nuño de la Rosa**

Madrid

The debate on the role of chance in evolution has experienced an intense revival coming from the theoretical and empirical advances undergone by evolutionary biology in the last few decades and the integration of their new research agendas into evolutionary theory. In this work, we intend to clarify some aspects of this debate with respect to variational evolutionary chance (VC) through an analysis of its metaphysical implications regarding the nature of causality and probability in evo-devo. When discussing evolutionary chance at the variational level, namely whether biological variation is random or not with respect to evolution, a focus on the statistical correlations between mutational rates and their fitness effects has prevailed (see Merlin 2010). This focus neglects the causal generation of variation through development, and deals with a very restricted notion of chance with little explanatory power. We intend to show how the causal and probabilistic notions implied in the research agenda of evo-devo lead to understanding VC in a broader, more inclusive way than that implied in the traditional, Modern Synthesis frame. Using the conceptual tools from the metaphysics of causality and the philosophy of probability, we argue that, since the generation of phenotypic variation through development is not random, the received view of VC needs to be reconsidered.

While the focus on variation endorsed by the statistical approach of evolutionary genetics in the Modern Synthesis justifies a correlation-based notion of VC applied to specific inputs and outcomes (namely, mutations and their corresponding phenotypic effects), the switch to the causal functions of evo-devo, in relating initial conditions to phenotypic outcomes, allows us to think of chance as a property of the causal processes involved in the generation of variation. The properties of developmental systems can be understood as the setup conditions of a probabilistic system (Abrams 2017), namely those that enable the occurrence of particular causal interactions between specific initial conditions and outcomes, modelling a set of possibilities (in this case, a space of possible phenotypic forms, or morphospace). By contrast, mutations play the role of triggering conditions, namely those factors that precipitate specific effects among those possibilities (the extant forms filling particular regions of the morphospace).

In evo-devo, setup conditions are usually described in dispositional terms, while the effects of these causes are seen as their manifestation (Austin and Nuño de la Rosa 2018). In the philosophy of causation literature, dispositional properties are those that allow their bearers to behave in a specific way when certain conditions are given. In our case, the “variability of a phenotypic trait describes the

way it changes in response to environmental and genetic influences” (Wagner and Altenberg 1996), and is manifested in particular distributions or frequencies of phenotypic variants. This contrast between “the realised differences among organisms” and “their underlying variational tendencies” (Wagner 2014) is patent in the way chance is invoked by population and developmental approaches to evolution. Population genetics models make use of frequencies and statistical correlations in order to account for how extant variation behaves probabilistically in a population, whereas evo-devo models derive their probabilistic predictions out of the causal structure of developmental systems. In other words, evo-devo is concerned with the *probability of generation* of phenotypic variants, instead of the *probability of fixation* of variants. In this regard, we argue that developmental dispositions are better understood as “propensities”, probabilistic or chancy dispositions—that is, as dispositions that, in contrast to deterministic ones, manifest themselves in the probability distribution of their possible effects. Some authors regard this shift from statistical correlations to causal (molecular and developmental) mechanisms as a consequence of the empirical developments experienced by evolutionary biology in the last few decades. Whereas the identification of specific mechanisms is an important factor in the understanding of the causal generation of variation, we believe that the characterization of such generation with respect to chance is better understood in dispositional terms. This is because the appeal to propensities has no further metaphysical commitments regarding the ontology (mechanistic or not) underlying the dispositionality of developmental properties. Instead, it commits to the fact that such dispositionality is causally responsible for the probabilistic behaviour of developmental systems. We therefore think this shift can be more precisely thought of as a shift from statistical correlations to propensities, insofar as evo-devo’s explanations of developmental properties such as robustness, plasticity or modularity, require not only to depict the actual behaviour of a mechanism, but also its dispositions, that is, the capacities of developmental systems to either react to perturbations or permit modifications.

The major assumption of the traditional, externalist approach to evolution is that the fitness probabilities associated to mutations exclusively depend on environmental contingencies. In contrast, evo-devo focuses on the connection between the propensities of developmental systems and chances, grounding the probability distributions of phenotypic variants on the internal properties of organisms. Understanding developmental properties as propensities entails a switch from an effect-centered view of chance, defined in terms of how the effects are distributed, to a process-centered one, where chance is defined in terms of how those effects are produced. In this sense, chance is understood as a property of the manifestation process of the dispositions involved, and not as a property of the statistical distributions of a set of their effects. This way, in order to judge whether variation is chancy or not, what the explanatory structure of evo-devo demands is a consideration of how the variational propensities of developmental systems behave in their manifestation process. We defend that this metaphysical distinction leads to a better understanding of some of the core aspects of the debate on VC.

## References

- Abrams, M. (2017). Probability and chance in mechanisms. *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, 169–184.
- Austin C. J., Nuño de la Rosa L. (2017). Dispositional concepts in evo-devo. In: *Evolutionary developmental biology: a reference guide*.
- Merlin F. (2010). Evolutionary Chance Mutation: A Defense of the Modern Synthesis' Consensus View. *Philos Theor Biol*, 2, 1–22.
- Wagner G. P. (2014). *Homology, Genes, and Evolutionary Innovation*. Princeton University Press.
- Wagner, G. P., Altenberg, L. (1996). Perspective: complex adaptations and the evolution of evolvability. *Evolution*, 50(3), 967–976.

SIMPOSIO:  
LA FILOSOFÍA DE LA  
ARGUMENTACIÓN

# Maniobras estratégicas y estrategias interpretativas en la argumentación

**Jesús Alcolea Banegas**  
Universitat de València  
alcolea@uv.es

## Introducción

Frans van Eemeren ha ampliado su teoría pragmatialéctica de la argumentación hasta incorporar elementos retóricos bajo la idea de “maniobra estratégica”, y un conjunto más rico de instrumentos conceptuales. Ahora bien, al analizar los diferentes movimientos que los argumentadores realizan en el proceso argumentativo para alcanzar sus objetivos se recurre a determinadas estrategias interpretativas que no se reducen a simples declarativos de uso, como ha indicado Van Eemeren, quien también ha explicado que su uso del término interpretación remite a cualquier asignación de sentido a cualquier parte del discurso –o artefacto cultural de otro tipo, dice él–, pero también a los distintos movimientos que en él se realizan. Lo que proponemos es que estos instrumentos son, en parte, estrategias interpretativas de las características relevantes del contexto (verbal y no verbal, y “poli-interpretable”), que son las que facilitan una “interpretación máximamente argumentativa” de las intervenciones de los agentes que podría concordar con sus intenciones.

## Maniobras estratégicas

Con la introducción de las maniobras estratégicas, se ha buscado combinar lo eficaz y lo razonable de forma equilibrada. Estas maniobras son estrategias retóricas en las que los hablantes utilizan palabras y otros recursos adaptándolos a sus necesidades para promover y defender sus posiciones en los intercambios que tienen lugar en un contexto colaborativo, mientras esperan que sus contrapartidos comprendan sus pretensiones y respondan en consecuencia, al tiempo que despliegan sus propias sutilezas y su creatividad (cfr. Alcolea, 2017).

Estas maniobras estratégicas siempre aparecen en una actividad comunicativa particular y a ellas se asocian tres aspectos relacionados con las elecciones que se realizan en el proceso: (1<sup>o</sup>) La elección del “potencial tópico” remite a la selección por parte del argumentador de los temas en diferentes movimientos y etapas. Por ejemplo, elegir alguno de los argumentos disponibles que mejor permiten defender su posición. (2<sup>o</sup>) Los movimientos se han de adaptar a las “demandas del auditorio” buscando su agrado. De modo más general, los movimientos deben



armonizar con las creencias, los valores y las preferencias del auditorio. (3º) El “dispositivo de presentación” conlleva la elección del mejor modo de presentación (estilo) de los movimientos a fin de rentabilizarlos estratégicamente y de que “parezcan más aceptables” al presentarlos de forma razonable y eficaz. En el análisis y evaluación del discurso argumentativo hay que identificar estos tres aspectos y sus interrelaciones en cada movimiento, y cómo se configuran en cada maniobra estratégica (van Eemeren 2012:158–159).

Lo ideal sería que todos los movimientos mantuvieran el referido equilibrio entre lo razonable y lo eficaz. Pero ello no siempre es el caso, pues hay argumentadores que en sus estrategias sobrecargan los aspectos razonables en detrimento de los eficaces, mientras que otros actúan a la inversa. En el primer caso, podemos encontrarnos con estrategias no falaces, pero que no siempre resultan beneficiosas para el argumentador. En el segundo, las estrategias parecen conducir irremediabilmente, según Van Eemeren, a una falacia.

Para evitar estos efectos perniciosos, el agente se preocupará de ser interpretado correctamente y, así, debe buscar la colaboración activa y máxima del coagente dejándole suficiente libertad para interpretar sus propuestas y sus movimientos de la forma más acertada posible. Es decir, debe cuidar sus actos ilocucionarios para dejar claro que sus pretensiones son sinceras y que hace lo que hace o dice lo que dice porque lo considera adecuado. Solo así conseguirá agradar a la otra parte y que ésta comprenda y acepte sus propuestas (el efecto perlocucionario) de forma activa y compartida. Es decir, la otra parte será *persuadida* de forma razonable (cfr. Van Eemeren, 2017:43), porque ha respondido de forma razonable a argumentos *persuasivos*.

Ahora bien, en el proceso los argumentadores se ven obligados a ofrecer explicaciones o interpretaciones, que, como declarativos de uso, pueden ocurrir en cualquier momento, pues la comprensión puede requerir de mayores precisiones. Llegar a la correcta no siempre es fácil. La consulta al contexto, con las observaciones que nos facilita, es imprescindible para facilitar una interpretación pertinente y concluir, por ejemplo, si estamos o no ante un argumento. Para ello conviene tener en cuenta el propósito del intercambio y las obligaciones pragmáticas de los agentes que presentaron aquel argumento y que estarán constreñidos por la búsqueda de una interpretación máximamente argumentativa (Van Eemeren & Grootendorst, 2013:248–249).

## Necesidad de la interpretación

A veces, cuando se nos dice algo podemos proceder a buscar su significado y podemos formular determinadas hipótesis interpretativas sirviéndonos de determinados recursos o estrategias para generarlo. En principio, el recurso más socorrido podría ser un “todo vale” (Dascal, 2003:641), de manera que en la interpretación se podrían emplear nuestros temores, anhelos, creencias... para armonizar el significado con lo ya conocido.

Por Eco (1995) sabemos que un texto o una proferencia se abren a un abanico de posibles interpretaciones, pues todo significado surge de la actividad interpretativa. Un agente no crea plenamente el significado, que depende de las

intenciones de los agentes a quienes trata de interpretar, y a quienes no puede ignorar, pues busca una base segura de entendimiento con ellos y avanzar en el intercambio argumentativo, de modo que algunas interpretaciones podrían no ser aceptadas como válidas.

Por otro lado, más allá del significado literal de una preferencia, debemos reparar en numerosos elementos contextuales para transmitir el significado. A veces usamos expresiones de forma no literal, lo que no conlleva que el intérprete sea libre de insuflar en una expresión la interpretación que más le apetezca. El margen de libertad en la interpretación no debe estar reñido con la *fidelidad* a lo que estamos interpretando (de acuerdo con Eco, según Dascal, 2003:646).

También cabe señalar que en una interacción el trasfondo no articulado es tan relevante para la comprensión como lo articulado, y es importante captar la disposición anímica, el talante del coagente y su estilo de comunicación para comprender sus intervenciones. Al expresar e interpretar un mensaje, vamos más allá de la información procurada por las convenciones sintácticas y semánticas. Como señala Jacobs (2016:2), la gente “va más allá de la información dada al construir un contexto de supuestos e inferencias que da sentido a lo que se dijo y a lo que *no* se dijo pero *podría haber* sido dicho, y que da sentido a *cómo* y *cuándo* todo *fue* dicho”. Las preferencias no son el mensaje con el que se transmiten los argumentos, sino que sirven como “señales” para construir el mensaje, que integra el conjunto de supuestos e inferencias y que tiene sentido porque “satisface criterios de acción racional”. Esto conlleva también que la conducta del agente sea interpretada maximizando esa racionalidad y que el coagente proceda seleccionando a través de determinadas estrategias la interpretación que sea más relevante a los objetivos manifiestos del agente. Comprender la semántica de sus palabras o sus gestos es necesario, pero no suficiente, pues la comprensión de los actos de habla indirectos, las implicaturas, su retórica, exige información de la situación concreta en que la comunicación tiene lugar.

Sin embargo, toda interpretación, incluso por razonable que parezca, tiene carácter provisional y debe quedar abierta a la presentación de otras posibles interpretaciones, buscando el equilibrio. Recurrir al contexto es necesario para facilitar la interpretación, como lo es tener en cuenta que podemos estar ante diferentes recursos multimodales para comunicarnos y dotar de sentido y significado nuestra comunicación. Como Jewitt (2014:1) resume acertadamente, “la multimodalidad se centra en la representación, comunicación e interacción como si fueran algo más que lenguaje”. Al vivir en una cultura multimodal, nos comunicamos con diversos modos semióticos. Así, para dotar de sentido a un texto debemos tener en cuenta otros elementos que los estrictamente lingüísticos, ir más allá de las tradicionales estrategias cognitivas y reparar en las estrategias interpretativas que nos permiten construir el significado. La multimodalidad se encargaría de analizar las reglas, las estrategias, etc. que permiten que los agentes comprendan el significado potencial de los elementos convergentes en el contexto. Todo modo semiótico procura un significado potencial y la multimodalidad las interrelaciones entre los diversos modos, su análisis y su evaluación.

Por otro lado, para comprender la interacción que se produce entre los agentes es preciso estudiar cómo los coagentes interpretan y responden a la interacción. Sus estrategias interpretativas serán tanto más importantes cuanto el contexto se presenta de una manera multimodal y, así, entimemática. Estas estrategias nos

ofrecerán el mejor significado del intercambio y la comprensión intersubjetiva que ha de primar en una comunidad de intérpretes. La máxima interpretación argumentativa se alcanza así yendo más allá de los simples “argumentos” del agente hasta el contexto, para, a través de las estrategias interpretativas, volver a la máxima interpretación logrando el objetivo de la persuasión argumentativa razonable.

## Estrategias interpretativas

Recordemos que el desacuerdo solo hace acto de presencia entre agentes que defienden diferentes posiciones. A partir de su conocimiento, del conocimiento que va obteniendo de las intervenciones de la otra parte y del uso de las estrategias a que pueda recurrir, un agente puede leer e interpretar de forma activa el contexto del intercambio.

Quien interpreta debe enfocar multimodalmente la situación, su forma de apreciarla y percibirla, para detectar obstáculos y elementos facilitadores. Entre esos obstáculos puede estar la voluntad del agente proponente de impedir que lleguemos a cumplir cabalmente el objetivo. Pero si aplicamos el principio de caridad, debemos rechazar tal pretensión, evitando las discrepancias y mostrando una disposición crítica para alcanzar una interpretación adecuada. La interpretación no tiene un objetivo difuso y desarticulado, sin falta de estrategias. Al contrario, una estrategia es el conjunto de acciones complementarias que se realizan para conseguir un determinado objetivo. Con una estrategia interpretativa se destacan los elementos que nos permiten alcanzar una interpretación adecuada. Las estrategias no son ni buenas ni malas sino adecuadas o inadecuadas. Entre las primeras encontramos:

(1<sup>a</sup>) *La protección de las intenciones*, según la cual podemos interpretar la interacción del argumentador para reforzar sus argumentos, pero no tanto para realzar su poder justificatorio, como para incrementar su poder persuasivo. Para ello, confiamos en la sinceridad del agente y en la selección de aquellos aspectos del contexto que se consideran más significativos, recurriendo al recuerdo o a la consideración de situaciones análogas. La interpretación debe orientarse a las intenciones del agente, a la comprensión general y a la utilidad del intercambio. Este solo puede resolverse si los participantes expresan sus intenciones e interpretan las intenciones de los demás con la mayor fidelidad, de modo que se minimicen los malentendidos. Cuando sea necesario, deben estar preparados para reemplazar sus formulaciones e interpretaciones por otras mejores, pues es efectivamente imposible, en principio y en la práctica, que una única interpretación pueda ser correcta u óptima en todos los aspectos.

(2<sup>a</sup>) *La corrección de las consideraciones en contra (contra-consideraciones)*. Se trata sobre todo de consideraciones pragmáticas de carácter marcadamente crítico y que el agente proponente no puede ignorar. El caso más simple podría ser una contra-consideración explicativa, en la que cabría exigir una explicación de un aspecto conflictivo del intercambio. O también las objeciones que conducen a (pedidos de) explicaciones. Pero las contra-consideraciones pueden plantearse a un nivel más dramático. La variedad de consideraciones no triviales que se

ponen de relieve en el intercambio y con las que debe llegar a un acuerdo son de tal naturaleza que a veces surgen tensiones internas y hasta inconsistencias. El desacuerdo puede superar el ámbito de lo que se puede mantener a la luz de consideraciones de consistencia y acabar en una situación complicada. Por supuesto que los agentes desean eliminar la inconsistencia en que se han enredado, pues compromete todo el intercambio argumentativo. Por ello, el coagente debe invitar al agente proponente a restaurar la consistencia entre sus compromisos incompatibles abandonando, por ejemplo, algunas de las creencias que engendran la dificultad.

(3<sup>a</sup>) *Lectura crítica del significado pretendido.* Se trata de la construcción multimodal del significado objetivo, leyendo críticamente el significado pretendido por el agente proponente. El intérprete debe tener en cuenta que su construcción del significado a partir de diferentes elementos contextuales, que actúan como genuinos recursos semióticos, depende de dos aspectos: del potencial significante y argumentativo de esos recursos, y del significado que el agente proponente ha podido proyectar sobre ellos con sus intervenciones, que obviamente traslucen su intencionalidad argumentativa. Es obvio que los diferentes recursos multimodales activan, en principio, respuestas diferentes para el agente argumentador y para la otra parte. Sin embargo, en la medida en que son complementarios, contribuyen de forma conjunta a facilitar un significado general y último que es algo más que la suma de los significados transmitidos por los recursos por separado. También es evidente que esta estrategia interpretativa pone a la otra parte en el camino de una respuesta crítica a las pretensiones del agente proponente, pues al ir más allá del debido análisis interpretativo se interna en el terreno de la evaluación en busca de una correcta interpretación de sus preferencias en base a la pretendida interpretación máximamente argumentativa que, sin duda, aguarda a un nivel más profundo.

(4<sup>a</sup>) *Percepción del sentido situacional del significado.* Es una estrategia relacionada con lo que el coagente ha podido ver y oír, y que procede no solo del agente, sino también del contexto, que a su vez dota de significado a los objetos a veces de forma cambiante. La percepción y la observación del entorno facilitan una serie de elementos particulares y directos que inician el proceso de comprensión del significado situacional, facilitando la interpretación visual y auditiva de la otra parte. En el intercambio, pueden entrar en juego nuevos elementos interpretativos que presuponen cambios perceptibles no solo en relación con los procesos mentales o emocionales, sino también nuevas ideas o sentimientos que pueden influir en la interpretación buscada. Cabe observar que en el proceso las palabras, las imágenes, los sonidos, etc., pueden sufrir alteraciones hasta el punto de recibir unos significados advenidos por cambios en asociaciones temporales nacidas del uso. La percepción del significado objetivo supone la captación de ciertos aspectos implícitos cuya relevancia para la interpretación deberá confirmarse. Estos significados intrínsecos dependen de la perspectiva que el agente asume sobre el tema y de los recursos que se utilizan en el intercambio. La percepción puede llevarnos a sutiles y válidas relaciones entre elementos del entorno que pueden o no haber estado en conexión en otras ocasiones. Así, según sean esas relaciones podríamos tener una posible interpretación u otra.

(5<sup>a</sup>) *El análisis de las interrelaciones.* Una vez se ha atendido a los diferentes recursos multimodales del contexto, conviene analizar las interrelaciones, pues

pueden desempeñar un papel clave en la interpretación. Es evidente que la otra parte se hará preguntas sobre lo que ve y oye y lo que cree que puede significar. Es lo que podríamos llamar *efecto ‘Ventana indiscreta’*. Así, si hablamos en general del contexto, puede preguntarse sobre cómo se disponen e interactúan los objetos en él. Lo mismo cabe decir de las imágenes o de los sonidos u otros recursos semióticos percibidos, y también de sus relaciones con las preferencias del agente. Cabe pensar que esta estrategia depende de cómo se consigue, por ejemplo, que determinados elementos visuales funcionen como ilustraciones de un tema y permiten a los coagentes la oportunidad de evaluar la adecuación de interpretaciones ya formuladas a las imágenes. En este caso, el esfuerzo interpretativo puede ser considerable, y el coagente verse implicado en un tipo de lectura visual “cultivada”, pues el análisis cuidadoso de los detalles puede llevarnos a una interpretación compleja y sutil.

(6<sup>a</sup>) *El análisis de las relaciones externas* nos permite identificar los significados de los objetos y de sus relaciones en ese contexto en función de sus relaciones con un contexto más amplio en el que podría enmarcarse el intercambio argumentativo, es decir, un contexto científico, filosófico, legal, etc., o incluso cultural todavía más amplio. Algunos objetos y algunos aspectos de los mismos pueden recordarnos esas conexiones, que también contribuirán a una interpretación adecuada. Quizás de forma más interesante la interpretación lograda siguiendo esta estrategia nos permite excluir alternativas e incluso optimizar su validez externa hasta el punto de sacar en claro las motivaciones y las razones de los participantes para entrar en el intercambio.

## Conclusión

La producción y la evaluación del discurso argumentativo son muy importantes, pero también lo es la interpretación como elemento del componente empírico de la pragmadialéctica. Dar la respuesta adecuada al sujeto argumentador no es tarea fácil y requiere realizar determinadas elecciones sobre los significados a adscribir a lo que él ha proferido, sobre todo porque algunos aspectos quedan ocultos en el nivel entimemático y contextual. Sacarlos a la luz puede contribuir a una interpretación más adecuada. El análisis cuidadoso de los detalles puede llevarnos a una interpretación difícil, compleja y sutil. Para ello, nuestra misión ha consistido en referenciar determinadas estrategias que sin duda pueden ayudar a alcanzar una interpretación que enriquecerá las propuestas del sujeto, pero que, por lo demás, nunca llegará a ser definitiva, pues se mueve en la dialéctica entre las estrategias y las competencias de los agentes argumentadores.

## Referencias bibliográficas

- ALCOLEA, J. (2017): “Ser agente argumentador creativo”. *Revista Iberoamericana de Argumentación*, 15, pp. 1–15.
- DASCAL, M. (2003): *Interpretation and Understanding*. Amsterdam: Benjamins.

- JACOBS, S. (2016): “Employing and exploiting the presumptions of communication in argumentation: An application of normative pragmatics”. In *Argumentation, Objectivity, and Bias: Proceedings of the 11th International Conference of OSSA*. Edited by L. Benacquista and P. Bondy. Windsor, ON: OSSA, pp. 1–24.
- JEWITT, C. (Ed.) (2014): *The Routledge Handbook of Multimodal Analysis*. London: Routledge.
- VAN EEMEREN, F.H. & GROOTENDORST, R. (2013): *Los actos de habla en las discusiones argumentativas*. Santiago de Chile: UDP.
- VAN EEMEREN, F.H. (2012): *Maniobras estratégicas en el discurso argumentativo*. Madrid: CSIC.
- VAN EEMEREN, F.H. (2017): “Fallacies as derailments of argumentative discourse”. In *Contextualizing Pragma-Dialectics*. Edited by F.H. van Eemeren and W. Peng. Amsterdam: Benjamins, pp. 37–58.

# Las preguntas “por qué” argumentativamente

**Joaquín Galindo Castañeda**

Universidad de Valladolid

kino\_gac@hotmail.com

Se ha sostenido que una función característica de determinados tipos de preguntas es pedir razones, por lo que, plausiblemente, en ocasiones pueden cumplir el papel de marcadores argumentativos (van Eemeren, F.H., Houtlosser, P. y A. F. Snoeck Henkemans: 2007). No obstante, con frecuencia, como sucede en el caso de las preguntas de tipo “por qué”, las razones solicitadas pueden ser explicaciones causales y no argumentos. Entonces cabe preguntar: ¿cuándo una pregunta pide razones, en el sentido argumentativo, y no explicaciones, clarificaciones u otro tipo de información? En este trabajo busco articular las condiciones que definen a las preguntas argumentativamente orientadas (PAO), esto es, preguntas que piden razones o evaluación de razones en el sentido argumentativo. Haré dos restricciones en mi objeto de estudio: por una parte, me centraré en un tipo de preguntas: las preguntas “por qué”; y, por otra parte, en una clase de argumentación: la argumentación filosófica. Estas dos restricciones obedecen a la siguiente estrategia heurística: deseo encontrar un claro contraste entre las PAO y preguntas de la misma forma, pero con función característicamente distinta. Las preguntas “por qué” han sido estudiadas en detalle por la lógica erotética, la lingüística formal y la filosofía de la ciencia. Su estudio ha sido motivado principalmente por su relación con las explicaciones causales. Ahora bien, hay un claro consenso (uno de los pocos) sobre que la filosofía, en contraste con la ciencia, no da explicaciones causales. Otro consenso es que la diferencia entre las explicaciones causales en el discurso cotidiano y las explicaciones científicas sólo es de grado, pero no de tipo (aunque hay toda clase de desacuerdo sobre los detalles). Así, parece promisorio en principio investigar el papel de las preguntas “por qué” en la argumentación filosófica, pues en ella es escasa o nula la expectativa de una explicación causal.

