



HAL
open science

Naturalness, Scientific Concepts, and the Substantivity of Social Metaphysics

Igor Douven

► **To cite this version:**

Igor Douven. Naturalness, Scientific Concepts, and the Substantivity of Social Metaphysics. *Philosophia*, 2024, 52, pp.849 - 863. 10.1007/s11406-024-00776-8 . hal-04859882

HAL Id: hal-04859882

<https://hal.science/hal-04859882v1>

Submitted on 31 Dec 2024

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.



Naturalness, Scientific Concepts, and the Substantivity of Social Metaphysics

Igor Douven¹

Received: 10 October 2023 / Revised: 14 August 2024 / Accepted: 22 August 2024 /
Published online: 25 September 2024

© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

We argue that concepts from the social sciences can be as natural as those from physics and chemistry, thereby answering in the positive the question of whether social metaphysics is or can be substantive. The argument takes as a starting point Douven & Gärdenfors' (*Mind & Language*, 35, 313–334 2020) optimality account of natural concepts, according to which natural concepts are represented by the cells of an optimally partitioned similarity space. While the account applies straightforwardly to perceptual concepts, it does not apply as straightforwardly to more abstract concepts, such as many scientific concepts. We propose an extension of Douven and Gärdenfors' optimality account that pertains to scientific concepts as well. The proposal reuses ideas from Lewis' Best Systems Account of laws of nature, but rather than *invoking* natural concepts (as Lewis does) *defines* natural scientific concepts as those that figure in the or a best systematization of the totality of facts. On this proposal, there is nothing especially different between concepts from the natural sciences and those from the social sciences.

Keywords Best systems account · Concepts · Naturalness · Similarity spaces · Social metaphysics

1 Introduction

There has been a recent surge of interest in social metaphysics (see, e.g., Khalidi 2015; Asta 2018; Griffith 2020; Taylor 2020; Taylor 2023; Bendifallah et al. 2023). While currently popular, however, social metaphysics faces a special challenge (see, e.g., Taylor 2023). According to a widely shared assumption, for a metaphysics to be substantive, it must deal with a reality whose structure is *objective*. But whereas the natural sciences study parts of the world and processes occurring in the world—

✉ Igor Douven
igor.douven@univ-paris1.fr

¹ IHPST/CNRS/Panthéon-Sorbonne University, Paris, France

biological species, chemical elements, phase transitions, and so on—that appear to be uncontroversially objective in that these parts and processes are not of our own making, the same cannot be said of the objects of study of social scientists. Indeed, the issues of interest to social scientists—class, race, gender, market cycles, national debt, inflation, unemployment, oppression, and much more—seem *clearly* dependent on us. Should we conclude that social metaphysics fails to be substantive, thereby (presumably) demoting it to a merely second-rate status?

The proposal to be presented in the following allows us to answer the foregoing question in the negative and, in effect, to maintain that social metaphysics is on a par with the metaphysics we take to emerge from the natural sciences. At the core of this proposal is a rethinking (begun in Douven (2023), building on Douven and Gärdenfors (2020) of what natural kinds are and what their metaphysical status is, which allows social concepts to be no less natural than the concepts central to physics or chemistry. In a nutshell, the two-fold claim is going to be that, *yes*, social concepts do depend on our thinking, but so do the concepts central to the natural sciences; and *no*, neither for the former nor for the latter does that imply that they lack objectivity and that hence a metaphysics concerned with either must fail to be substantive.

The background of our proposal is the conceptual spaces framework. This means, among other things, that we think about concepts as regions in similarity spaces (or as sets of such regions). We explain what naturalness amounts to for concepts thus understood, and think of natural kinds derivatively, as the worldly correlates of natural concepts. After outlining the basics of the conceptual spaces framework, we summarize a proposal from Douven and Gärdenfors (2020), which appeals to various principles of good engineering in order to demarcate the natural concepts from the nonnatural ones. The new work starts with noting that Douven and Gärdenfors' principles make sense for perceptual concepts, with which these authors were concerned, but do not apply readily, nor are readily extended to apply, to nonperceptual concepts, including (most notably, for our present purposes) concepts from the social sciences. To overcome this limitation, we recruit ideas central to Lewis' (1973) Best Systems Account (BSA) of laws, which understands lawhood by reference to a best summary of the totality of basic facts. Lewis' ideas were motivated by very much the same considerations that led Douven and Gärdenfors to their proposal, and they will help to extend that proposal to scientific concepts, including those from the social sciences.

To forestall misunderstanding, it should be emphasized that the new proposal is in no way meant to improve upon Lewis' BSA or to address concerns that some have raised about it. We reuse some Lewisian ideas for our own purposes, which are very different from Lewis'. As already mentioned, we take *concepts* to be primary and wish to delineate natural from nonnatural concepts, seeing natural kinds as *deriving* from natural concepts. For Lewis, by contrast, *properties* are basic. Not only that; he takes naturalness as it applies to properties as a primitive notion not standing in need of any explication. Thus, in particular, we do—in Lewis' view—not need the BSA to define naturalness. We do think naturalness (for concepts, and so derivatively for kinds) needs to be explicated and also see Douven and Gärdenfors' (2020) work as a first step in this explication, which defines naturalness for perceptual concepts in terms of a trade-off between informativeness and simplicity, in a way to be made more precise below. Ideas from the BSA, suitably adapted, will help to argue that

we can similarly think of naturalness for scientific concepts (and again, derivatively, for scientific kinds) as arising from a trade-off between those same theoretical virtues. This will yield a metaphysics of kinds that is more in line with Putnam's (1981) internal realism than with the metaphysical realism championed by Lewis.

2 Theoretical Background

2.1 The Conceptual Spaces Framework

The core idea underlying the conceptual spaces framework is that concepts can be represented geometrically, as regions in similarity spaces. Similarity spaces are mathematical constructs, often multi-dimensional metric spaces, where each dimension encodes some fundamental quality in terms of which items can be compared to each other. The distance defined on such a space is meant to reflect dissimilarity; a smaller distance indicates a greater similarity in the respect corresponding to the space, which could be color, taste, shape, and so on. While various metrics could be used for this, the prevalent ones in the psychological literature are the city-block (or "Manhattan") metric, which adds distances along each of the dimensions, and the Euclidean metric, familiar from high school geometry.

Similarity spaces are often constructed on the basis of similarity judgments, confusion probabilities (or, more generally, data on how frequently two distinct stimuli are mistaken for one another), or correlation coefficients (e.g., Douven 2021; Bendifallah et al. 2023). This raw data typically undergoes a transformation into distances, followed by application of methods like multidimensional scaling (MDS) or its variants (such as Principal Component Analysis or Nonnegative Matrix Factorization), aiming to generate a metrical space that is interpretable and fits the data while being low-dimensional (see, e.g., Abdi and Williams 2010; Borg and Groenen 2010; Hout et al. 2013).

While MDS and related techniques can be illuminating, they do not always yield a low-dimensional representation that aligns well with the data. However, numerous usable similarity spaces have been identified by cognitive psychologists. Not only that; psychologists have also developed ways to turn similarity spaces into *conceptual* spaces, mapping not just similarities but also concepts and conceptual connections.

Ideas vary as to how to build a conceptual space on top of a similarity space. A prominent approach integrates prototype theory with Voronoi tessellations (Gärdenfors 2000, 2014). The former holds that, within a concept, some instances are more representative, with the most representative being its prototype (Rosch 1973, 1978). A Voronoi tessellation partitions a space into discrete cells, each linked to a unique generator point and encompassing all points nearer to its own generator point than to any other.¹ In this approach, conceptual spaces emerge from similarity spaces by identifying in them the locations of the prototypes and using these to create a Voronoi tessellation, the cells of which then represent concepts in the space. An alternative approach to obtaining a conceptual space from a similarity space is to be found in

¹ For an in-depth explanation, see Okabe et al. (2000)

Bendifallah et al. (2023), where the authors use a standard clustering algorithm to partition a social similarity space. We come back to this below.

2.2 What are Natural Concepts?

From the onset of the conceptual spaces framework, it was understood that not every region in a similarity space will correspond to a concept, in any event not a *natural* one, like GREEN, GOLD, or TIGER—concepts which merit linguistic labeling and play, or could play, a pivotal role in our cognitive processes—as opposed to, for instance, GRUE (Goodman, 1955), or THINGS THAT ARE CURRENTLY ON YOUR MIND, or COMPANIES THAT WENT BROKE DURING THE PANDEMIC. So, what differentiates natural concepts from their nonnatural counterparts? Gärdenfors (2000, p. 71) posited that only *convex* regions can represent natural concepts, where a region is convex exactly if any two points within it can be connected by a line segment that also lies entirely within the region. However, Gärdenfors acknowledged that convexity might not be a sufficient condition. And indeed, as noted by Okabe et al. (2000, p. 58), every Voronoi tessellation of a Euclidean space, regardless of whether it is generated by a set of prototypes or by a set of arbitrarily chosen points, will yield convex cells.