Dividiré mi exposición en tres partes. La primera parte tiene el propósito de fijar la terminología y los conceptos básicos que utilizo en mi argumento principal. La terminología y los conceptos se refieren principalmente a la estructura de las preguntas “por qué” y a las escurridizas nociones de “razón” y “de dar razón”. Comienzo con una breve caracterización lingüística de las preguntas “por qué” y sus expresiones normales. Después procedo a analizar los modelos de explicación causal basados en preguntas o erotéticos de Bromberger (1992), van Fraassen (1987) y con más detalle de Bradford Skow (2016). Explico las distinciones canónicas entre “tema  $P_K$ ”, “clase contraste  $X=\{P_1, \dots, P_K, \dots\}$ ” y “relación de relevancia  $R$ ” y las preguntas “por qué” entendidas como el triplete  $Q=\langle P_K, X, R \rangle$ . Después me detengo en una noción que será crucial en mi argumentación: “la presuposición de una pregunta”. Se trata de un concepto clave

para dar cuenta del hecho de que una pregunta surja o se suscite en un contexto determinado (concepto modelado por primera vez en la lógica erotética de Belnap de 1972). En este caso, las condiciones para que surja una pregunta “por qué” con orientación explicativa son: 1) su tema es verdadero; 2) en su clase contraste, solamente su tema es verdadero; 3) por lo menos una de las proposiciones que está en relación de relevancia con su tema y su clase de contraste es también verdadera (van Fraassen: 1987). Por último, aludo a las teorías de trasfondo y al carácter pragmático de las preguntas “por qué” en su orientación explicativa causal.

El contraste que deseo trazar entre PAO y preguntas “por qué” explicativas no se entenderá si no esclarezco primero lo que quiero decir por “razón” y “dar y pedir razones”. Repaso algunas definiciones de “razón” en la tradición de la filosofía analítica: Parfit (1984), Bratman (1999), Alvarez (2010), Brandom (2001). Y por otra parte, algunas definiciones de “argumento” y “argumentación” provenientes del campo de Teoría de la Argumentación: Johnson (2000), Govier (2005), van Eemeren & R. Grootendorst (1984), Vega(2003), Marraud (2013). No se trata de elegir una de entre este vasto catálogo, sino de señalar el fenómeno al cual apuntan. Por último, introduzco la distinción de John Broome (2013) entre razones *pro toto* y razones *pro tanto*, que me parece es esclarecedora, como se verá en la tercera parte.

En la segunda parte, presento y explico la tesis que sostendré. Antes, en la primera parte, se ha enfatizado que el análisis de las preguntas “por qué” se centra en su orientación explicativa y no argumentativa. En teoría de la argumentación, por contraste, se ha enfatizado, no exactamente una diferencia entre preguntas PAO y preguntas explicativas, pero sí, un claro contraste entre explicación y argumentación. Nos encontramos con un lugar común en toda el área de teoría de la argumentación: desde enfoques en lógica informal (Govier:2005) hasta pragma-dialéctica (van Eemeren & R. Grootendorst:1984). A veces, hilando más fino, encontramos esfuerzos por identificar marcadores que distingan argumentación-explicación, por ejemplo: ciertas transformaciones que indican asimetrías (Marraud: 2013).

En esta investigación propongo buscar la diferencia entre preguntas “por qué” argumentativamente orientadas y las explicativas, en el tipo de contexto en el que surgen o se suscitan, es decir, en diferencias en cuanto a sus presuposiciones. En otros términos, la primera tesis que defiendo es que la demarcación entre preguntas “por qué” explicativas y argumentativamente orientadas se da en que sus presuposiciones son distintas. Respecto a cuáles son esas diferencias, planteo dos hipótesis de trabajo. Las hipótesis están formuladas en la terminología canónica que modela a las preguntas “por qué” como el triplete  $Q = \langle P_K, X, R \rangle$ , de tema, clase contraste y relación de relevancia. Las hipótesis son las siguientes:

- H1.** En una  $Q$  argumentativamente orientada, la clase contraste es un entramado de argumentos.
- H2.** En una  $Q$  argumentativamente orientada, la relación de relevancia no es a través de una teoría de trasfondo  $K$  sino de un entramado de argumentos  $K$ .

A fin de esclarecer el significado y alcance de cada hipótesis, presento un ejem-



plo estándar de cómo surge una pregunta “por qué” explicativa dadas sus presuposiciones características. Después defino la frase “entramado de argumentos”. Justifico esta definición estipulativa mediante el recurso de considerar diversos candidatos para clase contraste X o teoría de trasfondo K de una PAO. En particular: el modelo de las antinomias de Rescher (2016), *dialectical tier* de Johnson (2000), argumentos conductivos de Wellman (1971), la dialéctica de argumentos de Marraud (2015). También exploro, como relevante en esta discusión, la ponderación de razones (Chang: 1998) y las situaciones dilemáticas sin ponderación de Adam Morton (1991).

En la tercera parte, discuto los *desiderata* o criterios de adecuación CA para justificar una elección entre H1 o H2. Después procedo a analizar dos ejemplos: un argumento presentado en forma explícitamente dialógica (Perry: 1978) y otro en presentación directa (Williams: 1981). En ambos aparece explícitamente una pregunta “por qué”. Uno de los *desideratum* es que los métodos de análisis de argumentos deben hacer explícitas las presuposiciones de las preguntas “por qué” argumentativamente orientadas. En ambos casos el análisis utiliza técnicas de la dialéctica formal de Rescher (1977), los *profiles of dialogue* de Walton (1999) y las técnicas de diagramación de argumentos de Marraud (2013). Otro *desideratum* es que admita una interpretación *pro tanto* de las razones. El análisis realizado muestra que las presuposiciones de la pregunta “por qué” en los dos ejemplos deben ser entendidas en términos de H2.

Por último, discuto si los ejemplos analizados y su interpretación a la luz de H2 cumple los criterios de adecuación propuestos. Concluyo que sólo los cumplen parcialmente. Aunque la investigación se restringe a ejemplos de argumentación filosófica por las razones heurísticas antes señaladas, las técnicas de análisis y la interpretación H2 son fácilmente extrapolables a otros contextos argumentativos.

# Usos asertivos, directivos y expresivos de argumentación

**Hubert Marraud**

Universidad Autónoma de Madrid

hubert.marraud@uam.es

## 1

Si una buena argumentación es aquella que sirve para la finalidad con la que ha sido concebida, para dar criterios de buena argumentación habrá que empezar por especificar cuáles son las finalidades de argumentar. A este respecto, hay que distinguir la función constitutiva de argumentar de los propósitos de quien argumenta y de los fines del intercambio en el que participa y se enmarca su contribución. Aunque tomo esta distinción de Dima Mohammed (2016) –y también la distinción entre propósitos y fines extrínsecos e intrínsecos–, me aparto de ella al considerar que lo que Mohammed tiene por la función propia de la argumentación es en realidad parte de los propósitos intrínsecos de los argumentadores, y que lo que ella toma por propósitos intrínsecos son más bien propósitos extrínsecos (*Tabla 1*).

|                                              | Mohammed (2016)                                                                                 | Yo                                                     |
|----------------------------------------------|-------------------------------------------------------------------------------------------------|--------------------------------------------------------|
| Función                                      | Justificar una pretensión o demostrar racionalidad                                              | Presentar algo a alguien como una razón para otra cosa |
| Propósito intrínseco de los argumentadores   | Convencer o persuadir racionalmente                                                             | Mostrar a alguien que hay buenas razones para algo     |
| Propósitos extrínsecos de los argumentadores | Imponer su opinión, demostrar aptitudes, etc.                                                   | Persuadir al destinatario, ganar su asentimiento, etc. |
| Fin intrínseco del intercambio argumentativo | Examinar críticamente un asunto                                                                 |                                                        |
| Fin extrínseco del intercambio argumentativo | Ponerse de acuerdo, llegar a un acuerdo, tomar una decisión, establecer responsabilidades, etc. |                                                        |

*Tabla 1:* Función, propósito y fin de la argumentación

Mi posición es que la función constitutiva de argumentar es presentar algo a alguien como una razón para otra cosa; que quien argumenta lo hace con el

propósito de mostrar al destinatario que hay buenas razones para algo y lo hace porque existe un compromiso compartido de examinar críticamente un asunto. A su vez hay que distinguir entre propósitos y fines intrínsecos y extrínsecos de la argumentación. Los propósitos y fines intrínsecos lo son de la argumentación en y por sí misma y en cualquier contexto, mientras que los propósitos y fines extrínsecos de la argumentación dependen del contexto en el que se argumenta (Mohamed 2106:223). Por tanto, los propósitos intrínsecos están presentes en cualquier práctica argumentativa y son los propósitos extrínsecos lo que permiten distinguir unas prácticas de otras. Por ejemplo, mostrar que hay razones que avalan una tesis es un propósito intrínseco del argumentador mientras que inducir un estado mental u obtener el asentimiento del destinatario a esa tesis es un propósito extrínseco. Del mismo modo, criticar una posición (es decir, analizarla pormenorizadamente y valorarla según los criterios propios de la materia de que se trate) es un fin intrínseco de la deliberación y legitimar las decisiones resultantes sería un fin extrínseco de la deliberación.

## 2

Una caracterización de argumentar y de su función, fines y propósitos debe abarcar los distintos tipos de argumentaciones. Se suele distinguir entre argumentos y argumentaciones teóricas, prácticas y valorativas, diciendo que la argumentación teórica tiene que ver con los hechos, la argumentación práctica con las acciones y la argumentación valorativa con las actitudes. Aunque la expresión “tiene que ver” no es demasiado precisa, este modo de trazar la distinción se basa en aquello para lo que se dan razones (para creer, para actuar o para juzgar). La distinción se puede formular también en términos semánticos, como hacen Christian Kock (2007) y Luis Vega (2013), diciendo que mientras la argumentación teórica trata de proposiciones, la argumentación práctica trata de propuestas<sup>1</sup>. Una proposición es el contenido semántico de un acto asertivo, y puede ser juzgada como verdadera o falsa; una propuesta es el contenido semántico de un acto directivo de un cierto tipo, y no es ni verdadera ni falsa, sino más o menos conveniente o inconveniente<sup>2</sup>. Otra manera de presentar la distinción atiende, no al objeto de la discusión, sino del efecto que se busca inducir en el destinatario: la adopción de una creencia, de una intención o de una actitud, respectivamente. Pese a su proximidad, pasar de una caracterización a otra comporta pasar del propósito intrínseco del argumentador al propósito extrínseco.

La esfera pública designa al espacio social en el que los ciudadanos piden, dan y examinan razones sobre las medidas, políticas y leyes que deben adoptarse desde un amplio rango de perspectivas sociales. El análisis de la argumentación

---

<sup>1</sup>Ni Kock (2007) ni Vega (2013) toman en consideración la argumentación valorativa. En un artículo posterior Kock sitúa las “evaluaciones” (como *Hubble fue el mayor astrónomo del siglo XX*, *Picasso apesta* o *¡Puaj! Los campesinos son repugnantes*) en un lugar intermedio entre las proposiciones y las propuestas (Kock 2011:72).

<sup>2</sup>Con respecto a esta distinción, Kock señala que “la diferencia que he resaltado entre proposiciones y propuestas para la elección deliberada es básicamente una reflexión sobre distinciones reconocidas [...] también en la filosofía moderna, sobre todo en la distinción de la filosofía de los actos de habla entre asertivos, por una parte, y directivos, comisivos, etc., por la otra” (Kock 2011:73).

en la esfera pública revela que quienes argumentan en ese espacio lo hacen frecuentemente con propósitos distintos de la persuasión (es decir, de la inducción en el destinatario de estados mentales reflexivos). Pero si es así, la persuasión no puede ser un propósito intrínseco de la argumentación. Esta es la razón fundamental para rehacer el cuadro de Mohammed de la función, los propósitos y los fines de la argumentación. Para describir las prácticas argumentativas en la esfera pública hay que ampliar el dominio de la argumentación, como hace la propuesta de considerar que el propósito intrínseco de argumentar es mostrar que hay razones para algo.

La persuasión solo ha lugar cuando existe una diferencia de opinión y se asume que se puede resolver argumentando. Como consecuencia, si la persuasión fuera el propósito intrínseco de argumentar, no podría haber argumentación cuando no hay una diferencia de opinión o cuando no puede ser resuelta con razones. Para Kock (2007) y Asen (2005) la deliberación política en las sociedades contemporáneas discurre a menudo en situaciones en las que falla alguno de los dos presupuestos de la argumentación suasoria. Por tanto, o la persuasión no es el propósito intrínseco de la argumentación o parte de la argumentación pública es solo aparente. Robert Fogelin introdujo en 1985 la denominación “desacuerdos profundos” (*deep disagreements*) para aquellos desacuerdos que no pueden resolverse argumentando, etiqueta que con el tiempo ha pasado a designar todo un área de investigación en teoría de la argumentación. Cuando no se dan las condiciones que hacen posible resolver racionalmente una discrepancia, dice Fogelin, “puede persistir el lenguaje argumentativo, pero se vuelve inútil, puesto que apela algo que no existe: un trasfondo compartido de creencias y preferencias” (1985:5).

### 3

Abordaré a continuación las siguientes cuestiones: ¿Qué usos no suasorios de la argumentación se han identificado en la esfera pública? ¿Esos usos no suasorios son propios de la argumentación práctica o también podemos encontrarlos en la argumentación teórica? ¿Son casos genuinos de argumentación o en ellos “solo persiste en lenguaje argumentativo”?

Kock mantiene que el propósito de la deliberación pública, como forma típica de la argumentación práctica, no puede ser el consenso de los participantes. La deliberación trata de propuestas y puede haber (buenas) razones inconmensurables a favor y en contra de una propuesta. Que haya razones inconmensurables no quiere decir que sean incomparables, sino más bien que existen múltiples maneras de compararlas, con distintos resultados, sin que haya ningún modo objetivo de anteponer una de esas escalas a las demás. Para Kock el propósito extrínseco de la deliberación política es presentar y desarrollar las razones a favor y en contra de un curso de acción para ayudar al destinatario a tomar una decisión reflexiva de acuerdo con sus preferencias valorativas. Esas preferencias valorativas escapan del ámbito de la deliberación racional: “dado que no existe un razonamiento intersubjetivamente convincente que determine tales elecciones (y si lo hubiese, no serían elecciones), son de hecho subjetivas.” (Kock 2007:237). Así, cuando hay buenas razones a favor y en contra de una posición

y esas razones son inconmensurables, la argumentación no puede tener como propósito la persuasión, aunque tiene sentido porque ayuda a elegir la posición más apropiada de acuerdo con sus preferencias valorativas.

En la misma línea Asen aduce que la combinación del pluralismo de valores y los desacuerdos básicos hace que en muchos casos la persuasión sea imposible en las sociedades democráticas contemporáneas. Eso que muestra, según Asen, que el propósito intrínseco de la argumentación en la esfera pública no puede ser la persuasión. Asen identifica tres propósitos alternativos de la argumentación en una esfera pública caracterizada por el pluralismo de valores y los desacuerdos profundos (2005:119):

- ampliar la agenda para incluir los intereses y perspectivas de individuos, grupos y asuntos marginados;
- establecer responsabilidades en situaciones en las que la parte más débil obliga a la más fuerte a asumir la responsabilidad de las decisiones tomadas en situaciones de deliberación restringidas, y
- moldear identidades, individuales y colectivas.

## 4

Kock parece pensar que la función no suasoria de la argumentación es propia de la argumentación práctica, en tanto que esta trata de propuestas, y ajena a la argumentación teórica, que trata de proposiciones verdaderas o falsas. Se me ocurren dos interpretaciones del razonamiento de Kock. Según la primera, la ponderación es característica de la argumentación sobre propuestas y ajena a la argumentación sobre proposiciones. Según la segunda, la diferencia reside en que hay razones prácticas inconmensurables, pero no razones teóricas inconmensurables.

Ejemplos como el rombo de Nixon muestran que también hay ponderación en la argumentación teórica. Dado que Richard Nixon era cuáquero y republicano, las reglas *los cuáqueros suelen ser pacifistas* y *los republicanos no suelen ser pacifistas* permiten extraer las conclusiones contradictorias *Nixon era pacifista* y *Nixon no era pacifista*. Como la conclusión de estos dos argumentos es una afirmación de hecho, tenemos un conflicto entre dos argumentos teóricos aceptables. Como no es admisible concluir simultáneamente que Nixon era pacifista y que no era pacifista, el agente debe determinar cuál es el peso relativo de esas reglas en esta situación particular. No obstante, la posición de Kock podría salvarse asumiendo que las razones teóricas, a diferencia de las prácticas, son siempre conmensurables.

El supuesto de que las razones teóricas, a diferencia de las prácticas, son siempre conmensurables no es evidente, para empezar porque apenas hay estudios y discusiones de la ponderación de argumentos teóricos. Carl Wellman señala que la ponderación de razones aparece cuando concurren varios criterios para la aplicación de un término, de desigual importancia, y que pueden ser satisfechos en mayor o menor medida (1971:54). Imaginemos que se discute si *a* es un *P* o no y que hay varios criterios distintos para determinar si lo es o no. Para llegar

a una conclusión justificadamente habría que tener en cuenta (1) la importancia relativa de esos criterios, (2) el grado en el que *a* cumple cada uno de ellos, y (3) el riesgo de la inferencia y las consecuencias de acertar o equivocarse (Pinto 2001:120). Pero (1) y (3) parecen tan opinables como las elecciones entre valores que están en el trasfondo de la deliberación en la esfera pública.

## 5

Si argumentar es presentar algo a alguien con el propósito intrínseco de mostrar que hay buenas razones para otra cosa, las diferencias entre la argumentación teórica y la argumentación práctica son mucho menores de lo que muchos teóricos piensan. Concluyo pues que no hay una diferencia sustancial entre la ponderación de razones prácticas y teóricas y que tanto la argumentación teórica como la argumentación práctica pueden tener usos suasorios y no suasorios. Se sigue además de esta manera de entender la argumentación que tales usos no suasorios son casos genuinos de argumentación.

Se necesita entonces algún modo de clasificar los usos de la argumentación. Las tipologías de los actos de habla pueden servir de inspiración a este respecto: la división de los argumentos en teóricos, prácticos y valorativos se basa en el tipo de conclusión, y la diversidad de los usos de la argumentación invita a abandonar la tesis de que ésta es siempre el contenido de una aserción.

Siguiendo la tipología de los actos de habla de John Searle, se podrían distinguir cuatro tipos de usos de la argumentación: asertivos, directivos, compromisivos y expresivos<sup>3</sup>. Naturalmente, las pretensiones de validez son distintas en cada caso. Habermas señala que, dependiendo del tipo de argumento, un argumento es válido si da razones para creer que una aserción se ajusta a los hechos, que una acción se ajusta a normas reconocidas como legítimas, o que es la expresión veraz de una vivencia. No obstante el carácter representacional y no dialógico de las definiciones de Searle (1975:459–462) dificultan su extrapolación a los argumentos.

Una primera deficiencia es que parece haber actos que son a la vez compromisivos y directivos, como suponer, en el sentido expresado por “Supongamos que P”, o proponer.

... *proponer* envuelve (i) la descripción de una acción o un curso de acción; (ii) una actitud proactiva [comisiva] al respecto; y por lo regular (iii) una invitación a que el interlocutor o los destinatarios del discurso compartan el compromiso. (Vega 2013:3).

Otra deficiencia es que las aserciones tienen también un aspecto compromisivo. En el contexto apropiado no hay ninguna diferencia entre aseverar “Nixon no es pacifista” o decir “Créeme, Nixon no es pacifista”. Pero, según la descripción de Searle, en el primer caso se trataría de un acto asertivo, mientras que en el segundo, dado que se pide al oyente que haga algo, parece tratarse de un directivo. Una consecuencia, referida específicamente a la extrapolación de esas categorías a la teoría de la argumentación, es que la categorización de Searle

---

<sup>3</sup>Asumo que las declaraciones no pueden introducir la conclusión de un argumento.

hace que toda argumentación suasoria, teórica o práctica, sea directiva, puesto que su propósito es invitar al destinatario a adoptar un compromiso.

La caracterización y clasificación de los actos ilocucionarios de Bach y Harnish (1979) puede ser más apropiada para buscar una clasificación de los usos de la argumentación.

Para nosotros los *constatativos* [es decir, los asertivos] expresan la creencia del hablante y su intención o deseo de que el oyente tenga o se forme una creencia igual. Los *directivos* expresan la actitud del hablante hacia alguna posible acción futura del oyente y su intención de que su preferencia, o la actitud que expresa, sea tomada como una razón para la acción del oyente. Los *compromisivos* expresan la intención y la creencia del hablante de que su preferencia le obligue a hacer algo (quizá bajo ciertas condiciones). Y los *reconocimientos* expresan sentimientos acerca del oyente, o en casos en los que la preferencia es claramente superficial o formal, la intención del hablante de que su preferencia satisfaga la expectativa social de expresar ciertos sentimientos y su creencia de que lo hace. (Bach y Harnish 1979:41).

La caracterización de Bach y Harnish explica el carácter compromisivo de algunos asertivos: el hablante expresa su creencia de que algo es verdad e invita al oyente a creerlo. Además, su descripción de los directivos los hace argumentativos por naturaleza: el hablante expresa una actitud e intenta dar una razón al oyente, mezclando la argumentación práctica y valorativa.

## 6

Como sería prematuro intentar establecer una clasificación de las argumentaciones sobre las bases expuestas, me limitaré a una serie de apuntes.

1. La concepción lógica tradicional de la argumentación es asertiva en el sentido de Searle: las premisas y la conclusión son aserciones y argumentar es extraer información de las premisas.
2. La deliberación con razones inconmensurables descrita por Kock es otro ejemplo de argumentación asertiva.
3. Los usos suasorios de la argumentación teórica son argumentaciones asertivas, en el sentido de Bach y Harnish, igual que los usos suasorios de la argumentación valorativa. Cuando se argumenta así, la conclusión expresa una creencia del hablante y su intención o deseo de que el oyente la adopte.
4. Los usos de la argumentación práctica para convencer al auditorio de la conveniencia de una propuesta corresponden a usos compromisivo-directivos de la argumentación.
5. Los casos, mencionados por Asen, en los que se argumenta para ampliar la agenda o para reafirmar identidades son ejemplos plausibles de argumentación expresiva.

6. Cuando el objeto de la argumentación es establecer y asumir responsabilidades, el otro uso mencionado por Asen, parece compromisiva. Aunque para Searle el objeto de los actos compromisivos es comprometer al hablante con un curso futuro de acción y la asunción de responsabilidades se refiere a actos pasados, también es cierto que acciones como responsabilizarse o excusarse no tienen fácil acomodo en la clasificación de Searle. Bach y Harnish clasifican excusar a alguien entre los actos directivos (1979: 47), por lo que parece natural clasificar excusarse entre los actos compromisivos.

## Referencias

- Asen, Robert (2005): "Pluralism, Disagreement, and the Status of Argument in the Public Sphere". *Informal Logic*, 25(2):117–137.
- Bach, Kent & Harnish, Robert M. (1979): *Linguistic Communication and Speech-acts*. Cambridge, Mass.: MIT Press.
- Fogelin, Robert J. (1985): "The Logic of Deep Disagreements". Reimpreso en *Informal Logic*, 25(1)(2005):3–11.
- Kock, Christian (2007): "Dialectical Obligations in Political Debate". *Informal Logic*, 27(3):233–247.
- Kock, Christian (2007): "Why Argumentation Theory Should Differentiate Types of Claims". En Blair, J. Anthony y Johnson, Ralph H. (eds.) *Conductive Argument: An Overlooked Type of Defeasible Reasoning*, 62–78. Londres: College Publications 2011.
- Mohammed, Dima (2016): "Goals in Argumentation: A Proposal for the Analysis and Evaluation of Public Political Arguments". *Argumentation*, 30:221–245.
- Pinto, Robert C. (2011): "Weighing Evidence in the Context of Conductive Reasoning". En Blair, J. Anthony y Johnson, Ralph H. (eds.) *Conductive Argument: An Overlooked Type of Defeasible Reasoning*, 104–126. Londres: College Publications 2011.
- Searle, John (1975): "Una taxonomía de los actos ilocucionarios". En Luis M. Valdés Villanueva (ed.) *La búsqueda del significado*, 449–476. Madrid: Tecnos, 1991.
- Vega Reñón, Luis (2013): "Argumentando una innovación". *Revista Iberoamericana de Argumentación*, 7:1–17. <https://revistas.uam.es/index.php/ria/article/view/8182/8523>
- Wellman, Carl (1971): *Challenge and Response. Justification in Ethics*. Carbondale: Southern Illinois University Press.