A more detailed characterization of natural concepts was later explored by Douven and Gärdenfors (2020), who introduced the idea of optimal design, grounded in cognitive economy. Gärdenfors (2000, p. 70) already emphasized the cognitive simplicity of dealing with convex regions, but Douven and Gärdenfors (2020) expanded on this by suggesting a suite of criteria meant to embody optimal partitioning. They propose to understand the notion of optimal partitioning in terms of design criteria, a proposal which was inspired by work in biology explaining the existence of biological traits or processes in organisms, or the workings of biological networks, by reference to good engineering design. For example, Alon (2003) notes that principles of modularity, robustness to component tolerances, and use of recurring circuit elements that are found in many well-designed engineered networks are also present in many biological networks. Similarly, Nowak (2006) argues that a particular observed pattern of cell division in the human intestine appears as the perfect design for diminishing the chances of mutations in tumor suppressor genes and oncogenes.

In Douven and Gärdenfors (2020) proposal, optimal design pertains not to biological systems but to cognitive ones, specifically, systems that are to operate under the same pressures we are operating under, needing to succeed in a constant competition for scarce commodities, and limited in the way we are (most notably, having limited memory and limited discriminatory powers). As a first step, they fill in the notion of optimal design by stating a number of broad design goals, such as that the system is to be provided with a rich arsenal of concepts, that the scheme of concepts be such that it not strain the system's memory too much, that the scheme be easy to learn, and that it minimize the risk of making classification errors.

A next step then consists of making these goals more specific, assuming the conceptual spaces framework (so assuming, for instance, that conceptual schemes are families of concepts represented in similarity spaces). To do so, Douven and Gärdenfors propose a number of design principles meant to guide the construction of

conceptual spaces on the basis of similarity spaces. For instance, they argue that a conceptual system should be *informative*, facilitating adequate distinctions within the relevant similarity space. Yet, due to memory constraints, this informativeness must be weighed against *parsimony* to avoid overwhelming memory with exceedingly detailed distinctions. Additionally, according to the principle of *Contrast*, the conceptual system should be such that prototypes can be located at a sufficient distance from each other, to make them readily discernible from one another. At the same time, it should be possible to place them centrally enough in the concept they are supposed to represent for them to be indeed *representative* of items falling under that concept; this is called the *Representation* principle. According to the principle of *Learnability*, the conceptual system should also be easy to learn.

The most central claim in Douven and Gärdenfors (2020) proposal is that natural concepts reside within optimally partitioned similarity spaces, spaces, that is, which satisfy the outlined rational constraints. More explicitly, given that the said constraints can, and typically will, pull in different directions, the optimal partition should offer a best trade-off of the various constraints, aiming for Pareto-optimality. (A solution achieves Pareto-optimality when improving one objective invariably compromises another.)

Gärdenfors (2000) had already pointed at evidence for his convexity constraint (see also Jäger 2007, 2010), and Douven and Gärdenfors (2020) further demonstrate that there is backing for a broader optimal design view of naturalness, especially concerning Informativeness and Parsimony (see studies by Cook et al. 2005; Regier et al. 2007; 2009; Kemp and Regier 2012; Xu and Regier 2014; Xu et al. 2016). Douven (2019) presents evidence specifically for Contrast and Representation, and Douven (2024a, 2024b) does the same for the principle of Learnability.

2.3 What Remains of Objectivity?

Douven (2023) proposes to view natural *concepts* as primary and conceive of natural *kinds* as the worldly correlates of natural concepts, where the latter are understood in terms of optimally partitioned similarity spaces, in the manner of Douven and Gärdenfors (2020). This raises an immediate concern about the objectivity of natural kinds, which are traditionally thought of as constituting the fundamental, mind-independent building blocks of reality. The concern is that on our proposal natural kinds apparently *do* depend on our minds, if only because the idea of an optimal partition refers to *our* cognitive makeup.