# Presuposiciones, preguntas y falacias

**Andrei Moldovan**

Universidad de Salamanca  
mandreius@usal.es

La “falacia de la pregunta compleja” (FPC) se suele introducir a partir de ejemplos paradigmáticos, como: “¿Te arrepientes de haber sido injusto con tus oponentes políticos?” (Greco 2003). Tales preguntas conllevan una presuposición controvertida, en este caso de que el oyente fue injusto con sus oponentes políticos. Como es una pregunta de tipo “sí” o “no”, las dos respuestas posibles conllevan la misma presuposición. De allí la naturaleza engañosa de la pregunta, y el nombre de “falacia” que se le da.

¿Es realmente FPC una falacia? Es cuestionable que lo sea, si tomamos en consideración las definiciones más populares de falacia (Hanson 2002). Eso se debe a que no parece posible reconstruir la FPC como un argumento, a pesar de intentos como el de Jacquette (1994). Considero también la propuesta de ver la FPC como una táctica sofística de naturaleza dialéctica (Pilgram y Polcar 2007).

Argumentaré que la teoría de las presuposiciones de Stalnaker (1974, 2002) y la discusión de la “acomodación” de las presuposiciones en Lewis (1979) pueden arrojar luz sobre la relación entre la normatividad argumentativa (i.e. el carácter falaz) de la FPC, y la normatividad inherente y característica de los actos de habla de presuponer (García-Carpintero 2010).

# Una invitación a tomar con filosofía la argumentación

Luis Vega Reñón  
UNED  
lvegar@gmail.com

Pensemos en una biblioteca con una sección de Lógica, Teoría de la argumentación o estudios afines relativamente abastecida. Pongamos la biblioteca de Humanidades de la UNED sin ir más lejos. Los libros dedicados a filosofía de la lógica son más de 70 según el catálogo. Los expresamente dedicados a filosofía de la argumentación no pasan de 2. Es un indicador de la que podríamos llamar “invisibilidad” de la filosofía de la argumentación<sup>1</sup>.

Ser prácticamente invisible no implica ser prácticamente inexistente. De hecho, hay filosofías de la argumentación –en un sentido preciso– aunque por lo regular no sean muestras declaradas sino más bien subyacentes o implícitas e incluso encubiertas. Les propongo considerar algunas de estas cuestiones y concepciones filosóficas de interés para la teoría y la práctica de la argumentación. En particular, estas tres: 1/ La tensión entre la universalidad y la diversidad del discurso argumentativa; 2/ la cuestión de las señas de identidad de la interacción argumentativa, e. g. competitivas vs. cooperativas, 3/ la opción entre concepciones presuntamente alternativas como la analítica estática del argumento-producto vs. la construcción dinámica de la argumentación. Será una consideración provisional y esquemática, como corresponde a una aproximación inicial a este terreno movedizo.

## Sobre la universalidad o unidad y la diversidad del discurso argumentativo

Cabe suponer, de entrada, una capacidad común de los seres humanos de razonamiento y uso general de la inferencia inherentes a la comunicación y al entendimiento mutuo, a la interacción lingüística sin ir más lejos. En este sentido, se reconoce que la razón y el lenguaje discursivo son una facultad universal y distintiva del género humano.

Por otro lado, también es familiar la existencia de manifestaciones y desarrollos específicos distintos de estas competencias básicas en diversos medios socioculturales. Baste comparar el caso uzbeko (Luria 1976), con nuestro caso WEIRD

---

<sup>1</sup>Por lo que recuerdo, tampoco es muy generosa la literatura al uso en atención a la existencia de 3 títulos propios (Govier 1999, Blair 2012, Tindale 2015) y otros 2 poco significativos en el presente contexto (Natanson & Johnstone 1965, Perelman 1968).

de gente occidentales (W), educados (E), industrializados (I), ricos (R) y democráticos (D) (Mercier & Sperber 2017). En el marco de la preparación de una campaña de alfabetización y escolarización de los campesinos uzbekos, Luria estudió en los años 1931-32 su comportamiento discursivo a partir de unos tests como los siguientes (1976:108-112).

| I                                                                                          | II                                                                                                                                |
|--------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| El algodón solo puede crecer<br>donde hay calor y está seco<br>Inglaterra es fría y húmeda | En el lejano Norte, donde hay<br>nieve, todos los osos son blancos<br>Nueva Zembla está en el lejano<br>norte y siempre hay nieve |
| ¿Puede crecer el algodón allí?                                                             | ¿De qué color son los osos allí?                                                                                                  |

Para un adolescente WEIR serían ejercicios triviales de deducción; si fuera un escolar aplicado, vería en I y II un *modus tollens* y otro *ponens* respectivamente. En cambio, los campesinos, por regla general analfabetos y de mediana edad, coincidieron en tomar las preguntas no como parte de un examen, sino como peticiones de una información real que no estaban en condiciones de atender porque el caso nada tenía que ver con su experiencia personal o con la de sus paisanos. Solo un par de ellos, más jóvenes, aludieron a lo que parecía desprenderse de las palabras del encuestador<sup>2</sup>. ¿Se trata de una diferencia radical y sintomática de que nuestra competencia inferencial es un rasgo propio de nuestra cultura letrada, no compartido por otras culturas primitivas o pre-lógicas? En consecuencia, ¿la razón o el razonamiento no es universal?

Hay diversos planteamientos de esta cuestión, aunque algunos resulten eventualmente complementarios o incluso solapados. Muy esquemáticamente cabe mencionar algunos tipos principales<sup>3</sup>:

- (a) Los que descansan en una pretendida base biológica, e. g. Baron-Cohen (2003).
- (b) Los que adoptan una perspectiva psicosocial para contraponer las personalidades occidental y oriental, especialmente asiática (china, japonesa), conforme al cliché acuñado por Nisbett (2003): la occidental es categórica, individualista, competitiva e inconformista; la asiática, en cambio, es flexible, colectivista, cooperativa, ritualista. De ahí se desprenden otras diferencias discursivas.
- (c) Los que se fundan en consideraciones y análisis socioculturales. Un punto tradicional de debate ha sido la tesis de la “Great Divide”, de la divisoria entre la mentalidad pre-lógica y la mentalidad civilizada o letrada. Alcanzó a tener resonancia en las primeras décadas del s. XX a través de las publicaciones socio-antropológicas de Lévy-Bruhl, sin dejar de propiciar derivaciones delirantes: el propio Lévy-Bruhl, dado que un bororo

<sup>2</sup>Por otro lado, las formulaciones no son completamente determinantes, no son del tipo “¿De qué color es el caballo blanco de Santiago?”. En el caso II, un uzbeko griceano podría pensar que, dadas las premisas sentadas por el encuestador, su pregunta solo tenía sentido si le cabía alguna duda sobre el color de los osos de Nueva Zembla en particular.

<sup>3</sup>Dejo a un lado tesis más radicales y discutibles como la que vindica la unidad, universalidad y uniformidad de la lógica (cf. Deaño 1980:235-245, y su revisión crítica en Vega 2003, 2015:230-243), por no ser directamente pertinentes en el presente contexto.

podía considerarse al mismo tiempo hombre y loro Arara, entendía que esto suponía negar los principios de identidad y no contradicción; en los años 1970, el historiador de la lógica Dumitriu quiso salvar los muebles atribuyendo a la mentalidad primitiva una lógica polivalente. Hoy es una curiosidad, a partir de la idea misma de mentalidad sometida a un severo juicio crítico por especialistas en estudios históricos e interculturales, como Geoffrey Lloyd. Lloyd, una autoridad en las antiguas culturas griega y china, y en su proyección científica y filosófica, sostiene por su parte la innegable universalidad y unidad del razonamiento, así como la notoria diversidad sociocultural de las prácticas discursivas y cognitivas (Lloyd 2007, 2013). Baste reparar, por ejemplo, en la ausencia de programas y de pruebas axiomático-deductivas en la antigua cultura china, por contraste con la antigua matemática griega –aunque esta diferencia no haya sido un obstáculo para la incorporación posterior de Euclides y Arquímedes a su acervo lingüístico y cultural–. Algo parecido puede decirse a propósito de otras peculiaridades discursivas y cognitivas chinas que están lejos de implicar una razón singular. Más recientemente, Mercier (2011) y Mercier & Sperber (2017) vienen defendiendo la universalidad del razonamiento argumentativo como factor evolutivo principal y característico de la especie humana.

Estoy de acuerdo con los planteamientos de Lloyd y de Mercier & Sperber. Pero creo que, especialmente en estos últimos, aparte de su proyección filosófica problemática, hay una referencia equívoca a la argumentación que conviene despejar<sup>4</sup>. Para este efecto propongo distinguir entre el nivel (1), primario e individual, de las competencias y actividades de *razonar* y *argüir* que son la dotación común de los seres humanos –seres dotados de lenguaje, entendimiento y razón–, y el nivel (2), secundario y social, de las diversas prácticas específicas de *argumentar* –objeto de aprendizaje–. En el nivel (1) se discurre más bien por motivos; en el (2), por razones y compromisos públicos de los que los agentes discursivos son responsables y capaces de rendir cuentas. No constituyen compartimientos estancos, aunque admitan polarizaciones como los extremos blanco y negro de una gama de grises (pensemos en un debate académico estándar frente a la discusión que caricaturiza el sketch de los Monty Python, “The argument clinic”). En todo caso, el papel de mediadora entre la actividad de *argüir* y la práctica de *argumentar* corresponde a la conversación como medio discursivo donde ambas se desenvuelven. Como instancia de contrastación y corroboración de esta hipótesis sugiero considerar, al margen de la antropología de campo, la adquisición y el desarrollo de las competencias discursivas y argumentativas en la infancia, en particular durante el periodo 5–12 años (cf. Kuhn 2012, Pérez, Postigo y García-Mila 2016). Esta investigación psicopedagógica desmiente además la alternativa nula a mi distinción entre *argüir* y *argumentar*, representada por tesis como «La actividad de argumentar es coextensiva a la actividad de hablar y tan pronto como se habla, se argumenta» (vid. Plantin 1988:118).

---

<sup>4</sup>Es sintomático que, a propósito del caso uzbeko, ellos mismos señalen: “Aun si aceptamos que todo el mundo, escolarizado o no escolarizado, puede razonar, esto no significa que todo el mundo argumente” (2017:284).

## La cuestión de las señas de identidad

Como se sabe, la idea de que toda tesis o posición trae consigo la contraria se remonta a Protágoras y al movimiento sofístico del s. V a.n.e. Otro momento decisivo en esta tradición dialéctica fue la cultura medieval de la *disputatio*, un género no solo escolar y académico sino también literario. Estos antecedentes han sentado una concepción todavía dominante de las señas de identidad beligerantes de la argumentación: en esta línea se supone que la existencia de una discrepancia o una diferencia de opinión es un rasgo definitorio o, al menos, una condición necesaria de la argumentación. Según sostenía Willard: “*Argumentación* es una forma de interacción en la que dos o más personas sostienen lo que consideran ser posiciones incompatibles” (1989:1). Es una tesis enfatizada por famosas metáforas: “la argumentación es la guerra”, “un debate es un combate”, aunque por lo regular sus manifestaciones suelen ser más contenidas.

Pero desde finales del siglo pasado, esta identificación de la interacción argumentativa ha entrado en crisis. La crisis ha venido provocada por varios factores **internos** y **externos**. Mencionaré los más relevantes a mi juicio.

(a) Entre los factores **internos** destaca una suerte de autocrítica en el campo de la lógica informal y la filosofía de la argumentación que se alza contra la beligerancia de la metáfora bélica (e. g. Cohen 1995). Mayor importancia a medio y largo plazo tienen los planteamientos alternativos, como la idea de controversia avanzada por Dascal (1998) en el terreno del discurso filosófico y científico, y pronto asumida en otros círculos, e.g. por van Eemeren (2008). En este punto, conviene recordar la existencia histórica de diversas tradiciones occidentales de las controversias, frente al supuesto poder definitorio de la confrontación dialéctica. Cattani (2007) ha reconocido a este respecto las controversias que se traman como objeto de placer o disfrute, o como vindicación del derecho a discutir y cuestionar o, en fin, como respuesta a las demandas y obligaciones de dialogar o confrontar alternativas en determinadas situaciones. Pero aún cabría añadir las armadas para el ejercicio o el entrenamiento escolar o para las prácticas deliberativas, por ejemplo.

En todo caso, la identificación “guerrera” de la argumentación tiende a privilegiar una perspectiva dicotómica o bivalente, aunque procure no ser agresiva y renuncie a sistemas de perdedores y ganadores en favor de papeles pro/contra. El supuesto crítico («si yo sostengo X y creo estar en lo cierto, quien sostenga o sugiera lo contrario está equivocado») ignora tanto la posibilidad de alternativas parejamente razonables, como la constitución dinámica e interactiva de los procesos argumentativos. Pluralidad y contraste no implican necesariamente polémica.

(b) Entre los factores **externos** cabe mencionar la crítica feminista y los reparos procedentes de la que llamaría “economía psicosocial de la argumentación”. Estos últimos tienen que ver con los riesgos y las consiguientes exhibiciones o inhibiciones que conlleva la exposición a los debates públicos que acusan modos agresivos, por contraste con los beneficios que se esperan de las deliberaciones conjuntas y cooperativas. Se trata de un terreno cuyo estudio se halla todavía en fase de aproximación y exploración.

Más maduro y decidido se presenta el caso del feminismo (vid. Palzewsky 1996). A mi juicio –y a riesgo de meterme en un jardín–, algunas de sus propuestas acusan un exceso de categorización dicotómica y cierta falta de elaboración conceptual en cuestiones técnicas (e. g. ponderación, deliberación). Pero, desde luego, son inestimables sus servicios críticos. Por un lado, ha ampliado el campo de la argumentación con referencias nuevas o renovadas al valor del testimonio, a la experiencia personal y las historias de vida, a los ejemplos, etc. Por otro lado, ha abierto nuevas perspectivas filosóficas sobre el sentido no ya agresivo o competitivo, sino comprensivo y cooperativo de la práctica de la argumentación. Son perspectivas que pueden atribuirse incluso tradiciones históricas propias como la que podría representar la alternativa performativa de los salones parisinos del s. XVII y su cultura de la conversación (vid. Craveri 2003).

Cambiando de cultura y de siglos, cabe recordar el caso parecido del daoísmo, aunque tenga otros propósitos, no precisamente vindicativos. La concepción daoísta de la confrontación la considera una interacción discursiva inconclusa e inconcluyente, abierta. Trata y practica la argumentación como proceso no adversativo ni competitivo, sino en busca de esclarecimiento y comprensión o con la intención de dar que pensar. Por lo demás, también acude a experiencias personales e historias de vida, testimonios, ejemplos, metáforas y analogías (Combs 2004).

Otros intentos de establecer unas señas de identidad argumentativa discurren por las vías de las dimensiones tradicionales: en atención a la estructura lógica del argumento, o por referencia a los procesos de confrontación dialéctica o, en fin, a la luz del propósito de persuadir al interlocutor o los alocutarios involucrados, reales o virtuales. Pero me temo que son intentos descaminados porque tienden a privilegiar unas determinadas señas de identidad antes de atender a las señales concretas de identificación marcadas por los usos de la argumentación en diversos contextos. Una alternativa viene a ser considerar las metas o fines de la actividad de argumentar en el marco de una distinción capital entre fines intrínsecos, inherentes a la actividad de argumentar en cualquier contexto, y fines extrínsecos, dependientes del contexto (Mohammed 2016). Resultan, en suma, cinco tipos de fines. Tres intrínsecos: (i) el objetivo constitutivo de la argumentación: su justificación o racionalidad manifiesta; (ii) la función intrínseca del acto de argumentar consistente en la persuasión racional; (iii) la función intrínseca de la interacción argumentativa, a saber, la contrastación crítica de posiciones y puntos de vista. Y dos extrínsecos: (iv) los usos del acto de argumentar que responden a fines individuales; (v) los propósitos de la interacción argumentativa que responden a fines colectivos. Con todo parece latir una intención definitoria; es sintomático que la recapitulación final no recoja el objetivo intrínseco (i), quizás al darse por descontado en su calidad de objetivo determinante de la argumentación (vid. 2016:241). Es una propuesta arriesgada: envuelve la problemática idea de función<sup>5</sup> y, en particular, de *función definitoria* (cf. Goodwin 2007, Doury 2012). Otro aspecto discutible es la indistinción entre actividades y prácticas. Recordando en este punto la distinción aristotélica entre la producción (póiesis) que se rige por valores y fines externos a una actividad, y la acción (práxis) cuyo propósito se encuentra en ella misma. Cabe plantearse

---

<sup>5</sup>Es sintomático que Blair (2012) no hable de funciones, sino de usos intrínsecos e incidentales.

si las prácticas argumentativas son productivas o performativas (Tanesini 2017), o pueden ser las dos cosas en distinta medida según su soporte conversacional y su contexto de uso.

## Concepciones presuntamente alternativas como la analítica estática del argumento-producto vs. la construcción dinámica de la argumentación

Recogiendo el guante que acabo de lanzar, a una consideración productiva se inclina la concepción analítica y estática centrada en el argumento-producto, generalmente asociada a una concepción monológica de la argumentación. Según una idea hoy extendida, los planteamientos monológicos de la argumentación son derivados o parasitarios dentro del marco de los dialógicos; un argumento autista –“atómico”, en el sentido de Blair 2012–, solo alcanza a tener sentido discursivo cabal dentro de un proceso dialógico de argumentación. Así, la identificación de un argumento debe incluir no solo su composición estructural (e. g. su núcleo ilativo), sino su contribución al debate o la conversación en curso y la pretensión de persuadir a un interlocutor o unos alocutarios reales o potenciales. En suma, los productos, procesos y procedimientos argumentativos se mueven en un espacio de razones, compromisos y responsabilidades de exposición y rendición de cuentas por parte de los agentes involucrados.

Mencionaré solo dos casos paradigmáticos. Uno es el de los entimemas, donde el alocutario contribuye no solo a la comprensión y asunción del argumento, sino a su construcción cabal y expresa. Recordemos la escena iii del acto 3º de *Otelo*. Yago, sin dar pruebas, vierte insinuaciones sobre la fidelidad de Desdémona y como remate final se dirige a Otelo: «Sacad entonces la conclusión». A Otelo le corresponde explicitar y completar el argumento de los celos.

El otro caso paradigmático es una historia árabe recogida en el *Libro de las argucias* (II, c. viii). Un sultán tenía un visir envidiado por sus enemigos. Tanta fue la presión que al fin el sultán ordenó arrojar al visir a su jauría de caza y que los perros lo destrozaran. El visir rogó un plazo de diez días para arreglar sus asuntos, que le fue concedido. Entonces acudió al Montero mayor con una bolsa de cien monedas de oro y le pidió que le permitiera cuidar de los perros del sultán durante los diez días. En ese tiempo, logró que se familiarizaran con él hasta asegurarse el reconocimiento y la fidelidad de todos ellos. Vencido el plazo, los enemigos del visir recordaron al sultán su sentencia. El sultán ordenó atar al visir y echarlo a los perros. Pero éstos se pusieron a dar vueltas a su alrededor y a jugar con sus ropas. El sultán, asombrado, hizo comparecer al visir: «Dime la verdad. ¿Qué ha ocurrido para que mis perros te perdonen la vida?». «He servido a los perros durante diez días –respondió el visir– y el resultado ha sido el que has visto, señor. Te he servido durante treinta años. El resultado ha sido que me condenes a muerte, instigado por las insidias de mis enemigos». El sultán enrojó de vergüenza, y devolvió al visir su dignidad y su posición anterior. Está clara, aunque la historia no la refiera, la argumentación reflexiva y práctica, deliberativa, en que el sultán convierte el escueto pero contrastado relato de su visir. Es la interpretación del sultán la que identifica y construye las palabras del

visir como un argumento práctico efectivo. Supongamos que el sultán hubiera respondido: «Es curioso: ¡qué distintos somos los hombres y los perros!». En este caso, nos habríamos quedado sin argumento práctico y, seguramente, el visir sin vida.



SIMPOSIO:  
LA ARGUMENTACIÓN DE LA  
FILOSOFÍA

# Argumentación y vigilancia epistémica en entornos digitales

**J. Francisco Álvarez Álvarez**

UNED

jalvarez@fsf.uned.es

Se analizan las ventajas de un modelo de agente no optimizador para comprender las prácticas argumentativas y se muestran las similitudes y diferencias con el propuesto desde la noción de vigilancia epistémica (Mercier, Sperber). Asumir que los agentes argumentativos exhiben una racionalidad acotada, entendida como satisfacción de criterios y no como optimización sometida a constricciones, tiene diversas consecuencias para la comprensión de los procesos comunicativos en los espacios masivos de interconexión que permiten las actuales transformaciones tecnológicas. Los espacios digitales parecen generar tendencias que fortalecen distorsiones de la comunicación, consolidan sesgos y potencian los rumores y las noticias falsas, pero también abren posibilidades de integrar la variedad epistémica y facilitar el debate entre posiciones diferentes; incluso, pueden impulsar la producción de conocimiento de calidad (Scott Page 2008). Un diseño capaz de dotar de funcionalidades adecuadas a los individuos en los espacios digitales puede superar los efectos indeseados de los agrupamientos cuantitativos, comunes en las primeras fases de las redes sociales. Experiencias como las realizadas por la Fundación Mozilla, para mejorar la calidad de internet, sugiere que facilitar las prácticas argumentativas en la interacción e interconexión digital puede contribuir a corregir algunos de los supuestos inevitables males de la sociedad interconectada.

# What is wrong with deductivism?

**Lilian Bermejo Luque**  
Universidad de Granada  
lilian.bermejo@cchs.csic.es

## Introduction: Botting's criticism

In a recent paper, David Botting (2015) criticises my rejection of deductivism as a standard for argument evaluation (Bermejo-Luque, 2011). The main goal of this paper is to show that Botting's criticism misses the point of my argument against deductivism and to explain further why deductivism is wrong.

## Arguments, Inferences, Reasonings and Argumentations

Within Argumentation Theory, arguments are commonly characterized as 'the product of argumentation'. However, if this characterization were offered as a definition of argument, it would be too vague to be adequate. If we agree that argumentation is a communicative activity, then we can think of its "product" in many different ways: as particular pieces of speech or text, as the meaning conveyed by them, as their effects on addressees, etc...

Within the framework of formal logic, an argument is usually defined as a set of propositions, one of which—the conclusion—follows from the others—the premises. But the problem with such a definition is: if the premises of an argument do not follow from the conclusion, isn't such set of propositions just a set of propositions? Yet, as Fohr (1979: 5) had observed, the common usage of 'argument'—and the very business of appraising arguments—requires that there can be bad instances of it.

In facing the difficulties of defining deductions, inductions and conductions in terms of the support that the premises provide to the conclusion, Govier (1980) and Hitchcock (1980) suggested that these are not types of inferences but types of validity or types of standards of appraisal. Yet, this was not the moral that Fohr himself drew from his observation. Instead, his suggestion was to refrain from thinking of arguments as things that exist in vacuo, but as being person-related (Fohr, 1979: 5).

The linguistic normative model of argumentation, LNMA, developed in Bermejo-Luque (2011) captures Fohr's intuition that the best option for avoiding such kind of problems is to adopt a pragmatic perspective, and to give

up Platonism altogether. LNMA follows Bach and Harnish's (1979) Speech-act Schema in order to characterize the second order speech-act complex of arguing; that is, a speech-act composed of a speech-act of adducing (i.e., the reason) and a speech-act of concluding (i.e., the conclusion or target-claim). The illocution of arguing, so characterized, counts as an attempt at showing a target-claim to be correct. Correspondingly, in LNMA argumentation goodness is characterized as a matter of both semantic conditions determining the correctness of target-claims and pragmatic conditions determining how well an act of arguing plays as an act of showing. In this model, arguments are mere representations of the particular inferences that supervene on acts of arguing and also on acts of reasoning (i.e. particular inferential judgments, which are the mental counterparts of acts of arguing).

In contrast with acts of arguing and acts of reasoning, which are, so to speak, 'objects' of the world where inferences supervene, in LNMA arguments are mere theoretical constructions, instead of abstract objects from a Platonic world. From this perspective, we would not "use" arguments, but "produce" arguments in order to represent the inferences that we make. As such representations, arguments can be obtained by displaying a variety of models, such as those of the different formal systems or informal argumentative schemas.

Hitchcock (2013) dismisses this strategy: "But, as has been argued in the case of attempts to classify reasoning and argument as deductive or inductive, such appeals to the intentions or claims or beliefs of reasoners and arguers are vacuous in many cases and are unnecessary for argument appraisal (...). As one can confirm for oneself by immediate retrospection, reasoners who draw a conclusion for themselves from information at their disposal are typically unaware of whether they are drawing it conclusively or non-conclusively. Reasoners just draw their conclusions, and it is only after that inferential act, if at all, that they determine whether their conclusion follows conclusively or non-conclusively. As for arguers, they sometimes claim a qualitative degree of support for their conclusion by qualifying it with terms like 'must' or 'probably' or 'presumably' or 'may.' But they do so in a minority of cases. If we cannot discover an arguer's intentions in this respect, we must construe the argument as ambiguous and test it against both deductive and inductive (and conductive) standards." (Hitchcock, 2013: 200) Contrastingly, within LNMA's theoretical framework, whether we are conscious of it or not, we always put forward conclusions with a certain degree of epistemic force, because this force is what turns a mere act of claiming that A into an act of concluding that A: this force is essential to any speech-act of concluding as a specific type of constative speech-act. Thus, interpreting a speaker as having drawn a conclusion requires being in a position to determine the specific epistemic modal that she is supposed to be using for so concluding.

## **The evaluation of argumentation: Ontological and epistemic qualifiers**

LNMA follows Toulmin's intuition that modal qualifiers are key to the semantic appraisal of argumentation—that is, the appraisal of argumentation's semantic

conditions, which determine the correctness of a target-claim. Yet, in contrast with Toulmin's model of argument (CITA), LNMA's model of argument incorporates two types of modals: ontological and epistemic.

In everyday discourse, we can make explicit the variety of ways in which we can put forward a certain semantic content  $p$  in a first-order constative speech-act by saying, for example "p is true," "p is (more or less) probable," "p is (more or less) acceptable," "p is (more or less) verisimilar," "p is plausible," "p is necessary," "p is possible," etc. These ontological modals are terms that make explicit the type and degree of pragmatic force of the constatives comprising an act of arguing. When we put forward a propositional content with the appropriate pragmatic force, given the actual state of the world, we make first-order constatives that are semantically correct—like the correct assertions "(it is true that) snow is white", "(it is necessary that) a bachelor is an unmarried man", "(it is possible that) there is life in other planets", etc. Contrastingly, the modal that expresses the pragmatic force with which we draw a conclusion is an epistemic modal. This modal is meant to communicate our credentials for concluding, i.e. the type and degree of support that our reasons are meant to confer on our target-claims because of our inference-claims. For example, in saying that a claim holds truly, necessarily, possibly, plausibly, (more or less) probably, etc. (i.e. in saying things such as "certainly p," "necessarily p," "it might be the case that p," "plausibly p," "(more or less) probably p," etc.), we are expressing something about the status of this claim as knowledge, about the confidence that we may place on it. Thus, any second-order speech-act of concluding involves, either explicitly or implicitly, not only the ontological modal of the first-order constative that it is built on, but also the epistemic modal that indicates the force with which this first-order claim is concluded.

As representations of the inferences that supervene on acts of arguing and acts of reasoning, arguments in LNMA consist of the following elements: premises (corresponding either to the speech-act of adducing, R, or to the cognitive input in the act of reasoning, CI), conclusion (corresponding either to the speech-act of concluding, C, or to the cognitive output in the act of reasoning, CO), warrant (corresponding either to the inference-claim in the act of arguing, IC, or to its counterpart in the act of reasoning; i.e., the inference-motivation, IM) and the representations of the epistemic and ontological modals,  $em$  and  $om$ , of each of the speech-acts making up the act of arguing (corresponding to the type and degree of constative pragmatic force with which the speaker, either implicitly or explicitly, puts forward the propositional content of each constative) or of the judgments and beliefs constituting the act of reasoning (corresponding to the type and degree of assent to each propositional content constituting the act of reasoning). Thus, an ascription of both epistemic and ontological modals (ultimately, the ascription made by the arguer or the reasoner) is part of the layout of arguments, and the semantic appraisal of an act of arguing or reasoning results in the process of determining the right ascription of modals to each represented claim or judgement/belief (i.e., the process of ascertaining whether or not the ascription made by the arguer or the reasoner is correct after all). This model of argument can then be outlined as follows:

( $om_r/ci$ )Premise\_\_\_\_\_therefore\_\_\_\_\_( $em_x$ )( $om_c/co$ ) Conclusion

$$\begin{array}{c} | \\ \text{since} \\ (\text{om}_{ic/im})\text{Warrant: "if } R/CI, \text{ then } C/CO" \end{array}$$

(The contents of the antecedent and the consequent of the warrant correspond to the whole first-order constatives R and C of the act of arguing, i.e. to their propositional contents in conjunction with their (implicit or explicit) ontological modals, or to the whole cognitive input and output, CI and CO of the act of reasoning, i.e. to their propositional contents and their corresponding type and degree of assent).

Let  $\phi$  represent the idiomatic function that, for each ontological modal of a conditional, assigns the epistemic modal needed to draw a conclusion having this conditional as its warrant—or, in other words, the term that is used in a certain language for expressing either the pragmatic force of any speech-act of concluding having a conditional so qualified as its inference-claim or the type and degree of assent to the cognitive output having a conditional so qualified as its inference-motivation.

$$\phi(\text{omi}) = \text{emi}$$

In this account, an argument is valid (i.e., the inference is good, whatever its type) iff  $\text{emi} = \text{emx}$  and  $\text{omi}$  is correct—that is, if it is the ontological modal that actually corresponds to the inference-claim as a constative or to the inference-motivation as a belief or judgement, given the actual state of the world. In other words, an argument is valid if and only if the epistemic modal that the speaker (or reasoner) has used for concluding or coming to believe the cognitive output is the epistemic modal that  $\phi$  assigns to the ontological modal of the speaker's implicit inference-claim or inference-motivation, and this ontological modal is appropriate for this inference claim or inference motivation given the actual state of the world.

## What is deductivism?

Is deductivism the claim that we can make explicit the inference-claim in representing the inferences that supervene in acts of arguing and acts of reasoning? In that case, there would be no disagreement between Botting and me. But, of course, this is trivial. Instead, we should understand our deductivism as the thesis according to which the inference-claim of a piece of argumentation is a premise, so as to contend that: “the only valid arguments are deductively valid arguments, and that deductive logic is the only logic that we have or need. The deductivist ploy regarding so-called non-deductive or inductive or ampliative arguments is to recast them as deductive enthymemes with unstated or missing premises of one kind or another”. (Musgrave 2012: 125)

The key of our proposal is to get rid of regulative rules, which, for any type of normative proposal, pose the justification problem that Putnam pointed out. (Particularly, alternative formal systems of logic pose the problem of what to do in case of contradiction or double negation). Our account deals with inferential

normativity as constitutive of the activity of inferring. Inferring has constitutive correctness conditions, just as believing or asserting. That the normativity of inferring is constitutive of the practice of arguing explains why people are normally good at inferring even knowing nothing about formal logic: learning to infer is eo ipso learning to infer well. This substantial conception of inferential goodness is able to distinguish between “being valid” and “being deductive”: for us, validity is a matter of how the world actually is, not a matter of the alleged form of the argument. Thus, the task of determining the validity of our arguments happens to be crucial for determining their justificatory force. Deductivism takes for granted that any argument is deductive.

At times, we provide non-conclusive reasons for our claims. Deductivism deals with this fact by taking epistemic modals such as “probably”, “presumably”, etc. as being part of the propositional content of the target-claim, instead as making explicit its pragmatic force:

For example, an act of arguing like “John’s car is on the street; presumably, he is at home” would be analyzed as:

P1: John’s car is on the street.

P2: If John’s car is on the street, then presumably he is at home.

C: (necessarily) presumably he is at home.

This strategy makes a mystery of the semantics of epistemic modals and their role in inference.

## Conclusion

In considering the debate about deductivism, we can distinguish between a false controversy and a real controversy:

*False controversy:* can we incorporate inference-claims as conditionals in arguments?

*Real controversy:* is inferential normativity a matter of following regulative rules such as modus ponens?

As I have tried to show, there is no problem in acknowledging that it is harmless to make explicit inference-claims as conditionals when representing the inferences that we make in acts of arguing and acts of reasoning. In fact, that’s what LNMA prescribes for assessing argumentation’s semantic properties. But I think that there are good reasons not to deal with inferential normativity as a matter of rules instead of as a matter of the constitutive correctness conditions of drawing conclusions; and, for this reason, in my view, we should not think of these conditionals as premises and, therefore, we should not think that arguments are either deductive or not arguments at all.

# Teoría de actos de habla y teoría de la argumentación

**Cristina Corredor Lanas**  
Universidad de Valladolid  
corredor@fyl.uva.es

## Introducción

Destacadas aproximaciones teóricas al estudio de la argumentación han dado cuenta de la práctica de argumentar, o de los argumentos, a partir de la noción de acto de habla. Complementariamente, algunas propuestas lingüísticas y filosóficas han estudiado los intercambios comunicativos tomando en consideración su orientación argumentativa, o la forma en que las afirmaciones, o cualquier acto de habla en general, se basa en una precomprensión de la práctica de argumentar, entendida como una práctica de dar y valorar las razones que justifican, o podrían justificar, lo que se dice.

Me interesa avanzar en esta última orientación, tomando como punto de partida la teoría de actos de habla originalmente propuesta por Austin (1962). Los actos de habla judicativos (*veredictives*) consisten en emitir un juicio, estimación o valoración, sobre la base de datos, pruebas o razones. Caracterizar así este tipo de actos de habla conlleva, en mi consideración, asumir una articulación compleja entre su realización y la base justificatoria que los respalda. Mi propósito es argumentar mejor este punto de vista, y explorar en qué medida la misma aproximación podría hacerse extensiva a otros tipos de actos de habla.

## Hablar, inferir, argumentar

Para aproximar el problema de la relación entre el habla (los actos de habla) y la práctica de la argumentación, una primera distinción conceptual pertinente es la que cabe establecer entre inferir y argumentar. Inferimos cuando pasamos de una creencia a otra (en el pensamiento), y el razonamiento consiste en llevar a cabo inferencias. De manera estándar y desde un punto de vista lógico, estas inferencias se representan por medio de implicaciones entre proposiciones (unidades conceptuales con condiciones de verdad completas). La argumentación es una práctica intersubjetiva y social, además de epistémica, donde se aducen o presentan razones para justificar una afirmación de referencia. Siguiendo a Toulmin, la transición desde las razones a la afirmación se legitima por la aplicación de una licencia inferencial o *garantía*, habitualmente tácita o implícita, que así



lo autoriza (Toulmin 2003, p. 91). Mientras que, cuando razonamos, las inferencias que se llevan a cabo (en el pensamiento, o explícitamente en el habla) no necesitan legitimarse de este modo, cuando argumentamos desarrollamos una actividad comunicativa y entramos en interacción con otras personas, cuya valoración o evaluación tiene repercusiones para nuestro propio desempeño. Así, por ejemplo, nuestras razones pueden juzgarse como poco relevantes, o no bien establecidas; nuestro paso de esas razones a la afirmación de referencia puede considerarse sujeto a objeción, o a recusación (*rebuttal* en Toulmin); e incluso el conjunto de la comunicación puede rechazarse como no argumentativa. (Una presentación original, sistemática y ampliamente abarcadora de tipos de contra-argumentación crítica es la que ofrece Hubert Marraud, 2018).

Lo que me interesa resaltar aquí es que estas potenciales dimensiones de crítica conllevan, desde la posición de quien argumenta, la adquisición de determinados compromisos. Desde un punto de vista lógico, se compromete con la corrección, relevancia y fuerza de las razones, así como con la garantía que autoriza la inferencia; desde un punto de vista dialéctico, adquiere la obligación de responder a las críticas, mostrando que el movimiento argumentativo sigue siendo legítimo, o bien a retirar su argumentación (a retractarse). En el ámbito de la atención a los estados psicológicos, que tiene en cuenta el punto de vista retórico, cabe esperar que quien argumenta crea efectivamente que sus razones y su afirmación son correctas, y así mismo que crea que esas razones autorizan a esa afirmación.

La práctica de la argumentación no puede separarse, por tanto, de su desempeño en un ámbito intersubjetivo y social, así como tampoco de la actividad de dar o presentar razones para justificar una afirmación de referencia, sobre la base (posiblemente tácita) de un compromiso con la garantía o licencia inferencial que autoriza a dar ese paso.

## El habla como una actividad inferencial

La idea de que, en el habla, llevamos a cabo y motivamos inferencias, y que estas están conectadas de alguna forma esencial con el modo en que atribuimos significado a nuestras emisiones, es una hipótesis explicativa importantemente presente en la pragmática contemporánea. A partir de Grice, y dentro de la pragmática cognitiva que le sigue, la comunicación se hace posible en la práctica, y se explica teóricamente, como una actividad inferencial, que está además orientada conforme a fines o propósitos. Como es bien conocido, Grice define la noción de *significado del hablante* (*speaker meaning*) como una intención reflexiva y compleja, donde quien habla lo hace con la intención de inducir en su oyente una actitud (con un contenido proposicional); tiene además la intención de que esa primera intención se reconozca por su oyente; y, así mismo, quien habla tiene también la intención de que ese reconocimiento de su primera intención sea el que mueva a su oyente a tener la actitud que tenía la intención de inducir.

Es interesante tener en cuenta que, cuando Grice formula la tercera cláusula de su definición de *speaker meaning*, formula la condición en estos términos: “U intended the fulfillment of the intention mentioned in (2) to be at least in part

A's reason for fulfilling the intention mentioned in (1)" (Grice 1969, p. 153). Es decir, U (quien habla) tenía la intención de que el reconocimiento de su intención de inducir una actitud en A fuera, al menos en parte, la *razón* que tiene A para adoptar esa actitud.

El concepto de razón, o de tener una razón, al que apela Grice aquí no aparece más aclarado en este ensayo. Pero la capacidad de realizar inferencias sí desempeña una función imprescindible en su tratamiento de las implicaturas conversacionales. Pues, en su modelo, el significado pragmático de lo dicho se complementa con un significado comunicado mediante la activación de un proceso inferencial que hace uso del principio cooperativo (aplicable mediante cuatro máximas de cantidad, cualidad, pertinencia y modo).

Si bien los procesos inferenciales de obtención de implicaturas –y, en general, los que permiten a quien escucha recuperar las intenciones comunicativas de quien habla– no necesitan transcurrir en un plano consciente, personal en el sentido de Dennett, Grice creía que al menos las implicaturas conversacionales tenían que ser *calculables*, en el sentido de que fuera posible reconstruir, explícitamente, el proceso inferencial que llevaba desde la información disponible requerida (premisas) hasta el significado comunicado. Si bien quien escucha podría haber recuperado el significado comunicado sin haber recorrido exactamente ese proceso inferencial, el proceso que permite obtenerlo tiene que poder ser “reemplazado por un argumento”: “The presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replaceable by an argument, the implicature (if present at all) will not count as a CONVERSATIONAL implicature” (Grice 1975, p. 50)<sup>1</sup>. Aunque Grice propone este criterio, en principio, como criterio de demarcación para el significado comunicado mediante implicaturas conversacionales, su comprensión del significado que está codificado lingüísticamente (como codificación de las intenciones comunicativas) permitiría generalizarlo al conjunto del significado pragmático, al que adquieren las emisiones en contexto. Importa tener en cuenta que otras propuestas de pragmática cognitiva de carácter neo-griceano o post-griceano (teoría de la relevancia, contextualismo) han dejado de exigir esta calculabilidad, o reconstruibilidad explícita, como criterio. Sin embargo, convierten el recurso a la reconstrucción explícita de procesos inferenciales de obtención de significado explícito en una herramienta teórica para respaldar sus tesis relativas a la recuperación cognitiva del significado no sólo de lo comunicado, sino también de lo dicho o explicatura. Este proceder teórico permite ver sus propuestas en línea con Grice en lo que atañe a la exigencia de que el significado pragmático pueda mostrarse recuperable mediante un proceso inferencial explicitable lingüísticamente.

Lo anterior sugiere una posible hipótesis que conectaría las emisiones en contex-

---

<sup>1</sup>La noción de *argumento* que Grice puede estar teniendo en cuenta es una noción clásica, donde a partir de un conjunto de premisas, y por aplicación de unas reglas, se llega a una conclusión. En relación con el razonamiento, que considera explicitable lingüísticamente, inicialmente propone: “reasoning consists in the entertainment (and often acceptance) in thought or in speech of a set of initial ideas (propositions), together with a sequence of ideas each of which is derivable by an acceptable principle of inference from its predecessors in the set” (Grice 2001, p. 5). Tras la discusión subsiguiente, introduce el matiz de que esos principios o reglas de inferencia pueden ser vistos como “directives (the precise kind of which remains to be determined) observance, or non-violation, of which is a desideratum” (p. 22).

to, empleadas comunicativamente, con su carácter inferencial (eventualmente, explicitable argumentativamente). La hipótesis podría formularse del siguiente modo:

(H1). Lo que hace de la comunicación una actividad inferencial está dado con su calculabilidad: es decir, con la posibilidad de recuperar el significado pragmático (el que resulta comunicado en un contexto particular de uso) a través de una serie de inferencias o un razonamiento explícitos (eventualmente, mediante un acto de argumentación que justificaría la atribución de ese significado a la emisión, mostrando cómo inferirlo a partir de la emisión en contexto).

En contra de lo que pudiera parecer, esta hipótesis podría tener consecuencias que entraran en conflicto con el marco teórico del que parece haberse obtenido. Pues la explicitación del significado, su recuperación mediante un proceso inferencial explícito, es un criterio que no se sitúa ya en el ámbito cognitivo de la persona que habla (de sus actitudes proposicionales), ni tampoco en el ámbito cognitivo de su oyente o auditorio. El requisito sitúa la recuperación del significado de quien habla, y con ello su identificación en cuanto tal, en el ámbito interpersonal y social de lo que puede hacerse explícito mediante actos de razonamiento lingüístico, del tipo de los que podrían constituir, eventualmente, actos de argumentación.

## **El carácter argumentativo del habla**

La hipótesis anterior parecería verse puesta en cuestión de forma inmediata por otras aproximaciones al significado y a la comunicación lingüística. El inferencialismo defiende que el significado de una expresión lingüística puede explicarse teniendo en cuenta su contenido inferencial, es decir, la contribución que esa expresión hace a las relaciones inferenciales de los enunciados en los que aparece. En la versión normativa del inferencialismo debida a Brandom (1995), las relaciones inferenciales de nuestros enunciados, expresados mediante relaciones argumentativas entre afirmaciones, están instituidas por nuestras actitudes normativas (compromisos y autorizaciones). Otra aproximación en aparente conflicto es la teoría de la argumentatividad radical de Anscombe y Ducrot (1976). Para estos autores, las relaciones argumentativas que los enunciados establecen entre sí (y no meramente las emisiones de esos enunciados) no pueden deducirse de su contenido informativo. Por ello, proponen ver esas relaciones como un ‘hecho bruto’ de la lengua, y defienden que el valor semántico de un enunciado consiste en proponer e imponer puntos de vista argumentativos. En ambos casos, tanto en el inferencialismo normativo como en la teoría de la argumentatividad radical, el carácter inferencial o argumentativo de nuestras afirmaciones y enunciados forma parte de la naturaleza misma de su significado o valor semántico. En cierta medida (aunque con importantes diferencias), las aproximaciones teóricas de Brandom y de Anscombe y Ducrot guardan cierta proximidad con la pragmática universal defendida por Habermas (1981). De acuerdo con su tesis, entendemos lo que significa un acto de habla cuando sabemos qué lo hace aceptable, y esta aceptabilidad se puede hacer explícita dando las razones que permitirían justificar su validez. De este modo, el significado de nuestros ac-

tos de habla aparece internamente conectado con un potencial de razones que permitiría justificarlos.

Desde el punto de vista de nuestra hipótesis, sin embargo, cualquier consideración sobre la naturaleza inferencial o argumentativa de nuestros enunciados, afirmaciones y actos de habla es una consideración sobre la forma en que las correspondientes relaciones deberían reconstruirse. Este tipo de reconstrucción está guiada, a su vez, por el esfuerzo de entender y justificar nuestras acciones de habla. Creo que este punto de vista puede encontrar respaldo en una aproximación al habla que tenga en cuenta su dimensión normativa, es decir, el modo en que lo que decimos conlleva determinados compromisos y obligaciones, autorizaciones y derechos, y otras posiciones normativas, que incluyen compromisos y obligaciones justificatorios en relación con el habla misma.

## **Actos de habla y compromisos dialécticos**

De acuerdo con la aproximación Austiniana al habla que considero más acertada (Austin 1962, Sbisà 2006), el significado y la fuerza de nuestros actos de habla puede esclarecerse mostrando de qué manera afectan e introducen cambios en el contexto interpersonal y social de quienes participan en el intercambio comunicativo; estos cambios inciden en las posiciones normativas que esas personas participantes se reconocen y atribuyen mutuamente, y que cabe hacer explícitas en términos de compromisos y obligaciones, autorizaciones y derechos, y otros estados normativos análogos. Así, en el caso particular de una afirmación (en tanto que caso particular de acto de habla veredictivo), quien habla se presenta como cognitivamente competente y adquiere la obligación de ofrecer, si se le piden, las razones que podrían respaldar su afirmación. El reconocimiento otorgado por las personas interlocutoras les permite, a su vez, realizar otras afirmaciones (y otros actos de habla) que tomen como punto de partida su aceptación de la afirmación inicial. Esta articulación dialéctica procede de las posiciones normativas que se reconocen y atribuyen quienes participan en el intercambio. Las razones que podrían aducirse para respaldar la primera afirmación no están dadas, por así decirlo, previamente a que quien la ha hecho tenga que cumplir su deber de justificarla. Es este deber, que tiene en mi consideración un carácter dialéctico, lo que vincula la afirmación con un acto de argumentación subsiguiente. Creo además que este tipo de obligaciones justificatorias, y otros deberes y derechos dialécticos son parte de las posiciones normativas que caracterizan y definen a nuestros actos de habla en general, y no únicamente a los veredictivos. (Mi declaración en este punto, sin embargo, aún ha de verse como provisional y subordinada a que pueda ofrecer un mejor desarrollo y justificación de ella.)

Todo lo anterior sugiere una segunda hipótesis, que complementaría la primera. De acuerdo con esta segunda hipótesis,

(H2) Las posiciones normativas que reconocemos y atribuimos con nuestros actos de habla incluyen deberes y derechos de carácter dialéctico. Este carácter dialéctico del habla no conlleva, sin embargo, que quepa asignar una naturaleza inferencial o argumentativa al significado pragmático o al valor semántico de nuestras emisiones.

## Bibliografía

- Anscombe, J.-C. y O. Ducrot (1976). L'argumentation dans la langue. *Languages*, 10(42), pp. 5–27.
- Austin, J. (1962). *How to do things with words*. Oxford: Clarendon Press.
- Brandom, R. (1995). *Making it explicit*. Cambridge, MA: Harvard University Press.
- Grice, H.P. (1969). Utterer's meaning and intention. *The Philosophical Review*, 78(2), pp. 147–177.
- Grice, H.P. (1975). Logic and conversation. En P. Cole y J. Morgan (eds.), *Syntax and semantics 3: Speech acts*, Nueva York: Academic Press, pp. 41–58.
- Grice, H.P. (2001). *Aspects of reasons*. R. Wagner (ed.), Oxford: Clarendon Press.
- Habermas, J. (1981). *Theorie des kommunikativen Handelns*. Bd. 1. Suhrkamp: Frankfurt am Main.
- Marraud, H. (2018). On the many logical ways to counter an argument. *The 9th Conference of the International Society for the Study of Argumentation*. Disponible en: [https://www.academia.edu/37009245/On\\_the\\_many\\_logical\\_ways\\_to\\_counter\\_an\\_argument](https://www.academia.edu/37009245/On_the_many_logical_ways_to_counter_an_argument) (Último acceso: 25/09/2018)
- Sbisà, M. (2006). Communicating citizenship in verbal interaction. En H. Hausendorf y A. Bora (eds.), *Analysing Citizenship Talk*, pp. 151–180. Amsterdam: John Benjamins.
- Toulmin, S. (2003). *The uses of argument*. (Updated edition, orig. 1958). Cambridge: Cambridge University Press.

# Cómo argumentar con coherencia

**José A. Gascón Salvador**

Centro de Investigación en Educación y Desarrollo (CIEDE)  
Universidad Católica de la Santísima Concepción (UCSC), Chile  
jgascon@ucsc.cl

## Introducción

Como es bien sabido, la lógica informal proporciona tres criterios básicos de evaluación argumentativa. Según fueron propuestos inicialmente, tales criterios son (Johnson y Blair, 1994, pp. 54–55):

**Suficiencia:** las premisas del argumento deben proporcionar un apoyo suficiente a la conclusión.

**Pertinencia:** las premisas del argumento deben ser pertinentes para la conclusión.

**Aceptabilidad:** las premisas deben ser aceptables.

Con leves variaciones, o incluso con la adición de un criterio adicional —el “nivel dialéctico” en el caso de Ralph Johnson (2000)— estos criterios siguen siendo ampliamente aceptados en la actualidad. No obstante, tal como están formulados aquí, tales criterios adolecen de una vaguedad que deja mucho espacio para el desacuerdo sobre si, en un argumento determinado, se cumplen o no. No disponemos aquí de la precisión y objetividad de la noción de consecuencia lógica de los sistemas lógicos deductivos. La comprensión de los criterios de la lógica informal se apoya en gran medida en nuestras intuiciones.