As Douven (2023) points out, his proposal is in some ways close to Putnam's (1981) internal realism, a position according to which the human mind is actively involved in shaping the world, without however simply making up the world.² Internal realism never gained much traction, in large part due to its being perceived as a fancy reinstallation of relativism, where in the end what is true depends on what conceptual scheme we decide to adopt. Regardless of whether that is a fair assessment of Putnam's

² Decock and Douven (2012) had already shown how the conceptual spaces framework can be used to make internal realism, and in particular the notion of a conceptual scheme on which the position relies, more precise.

position, it is to be emphasized that our proposal is sufficiently different from Putnam's in ways that should make it evident that the former is compatible with natural kinds being objective even if the latter is not (which we leave open).

Two differences between Douven's (2023) proposal and Putnam's position are particularly important at this juncture. First, while Putnam's sometimes-held view that truth coincides with rational acceptability in the Peircean limit of science leaves little room for holding that there may be a gap between how theories we hold in that limit *represent* the world and the way the world actually is, it is perfectly consistent with our proposal that even such theories miss their mark entirely. That the notion of an optimal partitioning refers to our cognitive powers and limitations does not in any way guarantee that the concepts we use are all natural, or even that any of them are natural, nor that this will be the case in the Peircean limit of science. The key observation to make here is simply that, in our quest of an optimal partitioning, there is nothing to prevent us from getting forever stuck in a local optimum. Even if stuck in such an optimum, we might well go on to develop theories that are empirically adequate. It is just that the conceptual apparatus those theories rely on could be improved upon, in the sense that there would then exist conceptual schemes that do better, on balance, in light of Douven and Gärdenfors' criteria of naturalness.

A second difference between Douven's (2023) proposal and an internal realist metaphysics of kinds is related to the intuition that natural kinds are, in Lewis' (1986) terminology, *sparse*. For Putnam, every conceptual scheme that serves our purposes is acceptable. There is nothing in his writings to suggest that there might not be an abundance of such schemes and so if we declare natural kinds to be the ones recognized by some acceptable conceptual scheme, natural kinds could end up being abundant as well. That is not so on the alternative proposal considered here. Almost all partitionings of a similarity space will result in a conceptual space that is non-optimal in Douven and Gärdenfors' sense, and so we may expect natural kinds to still be sparse on this proposal.³

More generally, note that the mind-dependence that follows from our proposal does not entail that, at any point, there is something for us to choose or decide. The similarity spaces are what they are—the shape of color space, for instance, depends on our perceptual apparatus, which is not of our own making. And our cognitive capacities and limitations, which in the end determine how similarity spaces are best furnished with concepts, are not under our control either. So, in particular, to say that natural concepts are mind-dependent in the sense of Douven and Gärdenfors (2020) does not imply that it is dependent on our thinking and theorizing what the natural concepts are.

3 Social Concepts

So far, most of the work conducted in the conceptual spaces framework has concerned spaces for representing *perceptual* concepts, related, for instance, to the perception of colors, vowels, smells, or tastes. However, the framework is not intrinsically limited to

³ See Douven (2023) for more on this.

such concepts. And indeed, there is *some* work on more abstract concepts as well, such as physical concepts (e.g., Gärdenfors and Zenker 2011, 2013) and moral concepts (Verheyen & Peterson 2021). In this paper, we are mainly interested in *social* concepts, which until very recently researchers working within the conceptual spaces paradigm have paid scant attention to. Interestingly, however, social scientists appear to have been using conceptual spaces for some time, albeit not under that name, and apparently unaware of the work on conceptual spaces being done in cognitive science (just as cognitive scientists were unaware of the said research in social science).

The French sociologist Bourdieu pioneered the use of spaces for representing social structures and phenomena. Bourdieu (1979; see also his 1989) is a study of how the distribution of social and economic goods relates to cultural, moral, and political practices and preferences. A central tool in this study is what he refers to as *social spaces*, whose dimensions encode economic, cultural, and social capital relative to which individuals can be represented in the space. In this space, proximity indicates similarity, now in the respect of (broadly) socio-economic status. Bourdieu's spatial approach was taken on board by various other French social scientists. Here we mention in particular (Deauvieu et al., 2014) work, which was inspired by Bourdieu's, and which in turn inspired (Bendifallah et al., 2023) to connect work on social spaces carried out in the social sciences with the conceptual spaces framework as developed in cognitive science, applying the latter to concepts central to French feminist theorizing by using experimental methods developed in the context of the former.