Esta dependencia de nuestras intuiciones puede dar lugar a diferencias considerables sobre qué es aceptable, qué es pertinente y qué es suficiente. Imaginemos, por ejemplo, una discusión entre dos personas sobre si deberían reformarse las leyes relacionadas con el asesinato. Una de esas personas sostiene que debería reformarse porque así lo han pedido los familiares de varias víctimas asesinadas; para la otra persona, en cambio, la opinión de las víctimas *no es pertinente* para la cuestión.

Si es posible llegar a un acuerdo sobre la calidad de un argumento, parece evidente que la solución debe pasar por una *metaargumentación* (Marraud, 2016). Dentro de las posibilidades que ofrece la metaargumentación, aquí me centraré en lo que llamaré *coherencia argumentativa*. Mostraré cómo los argumentadores en ocasiones evalúan la coherencia de sus interlocutores y trataré de dilucidar la noción de coherencia en el ámbito de la argumentación.

## Comprobaciones de coherencia

En las discusiones cotidianas, especialmente cuando se trata de temas controvertidos o de posturas de difícil reconciliación, los hablantes suelen comprobar la coherencia de las opiniones de sus interlocutores. Puede tratarse de una búsqueda de incoherencias entre enunciados –por ejemplo, si alguien sostiene que mentir *siempre* está mal y, al mismo tiempo, que las normas de cortesía hacen que sea legítimo mentir en ocasiones–. Pero aquí me centraré en un tipo de coherencia que atañe específicamente al campo de la argumentación: la coherencia entre argumentos. Creo que pueden distinguirse dos casos: las acusaciones de incoherencia en lo referente a las evaluaciones argumentativas de una persona, y la búsqueda de coherencia entre un argumento que una persona ha defendido y otro argumento hipotético, que se realiza por medio de analogías.

## Acusaciones de incoherencia

La coherencia argumentativa exige que, si se considera que determinadas razones apoyan una afirmación, entonces, en circunstancias similares, debe considerarse que esas mismas razones (u otras análogas) apoyan una afirmación análoga. De este modo, si un argumentador acepta una afirmación sobre la base de unas razones, pero no acepta otra afirmación análoga que es sustentada por las mismas razones, puede ser acusado de incoherencia. Veamos un ejemplo que nos ofrece un usuario de Twitter<sup>1</sup>:

Os preocupan los manteros porque “dañan al pequeño comercio” vendiendo falsificaciones de Tous.

No como Starbucks, McDonald’s, Mercadona, Carrefour, etc que no hacen ningún daño al pequeño comercio del barrio.

En este ejemplo se sostiene que, si los perjuicios al pequeño comercio constituyen una razón pertinente y suficiente para oponerse a la actividad de los manteros, entonces esos perjuicios también deben reconocerse como una razón pertinente y suficiente para oponerse a los grandes comercios. Si alguien acepta el primer argumento pero no acepta el segundo, incurre en una incoherencia.

A menudo, las acusaciones de incoherencia se producen porque un argumentador acepta una fuente de información en ciertos casos pero la considera poco fiable en otros –en función de si está de acuerdo con lo que afirma dicha fuente–. Esto es lo que sucede en el siguiente hilo de Twitter, en el que un usuario denuncia la incoherencia con la que se recibe la información que nos proporciona la comunidad científica<sup>2</sup>:

Cuando la comunidad científica avisa sobre un riesgo, peligro, catástrofe, tiene credibilidad y es neutra\*: tabaco, cambio climático...

Cuando la comunidad científica no encuentra riesgo ni peligro, carece de credibilidad y está comprada: transgénicos, glifosato, wifi...

---

<sup>1</sup><https://twitter.com/Vicariusx/status/1028283413697110016>

<sup>2</sup><https://twitter.com/paleofreak/status/1028944075553206274>

\*Excepto si es [*sic*] riesgo o peligro está en algo cotidiano que nos gusta: alcohol, exceso de azúcar, exposición al sol... entonces la comunidad científica bah, cada día dicen una cosa distinta.

(Esto pasó también con el tabaco, al principio)

Si la comunidad científica dice algo que coincide con nuestro pensamiento, joder, es la ciencia. Si dice lo contrario, está vendida, o es patriarcal, o cuidado, que vienen otra vez los nazis, o tiene sesgos (más que tú), o sirve al sistema capitalista.

Naturalmente, no se quiere decir con esto que, si una fuente de información se considera fiable, entonces necesariamente deben aceptarse *todas* sus afirmaciones. Puede darse el caso de que una autoridad epistémica afirme algo que quede fuera de su campo de conocimiento o que entre en conflicto con otros conocimientos bien establecidos. Pero, en nuestro ejemplo, el usuario de Twitter critica que en determinadas ocasiones se considera que la comunidad científica “tiene credibilidad y es neutra”, mientras que en otras ocasiones “carece de credibilidad y está comprada”. Es decir, la incoherencia que se señala es que la evaluación de la credibilidad de la fuente es *diferente* en casos distintos, a pesar de tratarse de *la misma* fuente.

## Analogías

Las analogías son unos tipos especiales de argumentos. Como han mostrado varios teóricos, recurrir a una analogía para defender un punto de vista implica recurrir a la metaargumentación. Al usar una analogía para argumentar, se presenta un argumento que es similar a otro en el sentido de que, si se acepta aquel, debe aceptarse también el otro. De acuerdo con Woods y Hudak<sup>3</sup> (1989, p. 127):

Los argumentos de analogía son argumentos de lo que se conoce como paridad de razones. Son argumentos sobre argumentos, *metaargumentos*. En ellos se argumenta que dos o más argumentos diana se sostienen o se vienen abajo conjuntamente y que esto es así porque están en paridad de manera relevante, que poseen estructuras profundas similares en virtud de las cuales su forma lógica coincide.

Hubert Marraud (2007) propone un análisis similar de los argumentos que se basan en analogías. Según explica, en la argumentación por analogía se transfiere un argumento desde un campo a otro que es análogo al anterior. Quien argumenta por analogía propone “inferir la corrección de un argumento de la corrección de otro argumento”, con la pretensión de que “no puede tenerse por bueno el argumento origen sin hacer lo propio con el argumento término” (p. 182).

Mi propuesta es que presentar una analogía consiste en presentar un argumento que *presumiblemente* el interlocutor aceptará (o rechazará), con el fin de hacerle ver que, *por una cuestión de coherencia*, también debería aceptar otro argumento que se le ha presentado (o rechazar otro argumento que el interlocutor ha

---

<sup>3</sup>Cuando se citan textos en inglés, la traducción es mía.



defendido). Si la analogía es plausible y el interlocutor mantiene su rechazo (o defensa) del argumento, se expone a acusaciones de incoherencia.

Veamos algún ejemplo. El siguiente argumento aparece en un artículo de *Scientific American* sobre la hipótesis de que el cerebro humano funciona de acuerdo con el teorema de Bayes<sup>4</sup>:

Además, la tesis del cerebro bayesiano puede reducirse a un silogismo dudoso: Nuestros cerebros destacan en ciertas tareas. Los programas bayesianos destacan en tareas similares. Por lo tanto nuestros cerebros emplean programas bayesianos.

Existen unos límites obvios para esta lógica. Los halcones peregrinos destacan en el vuelo, como también los reactores F15. Nadie afirma que, por lo tanto, los halcones peregrinos deben de emplear la propulsión a chorro, porque hasta el más tonto puede ver que la mecánica del halcón peregrino y de la propulsión a chorro son totalmente diferentes.

El autor del artículo compara dos argumentos que, según él, siguen la misma lógica. El primero de ellos es el que sostienen sus oponentes: puesto que nuestros cerebros y los programas bayesianos destacan en tareas similares, nuestros cerebros y los programas bayesianos emplean los mismos mecanismos. Al autor no le convence este argumento, presumiblemente porque las razones no son suficientes. Para mostrar a sus adversarios esta insuficiencia, lo compara con otro argumento que tiene una estructura similar y que indudablemente suscitará menos controversia: puesto que los halcones peregrinos y los reactores F15 destacan en la misma tarea (volar), los halcones peregrinos y los F15 emplean el mismo mecanismo. El autor espera que sus oponentes capten la (evidente) debilidad de este argumento y ello les lleve, por una cuestión de coherencia, a concluir que el primer argumento es igualmente débil.

Veamos otro ejemplo. En *The Enigma of Reason*, Hugo Mercier y Dan Sperber hacen el siguiente comentario (Mercier y Sperber, 2017, p. 49):

La trompa del elefante es un tipo de nariz. Por muy impresionante que sea, no tendría sentido considerarla como el prototipo de las narices. De manera similar, la razón es un tipo de mecanismo de inferencia; no es el mejor ni el modelo de todos los demás.

En esta analogía, Mercier y Sperber se enfrentan al argumento de que, puesto que la razón es un impresionante mecanismo de inferencia, entonces debe ser considerada como el mejor y más perfecto. Para convencernos de que la capacidad de impresionarnos no es una razón suficiente para concluir que un objeto es el mejor de su clase, nos muestran un argumento análogo cuya insuficiencia todos podemos reconocer. Tal argumento sería: la trompa del elefante es un tipo de nariz impresionante, luego la trompa del elefante debe considerarse como el prototipo de las narices. Al percibir la debilidad de este argumento y la similitud entre los dos argumentos, podemos convencernos de que el argumento sobre la razón es igualmente débil.

---

<sup>4</sup><https://blogs.scientificamerican.com/cross-check/are-brains-bayesian/>

## ¿Qué es la coherencia?

Solo tiene sentido hablar de la coherencia en la argumentación como algo deseable si aceptamos que nuestros argumentos deben fundamentarse en algún tipo de principios generales de razonamiento. En la argumentación epistémica, la existencia de leyes y de relaciones causales es necesaria para que exista conocimiento; en la argumentación práctica, hay una concepción de la racionalidad que exige la generalidad. Las anomalías pueden entenderse, de este modo, como desviaciones de tales principios generales.

Para ver cómo pueden aplicarse los conceptos que hemos visto a la argumentación, usemos el siguiente ejemplo simple:

María: Tienes que lavar el coche esta semana.

Juan: ¿Por qué?

María: La semana pasada lo hice yo. Acordamos que una semana cada uno.

Juan: Pero yo he tenido una semana muy dura. No es justo.

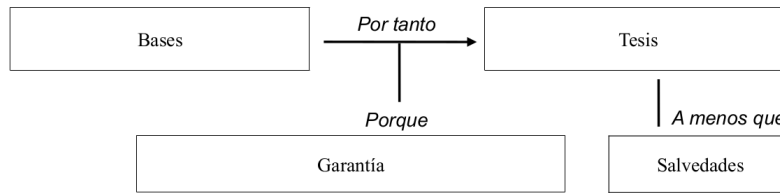
Cuando María argumenta ante Juan que él debe ser quien lave el coche esta semana porque la semana pasada lo hizo ella, su argumento se apoya en lo que podríamos llamar un principio. Tal principio, acordado por ambos, establece que deben alternarse cada semana para lavar el coche. Juan, sin embargo, presenta lo que podría considerarse como una excepción al principio acordado y que se aplica en el presente caso: ha tenido una semana muy dura. Pero, podríamos preguntarnos: ¿es generalizable esta diferencia? Generalizarlo a argumentos análogos supondría considerar que hay una razón para incumplir los acuerdos si alguna de las partes ha tenido una semana muy dura o está cansada por algún otro motivo. Parece difícil sostener que tal generalización es plausible. La excepción que Juan introduce es, pues, una anomalía, y por ello incurre en una incoherencia.

Para incorporar estas consideraciones a la teoría de la argumentación, podemos basarnos en el popular modelo propuesto por Stephen Toulmin. Según este modelo, unas *bases* (o razones) apoyan una *tesis* por medio de una *garantía*, que es la que autoriza el paso de aquellas a esta. A diferencia de lo que ocurre en la lógica deductiva, el paso de las bases a la tesis no es necesario sino plausible, así que puede haber excepciones. Tales excepciones –que habitualmente se introducen con la expresión “a menos que”– pueden incorporarse en el argumento como *salvedades*<sup>5</sup>. De este modo, los argumentos pueden representarse de la siguiente forma:

Las garantías son enunciados generales que explican el vínculo entre las bases y la tesis; o, como las presentan los propios autores, son “enunciados de unión” que “llaman la atención sobre las formas generales de argumentar” (Toulmin, Rieke y Janik, 2018, p. 79). Son reglas generales. Así pues, son muy similares a lo que

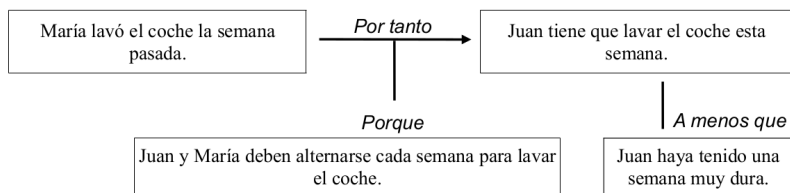
---

<sup>5</sup>El modelo de Toulmin incluye también un respaldo que fundamenta la garantía y un calificador que explicita la confianza que tiene el argumentador en la tesis. Omito estos elementos porque no desempeñan ningún papel en mi análisis de la coherencia. Véanse Toulmin (2007) y Toulmin, Rieke y Janik (2018).



antes he llamado “principios”. Las garantías deben servir no solo para sostener el argumento actual sino también otros argumentos. En uno de los ejemplos de la sección anterior, se denunciaba una incoherencia entre quienes en unas ocasiones se fiaban de la comunidad científica pero en otros casos afirmaban que carece de credibilidad. Ahora podemos explicar tal incoherencia como la aceptación, por parte de una misma persona, de dos argumentos con garantías incompatibles: “La comunidad científica es fiable” y “La comunidad científica carece de credibilidad”. Igualmente, en el caso de las analogías que vimos, lo que explica que dos argumentos análogos “se sostienen o se vienen abajo conjuntamente” es que poseen una garantía similar (Marraud, 2013, p. 200). Si la garantía se acepta en uno de los argumentos, entonces, por coherencia, debe aceptarse también la garantía en el segundo.

Las salvedades son necesarias porque habitualmente las garantías no son reglas universales sino que tienen excepciones. Lo importante es que, cuando un argumento con una garantía contiene ciertas salvedades, otros argumentos con la misma garantía también contendrán esas salvedades. El problema, lo que propongo aquí llamar “anomalía”, surge cuando se presentan en un argumento excepciones que no son generalizables a argumentos análogos. Hemos visto ya un caso de esto en el ejemplo que introducía esta sección. La respuesta de Juan era incoherente porque resulta poco plausible pensar que la salvedad del siguiente argumento es aceptable:



Este argumento no parece aceptable porque es poco probable que Juan y María acordasen en su momento esa salvedad. Y, más en general, el cumplimiento de los acuerdos no suele estar sujeto a que ninguna de las partes esté cansada en el momento de cumplir con sus obligaciones. La excepción que plantea Juan no es, por tanto, generalizable como una salvedad, así que se trata de una anomalía.

Puede decirse que una mayor generalidad de la garantía y de las salvedades conduce a una mayor coherencia en la evaluación de la pertinencia y la suficiencia de las razones de un argumento. A la inversa, cuando un argumentador solo acepta garantías o salvedades muy específicas que solo tienen validez para un número escaso de argumentos —o, en el peor de los casos, solo un argumento—, es por ello más incoherente. La coherencia es, por tanto, una noción gradual.

En conclusión, con el fin de proporcionar una orientación para aspirar a una mayor coherencia en la argumentación, propongo las siguientes directrices para la coherencia argumentativa:

1. Un argumentador es más coherente en la medida en que acepte un mayor número de argumentos sobre la base de la misma garantía o garantías similares.
2. Un argumentador es menos coherente en la medida en que acepte un menor número de argumentos con la misma garantía o garantías similares, o en la medida en que acepte un mayor número de garantías diferentes para argumentos análogos.
3. Un argumentador es más coherente en la medida en que considere las mismas salvedades en argumentos con la misma garantía o garantías similares.
4. Un argumentador es menos coherente en la medida en que considere diferentes salvedades en argumentos con la misma garantía o garantías similares.

## Referencias

- Johnson, R. H. (2000). *Manifest rationality: A pragmatic theory of argument*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Johnson, R. H., y Blair, J. A. (1994). *Logical self-defense*. New York: McGraw-Hill.
- Marraud, H. (2007). La analogía como transferencia. *Theoria*, 22(2), 167–188.
- Marraud, H. (2013). *¿Es lógic@? Análisis y evaluación de argumentos*. Madrid: Cátedra.
- Marraud, H. (2016). Usos y abusos de la metaargumentación. *Quadripartita Ratio*, 1(1), 90–99.
- Mercier, H., y Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.
- Toulmin, S. (2007). *Los usos de la argumentación*. Barcelona: Península.
- Toulmin, S., Rieke, R., y Janik, A. (2018). *Una introducción al razonamiento*. Lima: Palestra.
- Woods, J., y Hudak, B. (1989). By parity of reasoning. *Informal Logic*, 11(3), 125–139.

# Releyendo a Kuhn desde la teoría de la argumentación

**Paula Olmos Gómez**

Universidad Autónoma de Madrid

paula.olmos@uam.es

El texto de Thomas Kuhn “Objectivity, Value Judgment, and Theory Choice” (incluido en *The essential Tension*, 1977) continúa siendo, hoy en día, una referencia obligada en los debates filosóficos sobre la justificación de las conclusiones y resultados científicos. Kuhn señala allí una serie de cuestiones problemáticas en torno a tales procesos justificativos que, 40 años después, estamos en mejores condiciones de comprender y abordar gracias al desarrollo de los estudios de argumentación.

En dicho trabajo, Kuhn pretendía continuar una línea de respuesta a sus críticos que ya había ido esbozando en al menos dos textos anteriores: el “Postscript”, fechado en 1969, e incluido en la segunda edición de *La estructura de la Revolución Científica* (1970a) y las “Reflections on my Critics”, publicadas en el libro colectivo, editado por I. Lakatos y A. Musgrave, *Criticism and the Growth of Knowledge* (1970b). Kuhn intentó acallar las acusaciones de relativismo y psicologismo subjetivista de quienes leían escandalizados algunos de sus pronunciamientos en *La estructura*, como que los científicos utilizaban, en defensa de sus teorías: “técnicas de persuasión, o argumentos y contraargumentos en situaciones en las que no puede haber pruebas” (1970a, 151–152), entendiendo por pruebas, exclusivamente, la contrastación experimental de las consecuencias observacionales supuestamente “implicadas” por tales teorías que, según un conocido resultado, en todo caso, las *infradeterminarían* (Cf. Stanford, 2017).

El intento consistía en tratar de concretar qué tipo de argumentos y contraargumentos eran comunes en la ciencia y cómo funcionaban para justificar la elección teórica, poniendo de manifiesto hasta qué punto se mantenían, con ello, las pretensiones de racionalidad y objetividad. Pero es cierto que para ello tendría que hacer un esfuerzo de dilucidación conceptual de tipo disociativo (Cf. Tindale 2010) a partir del cual las nociones de *justificación racional* y de *objetividad* adquirirían un nuevo sentido que no acabó de tener la recepción deseada.

Creo que así puede entenderse la reseña crítica que H. Siegel publicó en 1980, en principio sobre el libro completo de *La tensión esencial*, pero centrada básicamente en el artículo que nos ocupa. Siegel muestra lo que podría parecer inicialmente un amplio acuerdo con Kuhn. El siguiente párrafo del “Postscript”, que contiene ya gran parte del planteamiento de “Objectivity...” le parece, por ejemplo, a Siegel “inobjetable” (1980, 364):

Esta tesis relativamente familiar [i.e. la ausencia de un algoritmo de decisión en la elección entre teorías científicas] no implica en absoluto ni que no haya buenas razones para sentirse persuadido ni que

estas no sean suficientemente concluyentes para la comunidad científica. Ni tampoco implica que las razones que sustentan la elección sean diferentes de las que suelen mencionar los filósofos de la ciencia: precisión, simplicidad, fecundidad y otras semejantes. Lo que sugiere dicha tesis, sin embargo, es que tales razones funcionan como valores y por ello pueden aplicarse de manera discrepante, tanto individual como colectivamente, por personas que coincidan en aceptarlas como pertinentes. Si dos científicos discrepan, por ejemplo, sobre la fecundidad de sus respectivas teorías, o si coinciden sobre este punto pero discrepan sobre el peso relativo que tal fecundidad debe tener frente a, digamos, el alcance [otro criterio] en la elección teórica, no podrá acusarse a ninguno de estar equivocado. Ni podrá decirse de ninguno de ellos que se está comportando de manera no científica. No existe un algoritmo neutral de elección teórica, un procedimiento de decisión sistemático que, aplicado correctamente, lleve a todos y cada uno de los individuos del grupo a adoptar la misma decisión. (Kuhn, 1970a, 199–200)

Sin embargo, Siegel le reprocha a Kuhn el que, al contrario que su maestro I. Scheffler (1967) no trate de *solucionar* de una manera más definitiva este *problema* (como la amenaza que supone a la racionalidad de la justificación en la ciencia), concretando los “meta-criterios” (deseablemente “externos” a los paradigmas) que permitirían superar las discrepancias en las “deliberaciones científicas” y alcanzar consensos presentablemente objetivos. Le reprocha, en particular, que siga hablando de *incommensurabilidad*, habiendo precisamente puesto las bases para comprender mejor la justificación de las elecciones teóricas y también que insista en el vínculo personal (asociado a una historia y una personalidad concretas) entre los individuos y sus criterios (sus razones, sus argumentos, sus ponderaciones). El veredicto de Siegel es que Kuhn “no puede sostener ambas tesis” (sería inconsistente): que hay *modos racionales* de argumentar en ausencia de pruebas definitivas o algoritmos sistemáticos y que, sin embargo, persisten las brechas infranqueables, los mundos diversos, las *incommensurabilidades*.

Aunque a mí tampoco me resulta convincente lo que podríamos llamar la “versión paralizante” de la *incommensurabilidad* y ya he tratado de señalar en otras ocasiones cómo la continuidad de la argumentación y la meta-argumentación (en la línea señalada por Siegel y Scheffler) pueden hacer avanzar una discusión en torno a posturas radicalmente enfrentadas (Olmos 2016), también simpatizo con la reticencia de Kuhn a asumir una lectura demasiado simplista u optimista (en el sentido de *clausurable*) de su planteamiento. Aunque Kuhn (sin contar, por otra parte, con el vocabulario de la teoría de la argumentación actual) da pistas de por dónde van los procesos de justificación, no puede entenderse su propuesta como la descripción (convenientemente compleja) de una “metodología” tradicional *completa* (como prácticamente hace R. Nola 2000) ni tiene realmente sentido una reescritura de su propuesta en términos bayesianos (como las que intentan Salmon o Earman, Cf. Farmakis 2008).

Un elemento importante de la originalidad, hasta ahora no asumida, de Kuhn está en que trata de hacer ver la *racionalidad* de los procesos de elección teórica no ya en ausencia de una metodología algorítmica de decisión sino precisamen-

te gracias a tal ausencia. Y con ello revela una concepción de la *justificación racional* que difiere de la de sus críticos, incluso de los más comprensivos. Una concepción más cercana a la defendida por J. Searle (2001, cap. 3) quien caracteriza la “racionalidad” como el ejercicio consciente del ofrecimiento y la captación de razones que se realiza *precisamente* en las brechas (*gaps*) de automatismos, algoritmos y supuestos “vínculos de necesidad”. No seríamos racionales por contar con algoritmos de decisión y seguirlos sino por ser capaces de tomar decisiones razonadas (y justificarlas ante los demás) en ausencia de ellos. La revisabilidad, no monotonía y fragilidad de las relaciones de consecuencia, la posibilidad de poder seguir “pidiendo razones” que refuercen o clarifiquen los vínculos propuestos entre razones y conclusiones, o muestren otros modos alternativos de apoyarlas, la falta de compulsión en la aceptación y en la puesta en práctica (*akrasia*) de los contenidos de nuestras propias conclusiones serían los *tipos de brechas* que, para Searle, se erigen en *condición constitutiva de la “racionalidad”*, en lugar de ser un problema para la comprensión de la misma.