Specifically, Bendifallah et al. (2023) used a French mailing list for researchers working in gender studies to recruit participants who were most likely knowledgeable about current work on feminism in France. The authors presented to their participants a list of 31 names of feminist authors and activists the participants could be expected to be familiar with. Then, using Deauvieu et al. (2014) design, they asked participants to sort the names they recognized into at most nine groups, also asking participants to provide an appropriate label for each group they created.

In their analysis, Bendifallah and co-authors calculated correlation coefficients for the feminists in their materials on the basis of co-grouping profiles (e.g., how often the author Beauvoir and the politician Hidalgo had been grouped together, and with which other feminists exactly). In their turn, these coefficients served as input for a classical MDS procedure, which resulted in a similarity space that met various goodness-of-fit criteria for such spaces.⁴ Finally, they used a popular clustering technique to carve up the similarity space into concepts. Specifically, they used Partitioning Around Medoids, which implements the idea that items falling under a concept should be as similar as possible to each other while at the same time being as dissimilar as possible to items falling under any of the other concepts represented in the space (Kaufman & Rousseeuw 2005).⁵ Interestingly, this technique works by creating a Voronoi tessellation of the space it is applied to and thus also guarantees convexity of the resulting concepts. The full conceptual space Bendifallah and co-authors arrived

⁴ We will not discuss these here. See, e.g., Dexter et al. (2018) for details.

⁵ The idea referred to here is what Douven and Gärdenfors (2020, Sect. 4.2.3) call "Well-formedness." As Douven and Gärdenfors (*ibid.*) note, this idea is implemented, in slightly different ways, by various clustering algorithms. See Bendifallah et al. (2023) for their specific reasons for choosing Partitioning Around Medoids.

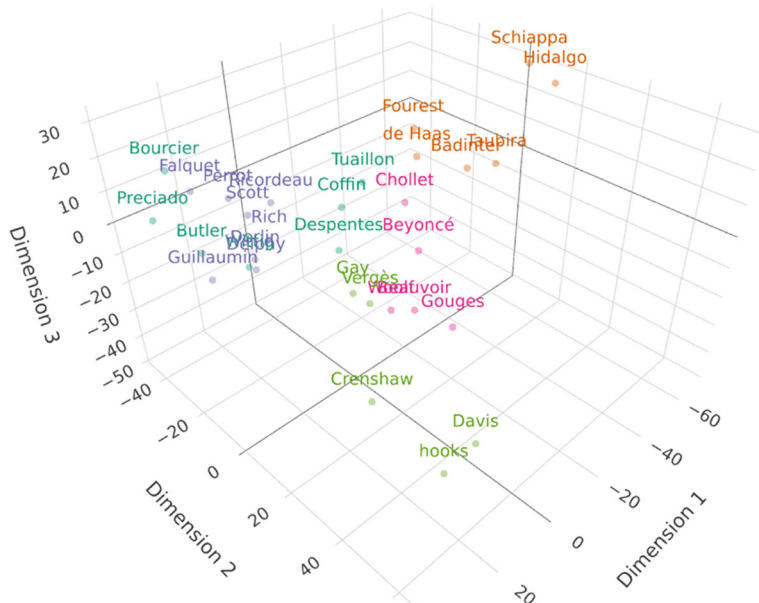


Fig. 1 The three-dimensional space reported in Bendifallah et al. (2023) representing concepts prevalent in French feminist theory. (Concepts are distinguished from each other by using different colors for items falling under them)

as is shown in Fig. 1. Using information from the labels participants had applied to the groups they created, the authors were able to link the concepts as emerging from their analysis to ones prevalent in debates about feminism, such as MATERIALIST, INTERSECTIONAL, and QUEER.

This is proof of principle that we can construct conceptual spaces for scientific concepts—in the above case, concepts pertaining to the social sciences—in basically the same way in which we can construct a space for color concepts, or shape concepts, or gustatory concepts.⁶ That, of course, is not enough to conclude that the scientific concepts thereby arrived at also merit the status of *natural* concepts. To be sure, in the case of Bendifallah et al. (2023) study, the concepts that emerged from their analysis were all represented by convex regions in the relevant similarity space. That was guaranteed by the clustering technique these authors used. Recall, however, that convexity is not enough for naturalness; we can easily come up with families of convex concepts that nonetheless all strike us as gerrymandered and not natural at all.