Como es sabido, Kuhn presenta en “Objectivity...” cinco criterios básicos, que ya mencionaba en el penúltimo capítulo de *La estructura*, características *deseables* que deben tener las teorías científicas y que pueden ser *fuentes de argumentos* a su favor. Una teoría científica debe ser:

1. **Precisa** (*Accurate*): i.e. empíricamente adecuada (especialmente en sus predicciones) de acuerdo con la observación y la experimentación. Parece el criterio más extendido entre los filósofos y muchas veces el *único* tenido en cuenta en planteamientos tradicionales de acuerdo con el principio del Empirismo del Conocimiento: “si los datos empíricos no son suficientes para determinar el valor de verdad de una teoría, nada puede determinarlo”.
2. **Coherente** (*Consistent*): tanto internamente (como *ausencia de contradicción* interna) como externamente (con vistas a una posible unificación con otras teorías). Se trata, en todo caso, de un criterio más complejo que su correlato lógico.
3. **Amplia** (*Broad Scope*): las consecuencias de una teoría deben poder extenderse más allá de los fenómenos para cuya explicación se concibió (con idea de evitar teorías *ad hoc*). De nuevo se trata de un *desideratum* propio del ideal de la unificación científica: evitar la fragmentación y multiplicación de principios explicativos.
4. **Simple** (*Simple*): que ofrezca las explicaciones más simples posibles (navaja de Occam).
5. **Fecunda** (*Fruitful*): debe “revelar” nuevos fenómenos, ampliar el campo de estudio, abrir nuevas perspectivas de investigación.

Lo primero que habría que decir –algo que se asume aunque no se hace explícito en el texto de Kuhn aunque se infiere de los ejemplos aducidos– es que no tiene sentido aplicar de manera “cualitativa” tales criterios, siempre estaremos en una situación “comparativa”. De lo que se trata es de que la teoría elegida sea *más* precisa, coherente, amplia, simple o fecunda que la(s) alternativa(s), aunque además se pueda señalar un “umbral de cumplimiento” mínimo (sobre todo en los dos primeros) para que las teorías entren en liza y sean susceptibles de comparación.

Por otro lado, Kuhn no se limita a presentar su listado de criterios (que explícitamente califica de *no exhaustivo*), sino que comenta algunos aspectos problemáticos de su uso, mencionando los dos *escollos* que aparecían ya en el “Postscript” para que se considere que, con su concurso, se *cierra la brecha* abierta por la *infradeterminación*:

- a) Los criterios son imprecisos, se desdoblan en posibles modos de cumplimiento, no son independientes entre sí; es decir, son de *difícil* aplicación.
- b) Su aplicación en todo caso, puede lugar a conclusiones contradictorias: “entran en conflicto los unos con los otros”.

Sorprendentemente, Kuhn dice que estos problemas son “familiares”, aunque lo cierto es que son problemas que no han sido abordados convenientemente hasta el desarrollo de la teoría de la argumentación. Respecto de a), podríamos decir que las dificultades de aplicación de los criterios, la falta de una *regla precisa, unívoca* para atribuir a una teoría tales características en mayor grado que a otra, convierte el acuerdo sobre su atribución en una situación argumentativa: es decir, hay que dar *razones* (sustantivas y no concluyentes, basadas en *garantías*) para su adjudicación, lo que puede llevar a controversias entre posiciones razonables. El ejemplo de Kuhn:

|                                                                                             |                                                                                                                                                                                     |
|---------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| La precisión de una teoría se mide por la extensión de su capacidad predictiva/explicativa: | La teoría del oxígeno predice correctamente (explica) las relaciones de peso. La teoría del flogisto no predice correctamente (explica) las relaciones de peso.                     |
|                                                                                             | Por tanto                                                                                                                                                                           |
|                                                                                             | La teoría del oxígeno es más precisa que la del flogisto.                                                                                                                           |
| La precisión de una teoría se mide por la extensión de su capacidad predictiva/explicativa: | La teoría del flogisto predice correctamente (explica) las semejanzas entre los metales. La teoría del oxígeno no predice correctamente (explica) las semejanzas entre los metales. |
|                                                                                             | Por tanto                                                                                                                                                                           |
|                                                                                             | La teoría del flogisto es más precisa que la del oxígeno.                                                                                                                           |

En una situación así, la racionalidad se sitúa a ambos lados de la discrepancia y sobre todo en la capacidad para seguir dando razones que permitan la elección. Kuhn habla de elegir la mayor significatividad de un área de precisión respecto de otra. Podríamos pensar en otras razones (alternativas, complementarias) como la representada en el diagrama:



|                                                        |                                                                                                                                       |                                                                                   |   |                                                                                         |                                                                                 |
|--------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|---|-----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
|                                                        | Las relaciones de peso entre componentes químicos son medibles.<br>Las semejanzas en comportamiento son características cualitativas. |                                                                                   |   |                                                                                         |                                                                                 |
| Las magnitudes caracterizan a la ciencia más avanzada: | Por tanto                                                                                                                             |                                                                                   |   |                                                                                         |                                                                                 |
|                                                        |                                                                                                                                       | La teoría del oxígeno predice las relaciones de peso y la teoría del flogisto no. |   | La teoría del flogisto predice las semejanzas entre metales y la teoría del oxígeno no. |                                                                                 |
|                                                        | La precisión de una teoría se mide por la extensión de su capacidad predictiva:                                                       | Por tanto                                                                         | > | Por tanto                                                                               | La precisión de una teoría se mide por la extensión de su capacidad predictiva: |
|                                                        |                                                                                                                                       | La teoría del flogisto es más precisa que la del oxígeno.                         |   | La teoría del flogisto es más precisa que la del oxígeno.                               |                                                                                 |
|                                                        | Por tanto                                                                                                                             |                                                                                   |   |                                                                                         |                                                                                 |
|                                                        | Debo elegir la teoría del oxígeno.                                                                                                    |                                                                                   |   |                                                                                         |                                                                                 |

Respecto de b), la situación es parecida, aunque el modo de funcionamiento de los criterios es aún más complejo al tener que ponderar argumentos menos semejantes entre sí, aduciendo razones también de mayor complejidad atributiva. Al ponderar argumentos con garantías muy distintas, lo que se compara, fundamentalmente, son dichas garantías. El ejemplo de Kuhn (teorías de Copérnico y Tolomeo) revela, además otras complejidades:

|                                                                                                              |                                                                                                                                                            |
|--------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
| La coherencia externa de una teoría se atribuye en función de la compatibilidad con otras teorías aceptadas: | La teoría geocéntrica (Tierra estacionaria) es coherente con la percepción humana y la física recibida. La teoría heliocéntrica (Tierra en movimiento) no. |
|                                                                                                              | Por tanto                                                                                                                                                  |
|                                                                                                              | La teoría geocéntrica tiene mayor coherencia externa que la heliocéntrica.                                                                                 |

En el caso de la sencillez de ambas teorías, podríamos no estar de acuerdo siquiera en su valoración, pues habría varios sentidos de sencillez, varias razones de su atribución:

|                                                                                 |                                                                                                                                              |
|---------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| La simplicidad se atribuye en función de la dificultad de cálculo:              | La teoría heliocéntrica y la geocéntrica permiten calcular con el mismo “esfuerzo de cálculo” las posiciones de los planetas.                |
|                                                                                 | Por tanto                                                                                                                                    |
|                                                                                 | La teoría geocéntrica y la heliocéntrica son igual de simples.                                                                               |
| La simplicidad se atribuye en función de la sencillez de la geometría manejada: | La teoría heliocéntrica asume una trayectoria planetaria circular, la geocéntrica, una trayectoria definida por varios círculos (epiciclos). |
|                                                                                 | Por tanto                                                                                                                                    |
|                                                                                 | La teoría heliocéntrica es más simple que la geocéntrica.                                                                                    |

La elección de la segunda opción supone, por cierto, el uso de un meta-criterio opuesto al manejado en el caso del oxígeno y el flogisto: en este caso se considerará más significativo un tipo de sencillez cualitativo que uno cuantitativo. En todo caso, la elección de Kepler y Galileo por el sistema heliocéntrico (elección personal y situada), parece poner en funcionamiento otros criterios de la lista de Kuhn (ver página siguiente).

Kuhn no utiliza ni el vocabulario ni los instrumentos de análisis de la teoría de la argumentación. Tan solo apunta que los criterios de elección funcionan como valores y no como reglas. Con ello resume características como las que hemos mostrado en nuestros análisis:

- El carácter debatible/justificable de su atribución comparativa: no solo basada en garantías sustantivas y no formales, sino además con posibilidades discutibles en cuanto a la propia garantía elegible.
- La necesidad de ponderar su contribución multidimensional al juicio electivo. Necesidad de ordenar (en un momento dado) los criterios y ponderar su significación conforme a nuevos criterios.

Si los criterios son *valores*, su cumplimiento no es una cuestión sujeta a un juicio bivalente sino que se sustancian en *preceptos de maximización* (algo que conocen bien los filósofos del derecho). Si además hay *varios valores en juego*, susceptibles de dar lugar a atribuciones comparativas divergentes, la complejidad de la argumentación en torno a ellos aumenta, sin que podamos definir una fórmula universal de orden, válida para todas las circunstancias y para todos los participantes en la discusión.

Ello ayuda a comprender el carácter insoslayable de los *agentes*, su conocimiento personal basado en sus experiencias (Polanyi 1958) y las circunstancias en las que se da la argumentación en el *reconocimiento* y la *aceptación diferencial* de razones. Como apunta J. Woods (2016), no estamos en un espacio impersonal de *relaciones lógicas* (basadas en características formales) sino en un espacio (que para él es *psicológico* y para otros, entre los que me encuentro, *público e intersubjetivo*) de *razones sustantivas* (basadas en garantías) que caracteriza la *racionalidad justificativa* entendida al modo de Kuhn, Searle y gran parte de los teóricos de la argumentación.

|                                                                   |                                                                                                                                                                  |                                                                                                                                                            |   |                                                                                                                                             |                                                                                 |
|-------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|---|---------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
|                                                                   | La simplicidad geométrica del universo permitirá el desarrollo de una astrofísica avanzada y la comprensión de los fenómenos terrestres en un marco más general. |                                                                                                                                                            |   |                                                                                                                                             |                                                                                 |
| La mayor amplitud y fecundidad de una teoría la hacen preferible: | Por tanto                                                                                                                                                        |                                                                                                                                                            |   |                                                                                                                                             |                                                                                 |
|                                                                   |                                                                                                                                                                  | La teoría geocéntrica (Tierra estacionaria) es coherente con la percepción humana y la física recibida. La teoría heliocéntrica (Tierra en movimiento) no. |   | La teoría heliocéntrica asume una trayectoria planetaria circular, la geocéntrica, una trayectoria definida por varios círculos (epiciclos) |                                                                                 |
|                                                                   | La coherencia externa de una teoría se atribuye en función de la compatibilidad con otras teorías aceptadas:                                                     | Por tanto                                                                                                                                                  | < | Por tanto                                                                                                                                   | La simplicidad se atribuye en función de la sencillez de la geometría manejada: |
|                                                                   |                                                                                                                                                                  | La teoría geocéntrica tiene mayor coherencia externa que la heliocéntrica.                                                                                 |   | La teoría heliocéntrica es más simple que la geocéntrica.                                                                                   |                                                                                 |
|                                                                   | Por tanto                                                                                                                                                        |                                                                                                                                                            |   |                                                                                                                                             |                                                                                 |
|                                                                   | Debemos elegir [hablan Kepler o Galileo] la teoría heliocéntrica.                                                                                                |                                                                                                                                                            |   |                                                                                                                                             |                                                                                 |

La dinámica racional de la ciencia, así entendida, más que pretender una “objetividad impersonal” se basaría en una “subjetividad justificable o enjuiciable” (*judgmental*) que para Kuhn es la que caracterizaría las discusiones científicas.

El carácter personal aunque enjuiciable (i.e. asociado a agentes racionales y argumentadores) de las atribuciones y evaluaciones de las teorías en relación con los criterios determina:

- la posibilidad razonable de que haya discrepancias (en distintos niveles argumentativos) que es lo que hemos desarrollado aquí, pero también,
- la constitución de la(s) ciencia(s) misma como una empresa racional y dinámica, en lugar de algorítmica y, por ello, sustraída a la discusión, lo que haría que *todos los científicos tomaran la misma decisión a la vez*, sin permitir que nuevas investigaciones mejorasen o empeorasen el estatus de las teorías propuestas.

- c) Finalmente, Kuhn apunta también (un tema que cabría explorar) que es precisamente el *carácter personal* del ejercicio del reconocimiento y concesión de razones y su asociación con experiencias y prácticas humanas (que podrían compartirse en un modo tácito y participativo) lo que puede lograr una cierta superación (no explicitable en términos proposicionales) de los problemas de *incommensurabilidad*.

## Referencias

- Farmakis, L. (2008). “Did Tom Kuhn actually meet Tom Bayes?” *Erkenntnis*, 68:41–53.
- Kuhn, T. (1970a). *The structure of Scientific Revolutions* (2nd ed.). Chicago: University of Chicago Press.
- Kuhn, T. (1970b). “Reflections on my critics.” En I. Lakatos y A. Musgrave, eds., *Criticism and the Growth of Knowledge*, pp. 231–278. Cambridge: CUP.
- Kuhn, T. (1977). *The essential tension*. Chicago: University of Chicago Press.
- Nola, R. (2000). “Saving Kuhn from the Sociologists of Science.” *Erkenntnis*, 68:41–53.
- Olmos, P. (2016). “La ponderación metaargumentativa al servicio de la racionalidad evaluativa. Otra lección aristotélica.” *Dilemata*, 22:257–281.
- Polanyi, M. (1958). *Personal knowledge*.
- Scheffler, I. (1967). *Science and Subjectivity*. Nueva York: Bobbs-Merrill.
- Searle, J. (2001). *Rationality in Action*. Cambridge MA: The MIT Press.
- Siegel, H. (1980). “Objectivity, Rationality, Incommensurability, and More.” (Review of *The Essential Tension*), *The British Journal for the Philosophy of Science* 31(4):359–375
- Stanford, P.K. (2017). “Underdetermination of Scientific Theory.” *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.): URL = <https://plato.stanford.edu/archives/win2017/entries/scientific-underdetermination/>
- Tindale, C.W. (2010). “Ways of Being Reasonable: Perelman and the Philosophers.” *Philosophy and Rhetoric*, 43(4):337–361.

SIMPOSIO:  
THE GENESIS OF GEOMETRIC  
KNOWLEDGE

# Introduction

**José Ferreirós**  
Universidad de Sevilla

The philosophy of geometry is a field with a long tradition, which is being enriched in recent times by new insights and proposals coming from the cognitive sciences. Some connections between philosophical reflections on geometry and space, and notions from the fields of physiology and psychology, can be traced back to the golden age of the philosophy of geometry. Thus, one can find interesting material in the classical writings of Helmholtz or Poincaré, to name only two key figures. Currently, however, a more sustained and sophisticated effort is being made to advance in this area, with contributions coming from many different disciplines. The special session we propose is thus interdisciplinary, combining history and philosophy of mathematics with reflections on current cognitive science, and even inputs from cognitive archaeology.

The main idea is to present some key lines of development of a research project which has started this same year at the Universidad de Sevilla: “The genesis of geometric knowledge” (FFI2017-84524-P). The three speakers are members of the project and represent different areas of expertise. Thus, the youngest researcher is doing a Ph.D. which combines history of mathematics and archaeology, with a view to understand proto-geometry as a source of knowledge. The most senior researcher is a historian and philosopher of mathematics, well acquainted with the history of geometry and with logical-foundational studies. And the mid-career one is an expert in the philosophical ideas of Poincaré, who works more generally on topics connecting mathematics and physics (e.g. the history of mechanics). All of these interdisciplinary connections, including also the links between early geometric ideas and astronomy, have important roles to play in understanding the emergence of geometric concepts, methods and results.

The three talks have been conceived so as to complement each other covering several key lines of thought. First will be the talk by José Ferreirós, offering a general presentation of the problem, arguments for a highly interdisciplinary approach, and a basic characterization of different levels of (proto-)geometric knowledge. Next the talk by María de Paz, focusing on Poincaré’s highly influential ideas, but especially on how they connect with insights about the psychological and physiological genesis of spatial ideas. Third the talk by Manuel J. García-Pérez, which will elaborate in more detail on the notion of proto-geometry taking into account results from comparative history and especially from the field of cognitive archaeology.

# The cognitive basis of Poincaré's philosophy of geometry

María de Paz  
Universidad de Sevilla

*«Let us now see the geometer at work,  
and try to surprise some of his methods»*  
Poincaré, 1902.

Poincaré's philosophy of geometry is usually known as conventionalism or geometrical conventionalism—in order to be distinguished from other specific aspects of his philosophy of science, in particular physical or mechanical conventionalism. The philosophical analyses of this position are based in his understanding that the axioms of geometry are not empirical or a priori propositions and that there are no spatial features that can determine the applicability of a geometry better than another. Hence, the axioms of geometry have a conventional status and the applicability of a determined geometry to space is a conventional decision. Traditionally, mostly in the interpretations developed by logical positivists and their successors, this is a philosophy used to decide which, among the different geometries, is the one that better suits a determinate physical conception of the world—or physical theory for short.

A corollary of the conventionalist perspective is that, given that the world or space has no particular features that can determine the choice among the different possible geometries, these geometries are equivalent because they are underdetermined by empirical data—meaning that there are no measurements which can establish the convenience of a geometry instead of any other. As a result, to some interpreters, this will make the choice *arbitrary*, equating this way the conventionalist perspective with an arbitrary choice among theories or, as in the present case, geometries.

This particular understanding of the conventionalist perspective fails to be faithful to Poincaré's position for two main reasons. First of all, and as some authors (Folina 2014) have recently remarked, it was never the intention of Poincaré to assert that the choice among theories—whether geometries or any other—is arbitrary. He stated repeatedly that “freedom is not arbitrariness” (Poincaré 1902), meaning that even if we are free to choose among different possibilities this does not imply that we can choose whatever. The second reason, which intends to be the focus of this talk, is that the above interpretation fails to capture the essence of Poincaré's conception as a philosophical position that interprets science not only from the point of view of completed theories—or from the context of justification in the language of logical positivists—but also from the point of view of the emergence and constitution of scientific developments, thus as a process. This is important to properly understand his philosophy of geometry because, as I will try to show, this philosophical position is linked to Poincaré's explanation

of the origin of geometry and of our geometrical conception of space. Besides, this explanation will connect with the first misunderstanding remarked above—the arbitrariness issue—because by understanding the connections between his philosophy of geometry and his explanation of the genesis of geometry, we will appreciate why the choice is never arbitrary, even if it cannot be determined by empirical data.

The genesis of our geometrical knowledge to Poincaré is heavily conditioned by the environment in which we live in. In several occasions he states that we have in the first place developed the geometry of solid bodies because “we are used to the properties of natural solids” (1902) and so, “doing geometry is studying the properties of our [measurement] instruments, that is of solid bodies” (1913). Those are the bodies surrounding us and we relate to them by means of our body, which acts as the first measurement instrument. It is by relationship to our body that we begin establishing spatial relations and measurements. Consequently, we need our perceptual apparatus and our body to create spatial relations which form the basis of geometrical concepts. Recent cognitive studies (Lakoff & Núñez 2000) explain the connection between the abstract body of mathematical knowledge and concrete bodily experience using terms such as conceptual metaphor or conceptual blending. This embodied conception of mathematical knowledge fits quite well with Poincaré’s ideas regarding the genesis of geometry as shown above.

The connection with the conventional position is precisely by means of abstraction. In the abstraction process we decide to eliminate, extract, or remove, certain practical, concrete empirical features of the bodies that serve as inspiration to ground geometrical concepts. We could have very well decided to eliminate other features resulting in different geometries, but mainly for convenience and simplicity reasons, we have in the first place created the geometry of the solid bodies, and that is why this geometry will always remain the most convenient one—at least in the eyes of Poincaré.

## Bibliografía

- Folina, Janet. 2014. “Poincaré and the Invention of Convention.” In *Poincaré, Philosopher of Science: Problems and Perspectives*, edited by María de Paz and Robert DiSalle, 25–45. New York: Springer.
- Lakoff, George y Rafael Núñez. 2000. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. New York, Basic Books.
- Poincaré, Henri. [1902] 1968. *La science et l’hypothèse*. Paris: Flammarion.
- Poincaré, Henri. 1913. *Dernières Pensées*. Paris: Flammarion.



# Mathematics and cognitive science: The case of geometry

José Ferreirós  
Universidad de Sevilla

The study of *mathematical cognition* is a well-established field that currently has societies and journals (like the MCLS and the *Journal of Numerical Cognition*), reference handbooks, and even best sellers such as *The Number Sense* by Dehaene, 2011. Yet the study of the cognitive and neural origins of geometry, arithmetic and symbolic mathematics is still in its infancy, and it is not equally well developed in all directions. The topic of numerical cognition has a well established tradition, from its origins in empirical and comparative psychology long ago, to recent neuroscientific studies; and although several competing paradigms are being discussed, one can find substantial levels of agreement among the experts. Important results are available concerning the basic systems and cognitive abilities that may (even innately) pave the way to “quantical” notions, providing initial access to the idea of cardinality of a collection. And quite a lot is known about the neural substrates of such systems.

In Spanish not much literature is available, but there exist publications such as *El árbol de los números*, ed. by Ferreirós & Lassalle Casanave in 2017, which include a substantial amount of material on numbers and cognition. Also noteworthy is the new special Issue of *Theoria* “From basic cognition to mathematical practice” (2018). At the University of Sevilla we are just starting to develop a research project on the topic “The genesis of geometric knowledge”, with the aim to promote interdisciplinary work on this issue.

The origins of geometric knowledge and its neurocognitive substrates are less clear than those of arithmetic, probably because of the greater complexity of the question. Some basic cognitive abilities linked with *numerosity* seem relatively easy to define, characterize, and to locate neurally (subitization & the analogue magnitude system). Meanwhile, visuo-spatial cognition is very prominent both neurally and mentally, widespread and highly complex. Nonetheless some proposals have been made concerning possible “core systems” for geometry (work by Spelke, Lee *et al.*), namely two: a system for ‘navigating’ the environment and a ‘landmark’ system for object features; together they would pave the way to “geometric” notions like direction, angle. Although the core systems paradigm (a more refined form of modularism) is contested, and alternative proposals have been offered, what seems well established is that the human cognitive system is biased toward the extraction of “spatial”, figural or proto-geometric invariants from the perceived features of the environment. This in itself is highly interesting: evidence that, already from birth, humans are biased toward figural or featural information that may be roughly described as “geometrical”.

In the case of numbers and arithmetic, it is quite obvious that symbols play

a great role in the emergence of precise concepts amenable to mathematical treatment (Carey, Menary). This seems to argue against innatism or intracranialism, speaking in favour of theses like the embedded mind or even extended cognition. Susan Carey, despite being prone to innatism and a proponent of core cognition, argues that the culturally-given series of numeral words (learnt and memorized by the child) is essential to allow the “bootstrapping” process that gives rise to the number concept from core systems of numerosity. The role of symbols and external representations makes a clear case for the need to tackle these issues interdisciplinarily. The kind of cognition evidenced already at the level of elementary mathematics (say, school-level problem solving) is highly complex, it is a locus where basic cognition meets semiotics, and where the role of enculturation in sculpting human cognitive processes must be considered (Menary).

The role of (external) symbolic representations has been highlighted in recent years by studies on the role of diagrams in Ancient geometry. Studies by Manders, Giaquinto and others have carefully scrutinized the way in which visual-diagrammatic thinking is an essential, constitutive part of Euclidean proof practices. The same happens also in other old geometric traditions, like the Chinese classic *Nine Chapters* of mathematical procedures. Thus there are historical reasons, substantial historical evidence, to believe that geometry properly speaking does not emerge without external symbols, cognitive tools that extend and reshape our abilities. And the question emerges to map the territory between basic cognitive abilities and the rise of geometric knowledge—of results about similar triangles (Thales) or about figures drawn on the sides of a rectangular triangle (Pythagoras).

In the spirit of Carey’s work on numbers, somebody who aims to work on geometric cognition must consider: 1) How to characterize the basic cognitive abilities having to do with visual and spatial cognition, which make our perception biased toward figural invariants. 2) How to identify the changes occurring in the course of development (either individual or historical), in order to highlight the differences between basic and advanced cognition (e.g. child and adult). And 3) how to explain the processes of conceptual change responsible for those developments, consistently with available evidence.

I will defend that, in order to achieve those goals, a highly interdisciplinary investigation must be undertaken, trying to combine experimental results in cognitive science (neural or not) with evidence from comparative history and even archaeology. Given the complexities involved in any attempt to bridge the gap from basic cognition to established mathematical practices, it seems inevitable to take into account philosophical and logical analyses of mathematical knowledge. It is through a combined effort to integrate all of those facets that we hope to contribute to the realization of Carey’s Program in connection with geometry.

The last part of the talk will be devoted to an attempt to characterize changes (point 2 above) inspired by an analysis of comparative historical evidence from ancient Greece and China, together with insights taken from the field of cognitive archaeology (Malafouris, Overmann). I will defend the thesis that one must distinguish at least three stages: i) basic *visuo-spatial* cognition, based on perception and biased toward figural traits such as shapes, angle, distance,

ii) *proto-geometric* knowledge, developed in prehistoric times and linked with the manufacture of tools, rudimentary measurement, rites associated with astronomical ideas, and iii) *geometric* knowledge in the strict sense, presented in symbolic form e.g. by means of diagrams, in different historical forms like e.g. the *gou-gu* geometry of the Chinese.

# The prehistorical roots of geometry: an overview

Manuel Jesús García-Pérez  
Universidad de Sevilla

Some authors have declared that the early history of mathematics could be divided into three main periods: i) proto-geometry or prehistory of geometry; ii) basic geometry oriented to practical goals; and iii) geometry in a proper sense. While some works on the history of mathematics recognise this tripartite division, little attention is paid to the prehistoric data. The birth of what can be called as “mathematics” is placed with the emergence of fluvial civilizations—Babylonia, Egypt and Mesopotamia. Nevertheless, the cornerstone of what mathematics is is situated some centuries later, with the elaboration of some key compilations such as Euclid’s *Elements*, or the compositions and commentaries from Liu Hui to the *Nine Chapters on the mathematical procedures*.

Although there is a good consensus in this genetic history, the correct interpretation and presentation of prehistoric data turn out to be a tricky venture. It is usual in the “A complete history of mathematics” type of works to present an introductory chapter about the prehistory of mathematics. Here, two possible and interconnected sources of mathematical knowledge are presented: on the one hand, the cognitive foundations of mathematics; on the other hand, material records testifying the existence of counting practices or the recognition and use of some spatial relations. However, these ideas are presented unsystematised and assume some problematic facts. I will show some common misconceptions related with them, and later I propose a relevant framework to understand correctly these issues.

In the first place, often an innatist conception of human cognition is presupposed. In that sense, the historical literature emphasizes that some practices as counting—tally marks in some bones or rocks—or drawing spatial motifs—body paintings or pot’s decoration—displayed by human past populations, gave rise naturally to later mathematical developments. However, this is not a trivial issue. One of the current discussions in some areas of cognitive science is centred on it, with some lively debates in the last year.

In the second place, we have some “retroprojection assumptions”; i.e., the—generally false—attribution of some geometric knowledge to past societies. Some material records have been interpreted as presenting mathematical ideas, assuming that some mathematical concepts or results flourished in Neolithic or even Palaeolithic period. One of the recurring topics is related with some “sacred” or “ritualistic” places—as Stonehenge or Woodhenge—and the use of some geometric knowledge to their construction—mainly related with the Pythagorean Theorem. Thus, it is stated that this material record shows that humankind possessed geometric knowledge as far back as the prehistorical times.

In our view, based in an interdisciplinary approach to the cognitive and archaeo-historical foundations of geometry, it is necessary to distinguish between three levels in the process that leads from our past cognitive abilities, to the emergence of explicit geometric knowledge. Briefly, these levels are: 1) visuo-spatial cognition, related with the human perception and recognition of shapes; 2) proto-geometry, where basic geometric concepts are formed, and some approximate results obtained; and 3) theoretical geometry, or geometry in a proper sense.

It is not our interest to give a detailed presentation of the three levels here. Just to mention that all past material record that is often geometrically interpreted, as pot's surface drawings, Palaeolithic art, burials' orientation, and so on, are related with our visuo-spatial perception.

However, there is a moment in the prehistory when some proto-geometric knowledge arises, being related with the achievement of some necessary conditions, such as: i) the establishment of the first sedentary communities, which allow populations to grow and to have some leisure; ii) the division, organization and specialization of labour, enabling such populations to preserve and improve any kind of knowledge—practical or theoretical; and iii) the use of some tools, enhancing our cognitive abilities—our memory, for example—and allowing to transmit information in a more accurate way, reaching more distant populations.

Particularly, in relation to the emergence of proto-geometry, it is important to notice that some Neolithic cultures began to perform some kind of activities where the use of “mathematical” tools and the development of some basic geometric concepts were important. This episode is essentially connected with the development of archaeoastronomy, especially with the relevance that some populations started to give to the possession and improvement of some proto-astronomical knowledge.

This body of knowledge became essential to the foundation of new kingdoms. Ideologies, a key element to agglutinate a group of people, were in some cases based on the performance of some shared ceremonies and rituals, and the use of some common symbols and myths. These elements were linked, in some sense, with the acknowledgment and correct interpretation of astronomical phenomena.

China, among other civilizations, offers relevant material to analyse this issue. For example, in the Liangzhu Culture (3300–2200 b.c.e.) some “ritualistic” associations between heaven and circle emerge, developing the *gaitian* concept (the canopy-heaven). Later, in the Zhou dynasty (1046–771 b.c.e.), a reliable calendar was needed to accurately perform some rituals. To this purpose, a *biao*, or gnomon, was used. Furthermore, some cities in the Longshan culture (3000–2000 b.c.e.) were planned in agreement with some astronomical phenomena, building the Taosi Observatory to find the correct place were to situate the city capital.

In these examples one can see that some mathematical tools and basic geometric concepts—e.g., circle or square—started to be used. However, a crucial step is made by these populations. Here, to perform correctly these rituals, or to build accurately the cities—essential ingredients to maintain a strong sense of community and celestial blessing of their chiefs—it was needed a more sophisticated, theoretically-oriented use of spatial relations. Consequently, proto-geometry begins to be present.

Therefore, prehistorical studies of geometry, cognitively oriented, are an indispensable source that helps us to understand the emergence of mathematics. Moreover, our work emphasizes that beyond possible innate cognitive foundations of geometry, there are cultural components that cannot be bypassed. In this sense, the anthropological, cognitive and cultural roots of geometry are essential to understand its genesis and evolution along history.

**SIMPOSIO:  
BIOLOGICAL INTER-IDENTITIES**

# Introduction

**Arantza Etxeberria Agiriano**  
**Kepa Ruiz-Mirazo**

This symposium is motivated by the aim of presenting some partial results of our Project on *Identity in interaction* (MINECO FFI2014-52173-P). It aims to review the concepts of identity and individuality in biology, mainly from an organizational framework, including fields like origins of life or ‘evodevo’. In line with previous work of the IAS-Research group, the contributed papers highlight the relevance of biological organization, the sciences of complexity, and autonomy, as well as a systemic aspects of evolutionary biology.

The received concepts of identity and individuality in philosophy have been criticised for not questioning the individualistic and particulate assumptions permeating research on living beings and their mental and social capacities. Naturalist descriptions of the organization of relevant entities make evident that there are two important challenges to be faced by the received concepts: 1) the massive interconnectivity makes it very difficult to establish which are the “primitive” entities or the starting point; 2) the heterogeneity of constituents shows that the conjunction of entities forms complex organizations instead of collections or populations of homogeneous elements.

In this context, we contend that new theoretical models and explanatory accounts are required for cases such as: metabolic interactions originating the first cells, lateral gene transfer, symbiotic relationships, major evolutionary transitions, the formation of multicellular organisms and biological reproduction.

We present 5 papers on this topic reviewing the interactive phenomenologies related to identity in the following aspects:

Prebiotic evolution. – How biological identities were—or can be—constructed for the first time will be here explored, from an interactivist perspective that takes into account both metabolic and evolutionary aspects.

Biological reproduction. – Biological organization involves material overlapping and reproducer transformation. Here pregnancy and the female-fetus relation in its developmental and evolutionary aspects motivate a philosophical study of interidentity.

Hierarchical organization. – Inter-identity compared to previous interactionism in evolutionary and developmental biology from an explicitly hierarchical perspective levels of organization and material constraints.

Collaborative interactions. – An organizational account of collaboration in which the collective dimension of life is acknowledged without ignoring the individual organization of living systems and its key role as the locus of mechanisms, of



adaptations, of selective-evolutionary dynamics, and its role in the evolution of biological complexity.

Biological Communication. – How minimal living and prebiotic systems can interact by exchanging signals, and how such interactions can be studied by synthetic biology through models involving protocells and synthetically modified cells, in such as way as to shed light on the phenomenon of biological communication in general.

# Collaboration and biological individuality

Argyris Arnellos

argyris.arnellos@ehu.es

Collaboration is a central characteristic of life that is not always appreciated sufficiently due to the Neo-Darwinist stress on competition. Still, there should be no doubt that collaboration is one of the main and central dimensions of life and its evolution. Examples of collaboration consist of bacterial biofilms, ecosystems and various multicellular organizations such as those of animals, plants and fungi, as well as of all these systems and their symbionts.

On the other hand, the emphasis on collaborative dimension of life, as it has so far been shaped and discussed under neo-Darwinian and adaptationist frameworks (see e.g., Queller and Strassmann 2009; 2016) as well as from a process-oriented perspective (Dupré and O'Malley 2009; Dupré 2012), overlooks the importance of biological individuals (conceived as integrated, self-maintaining organizations) in the build-up of more complex collaborative networks in the course of evolution. A common but important drawback of these approaches is that they either don't do justice to the individual status of the diverse biological organizations or they suggest an unnecessary pluralism that eventually blurs the concept of biological individuality altogether.

In this paper, I provide an organizational account of collaboration in biology that overcomes the main difficulties of previous accounts. Starting from earlier work on organizational accounts in multicellular organisms (Arnellos et al. 2014; Arnellos and Moreno 2015; 2016; Keijzer and Arnellos 2017; Arnellos in press) I introduce a concept of collaboration as roughly a function of different forms of integration between interaction (materialized through various forms of sensorimotor coordination), physiological maintenance (mainly related to metabolic as well as fluid-circulation and nutrition-related aspects), and constitution/development (either through aggregation or through egg-based development) that as I argue spans the whole biological realm. I apply this ontology to describe different types of collaborations in the unicellular and the multicellular realm, arguing that an organizational account of biological collaboration meets the real challenge of explaining the diversity of biological organizations by doing justice to the concept and thus providing a clear account of biological individuality.

## References

- Arnellos, A. (2018). From organizations of processes to organisms and other biological individuals. In: D. Nicholson and J Dupré (eds). *Everything Flows:*

*Towards a Processual Philosophy of Biology*. Oxford University Press (OUP).

- Arnellos, A. and Moreno, A. (2015). Multicellular agency: an organizational view. *Biology and Philosophy*, 30(3):333–357. doi: 10.1007/s10539-015-9484-0
- Arnellos, A. and Moreno, A. (2016). Integrating constitution and interaction in the transition from unicellular to multicellular organisms. In: Niklas K, Newman S (eds.) *The origins and consequences of multicellularity*. MIT Press, Cambridge MA, pp. 249–275.
- Arnellos, A., Moreno, A., and Ruiz-Mirazo, K. (2014). Organizational requirements for multicellular autonomy: insights from a comparative case study. *Biology and Philosophy*, 29(6):851–884.
- Dupré, J., (2012). *Processes of Life: Essays in Biology*. Oxford, Oxford University Press.
- Dupré, J. and O'Malley, M. A. (2009). Varieties of living things: life at the intersection of lineage and metabolism. *Philos Theory Biol*, 1:e003.
- Keijzer, F. and Arnellos, A. (2017). The animal sensorimotor organization: a challenge for the environmental complexity thesis. *Biology and Philosophy*, 32:421–441. doi:10.1007/s10539-017-9565-3
- Queller, D. C. and Strassmann, J. E. (2009). Beyond society: the evolution of organismality. *Philos Trans R Soc B Biol Sci*, 364:3143–3155.
- Queller, D. C. and Strassmann, J. E. (2016). Beyond society: the evolution of organismality. *Biology and Philosophy*, 31:855–873. doi:10.1007/s10539-016-9547-x

# An organisational framework for the synthetic modeling of biological communication

**Leonardo Bich**

IAS-Research Centre for Life, Mind and Society. Department of Logic and Philosophy of Science.  
University of the Basque Country. Avenida de Tolosa 70, Donostia-San Sebastian, 20018, Spain.

**Ramiro Frick**

Department of Sociology, Universidad Diego Portales, Manuel Rodríguez Sur 415, Santiago, Chile.  
Department of Philosophy, Universidad Alberto Hurtado, Alameda 1869, Santiago, Chile.

What is the role of synthetic models in the study of the origin and nature of minimal forms of cognition and communication? According to theoretical accounts built on the notions of biological organisation and autonomy, minimal cognitive properties are related with the interactive dimension of life. An autonomous organisation is capable to self-produce and self-maintain. It cannot exist unless it maintains a continuous coupling with its environment, made possible by an internal dynamical variability that enables different viable responses to a variety of environmental perturbations. In this context, minimal or proto-cognition emerges when distinctions in the environment make a difference for the organism. And these meaningful distinctions are realised through the activity of biological mechanisms of self-regulation, that are sensitive to specific features of the interaction between system and environment, and modify the internal dynamics of the system accordingly.

However, not all cognitive interactions are instances of communication. The phenomenon of communication includes only a restricted subset of interactions between biological systems, and it needs to satisfy specific requirements regarding the relationship between senders and receivers.

In order to characterise the distinctiveness of the phenomenon of communication and how it can be captured by synthetic models, we analyse conceptual and experimental work in synthetic biology on different types of interactions considered as examples or models of communication. We discuss the pertinence and relevance of these models for the wider understanding of this biological phenomenon and of its minimal instances, and we critically analyse their limits.

We argue that a precise conceptual framework is missing, and we provide a theoretical account of communication—based on the notion of organisation, and characterised in terms of influence of the sender upon the receiver. The influence account of communication is based on the idea that communication takes place

when a signal emitted by a sender triggers a change in the behaviour of the receiver that is functional for the sender itself. Yet, such functional influence has been characterised in an evolutionary framework according to which what is functional for the sender is interpreted in terms of adaptations: the signal is a functional trait because it allowed the ancestors of the sender to survive. Stated in these terms, this approach has several limits, one of which is that characterising communication as a product of natural selection rules out *a priori* the very possibility of an artificial, non-evolved communication system, making this approach useless in the context of synthetic biology.

However, this account can be reframed in terms of the current organisation of the system rather than in evolutionary terms, and put to work in synthetic biology. This operation can be done by grounding the notion of functional influence into the organisational account of biological functions according to which a function is understood as a contribution of a trait to the maintenance of an organisation (e.g. a living cell) that, in turn, contributes to produce and maintain the trait itself. Applied in the context of communication, according to an organisational approach to say that a signal is functional specifically means that, by triggering a reaction in the receiver, it contributes to the maintenance of the current organisation of the sender, without necessarily appealing to its evolutionary history.

This framework can be operationalised in synthetic biology, and can supply criteria and guidelines for the design and evaluation of synthetic models. By making explicit the distinctive features of communicative phenomena, it can provide tools and criteria to reframe the interactive dynamics between natural and artificial cells in order to model behaviours that are relevant for understanding the nature and roots of biological communication. In particular, this approach puts into evidence the necessity for modellers to shift the attention from designing protocells that can interact with cells, to protocells that can participate in functional loops with cells or among themselves.

# Biological individuality and reproduction: Intermittent interidentity in the evolution of pregnancy

**Arantza Etxeberria**

IAS-Research Centre for Life, Mind and Society. Department of Logic and Philosophy of Science. University of the Basque Country. Avenida de Tolosa 70, Donostia-San Sebastian, 20018, Spain.

**Laura Nuño de la Rosa**

Department of Logic and Theoretical Philosophy, Complutense University of Madrid

The relation of gestating females and embryos in mammals offers an interesting scenario to explore interidentity in biological reproduction. In our view, the current developmental and evolutionary knowledge on different modes of reproduction challenges usual assumptions of numerical biological individuality: although usually embryos and foetuses count as fully fledged individuals in many biological fields, including evolutionary biology, if we take into account the relative autonomy of pregnant females and embryos throughout development and the different reproductive strategies or kinds of relations between the related entities explored in evolution, the complex biological association formed by mother and embryos affords a manifold of variant cases worth examining.

From an organizational perspective, different situations can be found, some of which preclude the intuitive judgment of counting 2 (or more) individuals during pregnancy, and offer evidences to consider that the relation among mother and foetus resembles, at some stages, a host/resident or whole/part one, rather than a relation among two (or more) autonomous beings. The maternal/embryo interactions occurring in development has been analyzed from an organizational perspective (Nuño de la Rosa 2010), and recently Kingma (to appear) has explored the metaphysical consequences of these interdependencies. In this paper, we intend to retake this issue as a case study of biological individuality in reproduction from an evo-devo perspective, and will review literature on the developmental evolution of modes of reproduction, including the evolution of implantation and inflammation (Griffith et al. 2017), the role of the placenta in different taxa, and the literature on the maternal recognition of pregnancy.

The following are the main goals of the paper: 1) to question some intuitive ideas about biological individuality in reproductive processes; 2) to propose criteria for individuality that take into account phenomena appearing in reproduction; 3) to distinguish stages of autonomy and dependence during pregnancy and their evolutionary significance.

## References

- Griffith, O. W., Chavan, A. R., Protopapas, S., Maziarz, J., Romero, R., & Wagner, G. P. (2017). Embryo implantation evolved from an ancestral inflammatory attachment reaction. *Proceedings of the National Academy of Sciences*, 201701129.
- Kingma, E. (2018). Were You a Part of Your Mother? The Metaphysics of Pregnancy. Forthcoming.
- Nuño de la Rosa, L. N. (2010). Becoming organisms: the organisation of development and the development of organisation. *History and philosophy of the life sciences*, 289–315.

# The construction of biological ‘inter-identity’ as the outcome of a complex process of prebiotic evolutionary development

**Kepa Ruiz-Mirazo**

D. Logic and Philosophy of Science, University of the Basque Country, Spain  
Biofisika Institute (CSIC, UPV-EHU), Spain

**Ben Shirt-Ediss**

Interdisciplinary Computing and Complex BioSystems Group, Newcastle  
University, UK

**Álvaro Moreno**

D. Logic and Philosophy of Science, University of the Basque Country, Spain

The concept of ‘identity’ is intricate, philosophically speaking, but that makes it especially interesting to explore (or re-explore) in theoretical biology, where its explanatory and heuristic possibilities are numerous. In fact, this concept has a rather ambivalent character: it is used both to distinguish a system as a particular entity (among other similar ones) and to relate a system with other members of a group/set (i.e., to classify it within a common category). That double—and somewhat contradictory—epistemic role is probably linked to the two classical sides of the ‘problem of identity’: (i) what makes a system remain the same as it was before? (i.e., identity as *permanence* or *endurance* over time); and (ii) what makes a system be the same as some other system? (i.e., identity as a synchronic *equivalence relationship*). Providing an account about the stepwise emergence of complex dynamic systems that produce and maintain robustly an identity of their own—and not only in this double sense but also including a crucial interactive dimension, both ‘system-environment’ and ‘system-system’, or ‘multi-system’—could help us understand the overwhelming ‘hyper-complexity’ of biological phenomena.

Without delving into a debate that has taken the time and energies of many analytical philosophers in the past, our contribution will just claim that the ‘double-sidedness’ of the identity concept should *not* be resolved (in favor of one or the other side) but employed constructively (and complemented with other relevant sides of it—in particular, that unavoidable interactive side) to gain a better grasp on the multifarious nature of life. The discussion will be carried out in the context of current investigations on the process of biogenesis, conceived as a set of non-trivial transitions from the most intricate chemical systems towards the simplest biological ones (see Fig. 1). Accordingly, several relevant distinctions between various types of dynamic molecular entities/assemblies/organizations, at different stages of prebiotic evolution, will have to be drawn and argued for. But the main challenge of this essay will consist



in explaining that the development of systems with a proper biological ‘inter-identity’ necessarily involves integrating phenomena taking place at very different space and temporal scales and, thus, combining ‘actual properties’ of those systems with ‘potentialities’ intrinsic to their wider and longer-term collective-evolutionary behavior. In other words, living ‘inter-identities’ are highly complex forms of organization because they subsume (they reflect, in a nutshell) life’s inherent ecological and historical dimensions.

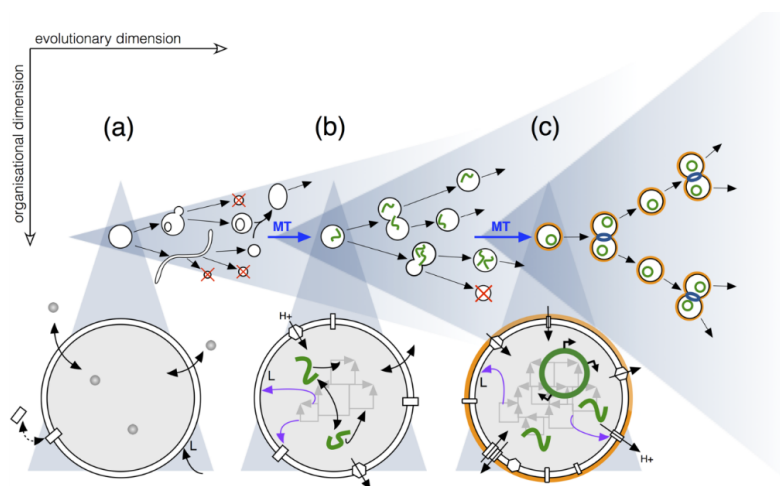


**Fig. 1:** Scheme depicting the process of origins of life as the unfolding of different types of protocellular organization with increasingly robust metabolic capacities, longer-term sustainability and higher evolutionary potential. Image taken from Ref. [1]

- [1] Shirt-Ediss, B., Murillo-Sánchez, S. & Ruiz-Mirazo, K. (2017): Framing major prebiotic transitions as stages of protocell development: three challenges for origins-of-life research. *Beilstein Journal of Organic Chemistry* 13: 1388–1395.

# Biological individuality and reproduction: Intermittent interidentity in the evolution of pregnancy

**Jon Umerez**

Dept. of Logic & Philosophy of Science, University of the Basque Country (UPV/EHU); Donostia

**M. J. Ferreira**

Institute of Philosophy, University of Buenos Aires (UBA) & Conicet; Buenos Aires

Since Oyama 1985 (continued in other works as, v.g., Oyama 2001 or several of those collected in Oyama 2000) and even earlier with Lewontin (v.g., 1982, 1983 or also 2000 and Lewontin et al. 1984)—just to mention two basic references without presuming the relevance of others—it is being noticed the difficulty to get beyond the mere statements, more or less rhetoric, asserting, for instance, that development is the result of the “interaction” between genes and environment. There is a certain “interactionist consensus” that almost no one defies but which is still poorly understood.

In Lewontin’s survey (2000), they are equally questioned both the relation gene-organism and the relation organism-environment. In the last term, this questioning brings us back to the same old issue exposed by Oyama in the frame of the more general dichotomy between “nature and nurture” and its criticism as Keller has also addressed more recently (2010). The reexamination made by Lewontin of the three items included in the title of his book (gene, organism, environment) puts them explicitly into relation with the necessity to develop an appropriate dialectics of “parts and wholes” or, in other terms, with the necessity to relate levels of organization and articulate what inter-level relations are and what do they imply (Umerez 1994, 2016).

The goal of this work is to explore potential connections of some of the ideas developed in the present team project addressing the concept of identity and individuality in terms of interaction, with those previous perspectives of interactionism in evolutionary and developmental biology. It is intended to examine to which extent the difficulty to elaborate a substantive notion of interaction (interactionism), beyond the mere appeal to the necessity of taking into account diverse factors when trying to understand the organism, has any reflection in the project to develop the idea of inter-identity in these and other fields. Additionally, we think it will be illuminating to link that analysis with the effort to understand biological organization from an explicitly hierarchical perspective in terms of levels of organization and the action of material constraints that instantiate their relation.

Thus, we will start by showing that there is a conceptual connection between the generic idea of identity in interaction and interactionist approaches and that

the development of the former may benefit from this overlap. Next, will next establish a map of those interactionist approaches more directly relevant to biological issues and try to define their scope and limitations. This typification is complemented with Pattee's approach to biological organization, which is not couched in the usual terms of interactionism but can be fruitfully interpreted in this sense ("...a perspective which we could call 'internally' interactionist...", Umerez 2001, p. 161). Finally, we will inscribe this account within an organizational perspective, explicitly hierarchical, developed around the concept of constraint (Umerez 1994, Umerez & Mossio 2013) as the tool to instantiate materially the connections assumed.

## Project reference

FFI2014-52173-P – Identity in interaction: ontological and normative aspects of biological, cognitive and social individuality. FFI2014-52173-P, from the Ministerio de Economía y Competitividad (MEC) and FEDER funds from the EC.

## References

- Keller, E.F. (2010). *The Mirage of a Space between Nature and Nurture*. Durham, NC: Duke University Press.
- Lewontin, R.C. (1982). Organism and environment. In: Plotkin, H.C. (Ed.), *Learning, Development, and Culture*. Wiley, New York, pp. 151–170.
- Lewontin, R.C. (1983). The organism as the subject and object of evolution. *Scientia*, 118, 65–82.
- Lewontin, R.C., Rose, S., Kamin, L.J. (1984). *Not in Our Genes*. Pantheon Books, New York.
- Lewontin, R.C., Rose, S., Kamin, L.J. (2000). *The Triple Helix. Gene, Organism, and Environment*. Cambridge, MA: Harvard University Press.
- Oyama, S. (1985). *The Ontogeny of Information*. Cambridge: Cambridge University Press.
- Oyama, S. (2000). *Evolution's Eye. A Systems view of the Biology-Culture Divide*. Durham, NC: Duke University Press.
- Oyama, S. (2001). *Terms in tensión: what do you do when all the good words are taken?* In S. Oyama, P. Griffiths & R.D. Gray (eds.) *Cycles of contingency. Developmental Systems in Evolution*. Cambridge, MA: The MIT Press, pp.: 177–193.
- Umerez, J. (1994). *Jerarquías autónomas. Un estudio sobre el origen y la naturaleza de los procesos de control y la formación de niveles en sistemas naturales complejos*. Tesis Doctoral, UPV/EHU.

- Umerez, J. (2001). H. Pattee's theoretical biology. A radical epistemological stance to approach life, evolution and complexity. *BioSystems*, 60(1/3), 159–177.
- Umerez, J. & Mossio, M. (2013). Constraint. In Dubitzky, W., Wolkenhauer, O., Cho, K-H. & Yokota, H. (eds.) *Encyclopedia of Systems Biology*, pp. 490–493. New York: Springer.
- Umerez, J. (2016). Biological Organization from a Hierarchical Perspective. Articulation of Concepts and Inter-level Relation. In N. Eldredge, T. Pievani, E. Serrelli & I. Tëmkin (eds.) *Evolutionary Theory. A Hierarchical Perspective*, 63–85. Chicago, ILL: The University of Chicago Press.

**SIMPOSIO:  
EVIDENCIA CIENTÍFICA Y  
REGULACIÓN: ALIMENTOS Y  
FÁRMACOS**

# Introducción<sup>1</sup>

La regulación de las tecnologías en las sociedades actuales utiliza el conocimiento científico como base de la toma de decisiones. La actividad científica que se ocupa de suministrar este conocimiento es lo que se conoce desde hace algún tiempo como ciencia reguladora, que mantiene ciertas características distintivas respecto a la ciencia académica. Este simposio se ocupa de analizar cómo se caracteriza la evidencia científica en la ciencia reguladora. Además, se tratará de identificar el modo en el que el contexto regulador influye en la toma de decisiones metodológicas y en el desarrollo de la investigación científica. El simposio se centrará en alimentos y fármacos, dado que constituyen dos áreas en las que se plantean importantes cuestiones filosóficas relativas a la evidencia, el establecimiento de nexos causales, la determinación de los estándares de prueba, así como los efectos no epistémicos relevantes para la salud pública. Uno de los debates principales respecto de las investigaciones sobre alimentos y fármacos es la contraposición entre, por un lado, las propuestas del establecimiento de jerarquías evidenciarias y, por otro, de la adopción de un pluralismo metodológico en la generación de datos. Un aspecto particular de estos debates es la función de las pruebas controladas aleatorizadas (Randomized Controlled Trials), que plantea problemas importantes relacionados con los análisis en filosofía de la ciencia sobre los requisitos para las afirmaciones causales.

---

<sup>1</sup>Simposio parcialmente financiado por Secretaría de Estado de Investigación, Desarrollo e Innovación del Gobierno de España: Proyecto de Investigación: Estándares de prueba y elecciones metodológicas en la fundamentación científica de las declaraciones de salud (FFI2017-83543-P).

# Health claims. La frontera entre alimento y medicamento

**Montaña Cámara Hurtado**

Prof Titular Universidad

Dpto. Nutrición y Ciencia de los Alimentos

Fac. Farmacia, Universidad Complutense de Madrid

El etiquetado de los alimentos es una herramienta de gestión y comunicación del riesgo alimentario que constituye, en última instancia, una herramienta clave para permitir que el consumidor realice elecciones informadas sobre los alimentos que compra y consume. Nunca hasta ahora se ha tenido tal conocimiento de la relación existente entre alimentación y salud, ni se han generado tantas situaciones de incertidumbre científica, todo ello a pesar de que los alimentos nunca han sido tan seguros como lo son hoy en día. Además, aunque la oferta de alimentos saludables es muy amplia, nuestras elecciones dietéticas no siempre son las más adecuadas. Dando por hecho que unos malos hábitos dietéticos constituyen uno de los múltiples factores de riesgo de aparición de enfermedades crónicas y con el fin de no confundir a la población, deben aplicarse requisitos específicos de etiquetado de los alimentos. Este es el principal objetivo del Reglamento europeo de alegaciones nutricionales “Health Claims” (Reglamento (CE) 1924/2006) que prohíbe que un alimento (y/o complemento alimenticio) pueda promocionarse como poseedor de propiedades terapéuticas o curativas y establece distintas categorías de declaraciones: “declaraciones nutricionales” o “de contenido”; “declaraciones de propiedades saludables”, “declaraciones de reducción del riesgo de enfermedad”. La exigencia del citado Reglamento es que cualquier declaración se base en evidencias científicas contrastadas y reales. Se aplicará a las declaraciones nutricionales y de propiedades saludables efectuadas en las comunicaciones comerciales, ya sea en el etiquetado, la presentación o la publicidad de los alimentos que se suministren como tales al consumidor final y en ningún caso darán a entender que el producto en cuestión posee propiedades preventivas, terapéuticas o curativas de una enfermedad humana. La aplicación de este Reglamento con las distintas autorizaciones publicadas, no está exenta de polémica, dejando a veces en cuestión el cumplimiento de su objetivo de velar por la protección del derecho de los consumidores a una información veraz, contrastada y con un riguroso fundamento científico, aspecto particularmente relevante en el caso de los alimentos, y evitando su confusión con los medicamentos. El binomio alimentación – salud es un caso claro en el que los científicos y agencias reguladoras (la evidencia científica y la regulación) no siempre dan las respuestas esperadas, y donde aún hay mucho trabajo que realizar en pro de la cultura científica de los ciudadanos.

# Estándares de prueba y valores no cognitivos en la ciencia reguladora: el caso de las declaraciones de salud<sup>1</sup>

**José Luis Luján**  
**Juan Bautista Bengoetxea**  
Universitat de les Illes Balears

En este trabajo se analiza la influencia de los valores no cognitivos en la determinación de los estándares de prueba en la ciencia reguladora. Con esta finalidad se examina la controversia en torno a la regulación de las declaraciones de salud en la UE, controversia en la que surge una divergencia respecto de los objetivos que ha de perseguir dicha regulación. Para algunos autores, se trata de proteger a los consumidores de la posible información falsa respecto de las propiedades saludables de ciertos alimentos, objetivo que también es fundamental para los reguladores. Para otros, en cambio, la meta principal consiste en proporcionar información sobre la mayor cantidad posible de alimentos con propiedades saludables. Ambos tipos de objetivos regulatorios pueden entenderse como valores no cognitivos. En este trabajo se muestra que la controversia científica relativa a los estándares de prueba en la investigación científica sobre las propiedades saludables de los alimentos está directamente relacionada con dichos objetivos (valores no cognitivos) de la regulación. Por lo tanto, consideramos que la determinación de los estándares de prueba no depende exclusivamente de consideraciones generales relativas al mejor conocimiento disponible, sino también, y de forma especialmente pertinente, de los objetivos de la regulación.

---

<sup>1</sup>Trabajo financiado por Secretaría de Estado de Investigación, Desarrollo e Innovación del Gobierno de España: Proyecto de Investigación: Estándares de prueba y elecciones metodológicas en la fundamentación científica de las declaraciones de salud (FFI2017-83543-P).



# Los registros electrónicos sanitarios como evidencia regulatoria

**Niccolo Tempini**

U. Exeter

**David Teira**

UNED

En diciembre de 2016 el congreso estadounidense aprobó la 21st Century Cures Act (21CCA), una de cuyas secciones introducía, de facto, el pluralismo metodológico en la evaluación de medicamentos. Desde 1962, la Food and Drug Administration (FDA) sólo admitía ensayos clínicos frecuentistas como prueba de la seguridad y eficacia de los nuevos tratamientos. La 21CCA invita a la FDA a servirse de nuevas fuentes de evidencia para efectuar la evaluación de medicamentos, tales como los ensayos clínicos bayesianos o los registros electrónicos sanitarios (electronic health records). Durante la última década, numerosos filósofos de la ciencia han defendido el pluralismo metodológico contra las jerarquías de evidencia coronadas por los ensayos clínicos convencionales. Sin embargo, los críticos de la 21CCA consideran que semejante pluralismo, en la esfera regulatoria, sólo perjudica los intereses del paciente. A través de un análisis de los registros electrónico sanitarios, discutiremos qué garantías de imparcialidad deben de ofrecer estas bases de datos para que puedan usarse imparcialmente como fuente de evidencia regulatoria.

# La regulación de las declaraciones de salud y el debate público<sup>1</sup>

Oliver Todt

Noemí Sanz

Universitat de les Illes Balears

El objetivo de la presente comunicación es el análisis de las controversias sobre el nivel de evidencia científica que se considera adecuado para la toma de decisiones en la regulación europea de las declaraciones de salud (o declaraciones de propiedades saludables, *health claims*), así como los efectos que estas controversias puedan tener para la confianza de los consumidores en los reguladores.

Las declaraciones de salud contienen informaciones sobre los beneficios (particularmente relacionados con la salud) que un alimento aporta a sus consumidores, más allá de sus meras propiedades alimenticias. Según el tipo de regulación a la que estén sujetos, esas declaraciones incluyen, además, una justificación científica de los beneficios declarados, y en muchos casos indicaciones sobre las limitaciones de su consumo. En la Unión Europea, las autorizaciones sobre la base de evidencia científica de las declaraciones de salud son llevados a cabo por la Agencia Europea de Seguridad Alimentaria (EFSA, en sus siglas en inglés). La puesta en el mercado de alimentos con declaraciones de salud puede tener efectos importantes no solamente para la salud individual de sus consumidores, sino (por los efectos agregados de un consumo extendido) sobre la salud pública.

La regulación europea establece una jerarquía evidenciaria en la que las pruebas controladas aleatorizadas (*randomized controlled trials*, RCT) se convierten en una condición necesaria y suficiente para la autorización de las declaraciones de salud. Otro tipo de evidencia, por ejemplo, de estudios epidemiológicos, ensayos con animales o estudios *in vitro*, constituyen evidencia complementaria, pero nunca decisiva. En otras palabras, en la práctica, EFSA exige datos que establezcan relaciones causales entre la ingesta de un alimento y los efectos deseados.

El nivel muy alto de exigencia con respecto a la evidencia científica ha generado controversia entre los investigadores que trabajan en este ámbito (en parte porque el alto listón evidenciario ha limitado mucho el número de autorizaciones otorgadas hasta el momento). La controversia se centra en las ventajas y limitaciones de las diferentes metodologías científicas usadas para la generación de datos en la investigación sobre alimentos, que posteriormente pueden convertirse en la base de las decisiones reguladoras.

En nuestro análisis hemos identificado la concurrencia de varios tipos de intereses: por un lado, el objetivo de la EFSA de proteger a los consumidores de

---

<sup>1</sup>Trabajo financiado por Secretaría de Estado de Investigación, Desarrollo e Innovación del Gobierno de España: Proyecto de Investigación: Estándares de prueba y elecciones metodológicas en la fundamentación científica de las declaraciones de salud (FFI2017-83543-P).

posibles declaraciones falsas; por otro, el posible freno a la innovación tecnológica en el ámbito de los alimentos, y al desarrollo de un mercado para alimentos con declaraciones de salud.

Nuestro análisis muestra que parece poco probable que la exigencia de los reguladores de un nivel muy alto de evidencia (datos causales) evitará el debate, no sólo académico sino también público, respecto de la autorización de las declaraciones de salud. No está claro que la estrategia de la EFSA obtendrá la confianza de los consumidores. Esto es porque las decisiones sobre las metodologías y niveles de exigencia evidenciaria (lo que denominamos la “política evidenciaria” elegida por la EFSA) no han sido consensuadas con los diferentes actores sociales, a pesar de los importantes efectos no epistémicos que la elección de una determinada política evidenciaria puede acarrear. En el caso de las declaraciones de salud, como hemos visto, estos efectos no epistémicos se pueden encontrar particularmente en el ámbito de la salud pública (e individual). Una mejor estrategia consistiría en intentar consensuar las políticas evidenciaras con los diferentes actores sociales.

SIMPOSIO:  
THE BODILY SELF

# Introduction

**Michele Palmira**

**Víctor M. Verdejo**

Logos Research Group/BIAP (Barcelona Institute for Analytic Philosophy)  
University of Barcelona

Descartes famously maintained that introspection teaches us that we are thinking beings. Hume, however, observed that we do not introspect anything like a substantial self over and above particular mental states, such as sensations. These two different perspectives focus on the contention that we can gain fundamental insight into the nature of the self by looking at whether, and if so how, a specific judgement-forming method, viz. introspection, puts us in a position to discover our selves.

This purely mentalist approach to the nature of the self, however, has struck many as incomplete. Philosophers such as Gareth Evans, Thomas Nagel, and P. F. Strawson have countered Hume's sceptical stance on the self by urging that the very possibility of self-consciousness rests on our thinking of ourselves in a first-person way *qua* spatially located bodies and hence opening the door to the consideration of a *bodily or embodied self*.

This symposium will investigate the nature and philosophical significance of the bodily self. Taking into account the latest developments on the topic, it will address the following questions: What is the relation, if any, between a mentalist and a bodily picture of the self? What special epistemic properties, if any, are exhibited by self-ascriptions of bodily properties, as opposed to self-ascriptions of mental properties? What is exactly the connection between first-person thought, self-representation, and bodily self-awareness? How does crossmodal bodily self-perception result in an integrated first-person conception of ourselves?

In "The Inside-Out Binding Problem" Léa Salje (University of Leeds) attempts at strengthening the very idea of there being observational awareness of oneself whereby we can ground a substantive conception of the self which does not fall prey to Hume's scepticism. According to an influential strand of thought originally due to Gareth Evans and P. F. Strawson, we can have perceptual awareness of ourselves, as ourselves, in virtue of bodily self-perception. To this end, philosophers have distinguished between getting information about one's own body in a first-person (interoceptive) and third-person (exteroceptive) way. However, in this paper Salje argues that a *de se* conception of oneself as an object spatially located among others encompasses both of these (interoceptive and exteroceptive) ways of being bodily aware of oneself at once. Hence, the author offers a clear formulation of what can be arguably seen as one of the central challenges that Anti-Cartesian conceptions of the self have to face in order for them to make room for a substantive conception thereof. In the second part of the talk, Salje offers an as yet unexplored defence of the idea that both

exteroception and interoception can contribute to determining an integrated bodily conception of ourselves.

Frédérique de Vignemont (Institut Jean-Nicod) deals in “The Phenomenal Contrast of Ownership” with one of the main bones of contention in the debate on how to account for the sense of bodily ownership, namely, whether one feels one’s body as one’s own in virtue of a *sui generis* phenomenal feature of one’s bodily experiences or not. The debate has reached a stalemate between advocates and defenders of a rich view of phenomenology. This contribution proposes a new way to get out of the impasse which consists in harnessing the method of phenomenal contrast. This method has been primarily introduced to shed light on the nature perceptual experiences by focusing on their phenomenal character, namely the what is like-ness of the experiences. Some authors (e.g. García-Carpintero) have recently extended the phenomenal contrast method to explain why certain kinds of action require distinctively first-personal thoughts. In this talk, de Vignemont offers a novel application of the method by deploying it in the case of bodily experiences to show whether they exhibit a distinctive first-person character, or not.

Finally, the panel is completed with “Ownership and Immunity” by Carlota Serrahima (University of Barcelona). The paper clarifies a largely overlooked distinction between a sense of experience ownership and a sense of bodily experiences. This enables Serrahima to pursue a comparison between mental and bodily self-ascriptions under a new light. Importantly, she takes up the question of what the right order of philosophical explanation of the specialness of the first-person perspective is and focuses on the phenomenon of immunity to error through misidentification relative to the first-person concept. The phenomenon arises for any kind of introspection-based mental self-ascriptions, such as “I am thinking that it is sunny”. The idea is that I cannot be mistaken that it is *I* who is instantiating the relevant mental property. Noticing that even bodily self-ascriptions can be immune to error through misidentification, the author defends the contention that without an account of the sense of experience ownership and the sense of bodily ownership, we cannot give an account of immunity to error through misidentification. On these grounds, Serrahima will go on to uncover as yet unexplored dependence relations between the sense of experience ownership and the sense of bodily ownership and work out the implications that such dependency might have for the question of immunity.

The organisation of this symposium has received the support of the Secretary for Universities and Research of the Department of Economy and Knowledge (Government of Catalonia) as well as the Ministry of Economy and Competitiveness (Government of Spain) and the European Union through the research projects FFI2016-80588-R and FFI2015-63892-P (MINECO, AEI/FEDER, EU).

# The phenomenal contrast of ownership

Frédérique de Vignemont

Institut Jean-Nicod

Anscombe (1962) famously claimed that we do not feel our legs as being crossed; we simply know that they are that way. What about the rest of the knowledge that we have of our body, and more specifically the knowledge of high-level bodily properties such as the sense of bodily ownership? The debate is structured between those who defend a liberal or rich view of phenomenology (Vignemont, 2018; Gallagher, 2017; Peacocke, 2014), and those who defend a sparse and thus conservative view (Alsmith, 2015; Bermúdez, 2011, 2015; Martin, 1992, 1995; Wu, forthcoming).

One cannot settle this debate by a direct use of introspection in everyday life but one may be able to apply the method of phenomenal contrast proposed by Siegel (2010). It proceeds in two steps. First, one describes a situation in which there is intuitively a phenomenal contrast between two experiences, one of which only instantiating the high-level property. For instance, it has been suggested that it does not feel the same when one hears a sentence in a foreign language before and after learning to speak the language. The second step consists in drawing an inference to the best explanation of this contrast by ruling out alternative explanations.

Most interest in the debate on the degree of richness of perceptual content has focused on visual awareness but it may as well be applied to bodily awareness. The only difference here is that it is more difficult to find scenarios in which one is not aware that this is one's own hand than scenarios in which one is not aware that this is a pine tree. Nonetheless, there are some borderline cases in which one can manipulate the awareness of this specific high-level property. The questions we need to ask then are the following. Does it feel different when I am aware that the hand is my own and when I am not? And if it does, is it because of a phenomenal property of ownership, because of attention, or because of other phenomenal properties of bodily awareness?

## References

- Alsmith, A. J. T. (2015). Mental Activity & The Sense of Ownership. *Review of Philosophy and Psychology*, 6(4), 881–896.
- Anscombe, G. E. M. (1962). On Sensations of Position. *Analysis*, 22, 55–58.
- Bermúdez, J. L. (2011). Bodily Awareness and Self-Consciousness. In: S. Gallagher, ed., *The Oxford Handbook of the Self*. New York: Oxford University Press, 157–179.

- Bermúdez, J. L. (2015). Bodily Ownership, Bodily Awareness and Knowledge without Observation. *Analysis*, 75(1), 37–45.
- Gallagher, S. (2017). Self-defense: Deflecting Deflationary and Eliminativist Critiques of the Sense of Ownership. *Frontiers in Psychology*, 21.
- Martin, M. G. F. (1992). Sight and Touch. In: T. Crane, ed., *The Content of Experience*. Cambridge: Cambridge University Press, 199–201.
- Martin, M. G. F. (1995). Bodily Awareness: A Sense of Ownership. In: J. L. Bermúdez, T. Marcel, N. Eilan, eds., *The Body and the Self*. Cambridge (Mass.): MIT Press, 268–289.
- Peacocke, C. (2014). *The Mirror of the World*. Oxford: Oxford University Press.
- Siegel, S. (2010). *The Contents of Visual Experience*. Oxford: Oxford University Press.
- de Vignemont, F. (2018). *Mind the Body*. Oxford: Oxford University Press.



# The inside-out binding problem

Léa Salje

University of Leeds

The idea that there is something about self-conscious thought that requires a capacity to think of ourselves, *as* ourselves, as objects spatially located among others is a theme that finds echoes across the work of P. F. Strawson (1966), Thomas Nagel (1986), Gareth Evans (1982), and more recently John Campbell (1994), Quassim Cassam (1997) and J. L. Bermúdez (1995, 2017)—variously developed, but always returning to the basic idea that there are demands on self-conscious thought that require us to conceive of ourselves, in a *de se* way, as objects spatially located among others. How might we develop such a self-conception? One plausible suggestion is by *perceiving* oneself as such. That is, by bodily self-perception.

For the animals that we are, our modes of bodily self-perception can roughly be grouped in two: the *exteroceptive* senses like vision, touch and smell, through which one's body is perceived as an object among others; and the *interoceptive* senses like proprioception, kinaesthesia and nociception, through which a subject can perceive her own body and its parts only. Through interoception one perceives a body *as one's own*, while through exteroception one perceives a body as one worldly object among many. In this talk I ask what grounds our perceptual sensitivity to the sameness of the body (or body-parts) perceived through both sets of senses, such that such crossmodal bodily perception can feed into our *de se* conceptions of ourselves as ordinary worldly objects.

## References

- Bermúdez, J. L. (1995), 'Ecological Perception and the Notion of a Nonconceptual Point of View', in *The Body and the Self*, N. Eilan, J. L. Bermúdez and T. Marcel eds., MIT press.
- Bermúdez, J. L. (2017), *Understanding "I"*, Oxford University Press.
- Campbell, J. (1994), *Past, Space and Self*, MIT Press.
- Cassam, Q. (1997), *Self and World*, Oxford University Press.
- Strawson, P. F. (1966/1990), *Bounds of Sense*, Routledge.
- Evans, G. (1982), *The Varieties of Reference*, Oxford University Press.
- Nagel, T. (1986), *The View from Nowhere*, Oxford University Press.

# Ownership and immunity

**Carlota Serrahima**

Logos Research Group/BIAP (Barcelona Institute for Analytic Philosophy)  
University of Barcelona

Phenomenally conscious mental states are such that, may the subject that undergoes them have the capacity to make judgments in which she reports them, she will use a first-person pronoun in the subject position seemingly compellingly. Bodily experiences, in turn, are a type of mental state such that subjects endowed with a conceptual system or language express by qualifying their content, namely the felt body, with a first-person indexical, in an equally seemingly compelling way.

We can express this by saying that subjects typically have a *sense of experience ownership* (SEO) for the experiences they are phenomenally conscious of (e.g. Zahavi & Kriegel, 2015; Dainton, 2016; Howell & Thompson, 2016; Nida-Rümelin, 2016), and a *sense of bodily ownership* (SBO) for the body they feel in bodily experiences, respectively (e.g. Martin, 1995; Dokic, 2003; de Vignemont, 2007, 2018; Bermúdez, 2017).

Besides, judgments of mental self-ascription made on phenomenal grounds are typically said to be (logically) *immune to error through misidentification* (IEM; Shoemaker, 1968). In turn, judgments of bodily self-attribution made on the grounds of bodily experiences are typically said to be (de facto) *immune to error through misidentification* (IEM; Evans, 1982).

In this talk I aim at defining the connections between SEO, SBO and IEM for mental and bodily self-ascriptions. Firstly, I will notice that, given that IEM is a feature of judgments relative to their grounds, an explanation of this phenomenon both for mental and bodily self-ascriptions depends on, and in this sense is less fundamental than, an account on the SEO and the SBO respectively. Secondly, I will explore the prospects of a dependence of the SBO on the SEO, and the implications that this dependence would have for an explanation of both mental and bodily IEM.

## References

- Bermúdez, J. L. (2017). Ownership and the Space of the Body. In: F. de Vignemont and A. J. T. Alsmith, eds., *The Subject's Matter: Self-Consciousness and the Body*. Cambridge (Mass.): MIT Press.
- Dainton, B. (2016). The Sense of Self. *Aristotelian Society Supplementary Volume*, XC, 113–143.

- de Vignemont, F. (2007). Habeas Corpus: The Sense of Ownership of One's Own Body. *Mind & Language*, 22(4), 427–449.
- Dokic, J. (2003). The Sense of Ownership: An Analogy between Sensation and Action. In: J. Roessler and N. Eilan, eds., *Agency and Self-Awareness: Issues in Philosophy and Psychology*. New York: Oxford University Press, 320–344.
- Evans, G. (1982), *The Varieties of Reference*, Oxford University Press.
- Howell, R. J. & Thompson, B. (2016). Phenomenally Mine: In Search of the Subjective Character of Consciousness. *Review of Philosophy and Psychology*, 8(1), 103–127.
- Martin, M. G. F. (1995). Bodily Awareness: A Sense of Ownership. In: J. L. Bermúdez, T. Marcel, N. Eilan, eds., *The Body and the Self*. Cambridge (Mass.): MIT Press, 268–289.
- Nida-Rümelin, M. (2016). Self-awareness. *Review of Philosophy and Psychology*, 8(1), 55–82.
- Shoemaker, S. (1968). Self-reference and Self-awareness. *The Journal of Philosophy*, LXV(19), 555–567.
- Zahavi, D., and U. Kriegel. (2015). For-me-ness: What it is and what it is not. In: D. O. Dahlstrom, A. Elpidorou, and W. Hopp, eds., *Philosophy of mind and phenomenology: Conceptual and empirical approaches*. London: Routledge.