Douven and Gärdenfors (2020) proposed a number of criteria going beyond convexity. Does Bendifallah et al.’s conceptual space satisfy all of those, so that we could claim that it offers an optimal partitioning of the similarity space that came out of the MDS procedure these authors conducted? Here, a problem is that Douven and Gärdenfors (2020) were mainly concerned with *perceptual* concepts, not scientific concepts (or, better: scientific concepts more broadly understood, given that perceptual concepts can be said to be scientific as well). And indeed, the optimality criteria

⁶ It is proof of principle only, because it remains to be seen how broadly applicable (Bendifallah et al.’s 2023) approach is in science, or even just in the social sciences.

they propose do not readily apply to scientific concepts, at least not all of them. For instance, for such concepts it will often not be our first concern whether they can be quickly learned, or learned from few instances, or are easily retrievable from memory, or are easily communicable, or have easily discernible prototypes (if they have prototypes at all). As is shown in Galison's (1997, Ch. 8) work on how theoretical and experimental physicists communicate with each other in their collaborations, scientific concepts can be so complex and hard to grasp that scientists from different disciplines (e.g., theoretical and experimental physicists) will often be able to communicate only by using impoverished or simplified concepts. As Galison also shows, the use of such "pidgin" concepts is not necessarily an impediment to successful research; it is very often a prerequisite. In short, the question of how we are to demarcate bona fide scientific concepts from gerrymandered ones is not readily provided by Douven and Gärdenfors' optimal design principles. The following proposes an answer to this question that retains the spirit of Douven and Gärdenfors' optimality approach by adapting Lewis' Best Systems Account.

4 Best Systems and Naturalness

Lewis (1973) asks us to consider the "Humean mosaic," which is the "world's total history of manifest matters of particular fact," by which, more specifically, he means the spatiotemporal arrangement of instantiations of locally occurring properties, so, excluding causal connections among those instantiations as well as modal properties. A *best system* is a "summary" of that mosaic that optimally balances informativeness and simplicity. This notion helps Lewis explain what a law of nature is, and in particular what distinguishes laws of nature from "mere" regularities, viz., laws of nature are those regularities that are entailed by a deductively closed best system all of whose theorems are true (i.e., the summary is correct). Thus, this Best Systems Account (BSA) offers a reductive and non-modal account of lawhood, grounding laws of nature in the mosaic of particular facts in conjunction with certain optimality criteria.

Lewis realizes that a worry one may have about his proposal is that if we allow any predicate or property to be used in formulating a best system, we could end up with gerrymandered or gruesome systems that, while simple and strong, do not match our intuitions about what the laws of nature should be. Lewis' response to this concern is to point at an additional constraint for best systems, to wit, that they must refer to *natural* properties, which he conceives as the fundamental or basic properties that form the building blocks of reality. By introducing natural properties, Lewis aimed to constrain the BSA. The idea is that the best system should be formulated using only, or primarily, natural properties. This is supposed to help rule out systems based on gerrymandered or gruesome predicates. Indeed, *if nature is kind to us*, there may be no more than one best system after all; that is to say, the constraints might be such that only one system satisfies them all.

Lewis takes naturalness more or less for granted. "More or less," because he does make the case that assuming the existence of natural properties offers all sorts of philosophical benefits (Lewis, 1983), thereby in effect mounting an abductive argument in support of natural properties (Douven 2022, Ch. 1). Nominalists might retort that there

is no reason why the world would be friendly to philosophers' theoretical concerns, and they might add to this that it is unclear why pragmatic kinds—kinds of things grouped together precisely because doing so helps to serve our purposes—will not equally accommodate those concerns (which would amount to denying that Lewis has identified the *best* explanation for the phenomena he is interested in).

Admittedly, at least since the work of Putnam (1975) and Kripke (1980), mainstream philosophical thinking is in line with Lewis. As both Kripke and Putnam suggested, the strongest support for the existence of natural kinds (or properties) is coming from science itself, which appears to have been able to uncover the micro-essences of kinds of things that, on a more intuitive basis, we had always thought to belong together naturally.

That Kripke and Putnam, and with them the mainstream, may have erred on this point was forcefully argued by Leslie (2013), who questions the source of our essentialist intuitions and their epistemological status. According to her, such intuitions may stem from cognitive biases developed early in life and may not reflect anything about the true nature of reality. Leslie not only musters a wealth of evidence from developmental and cognitive psychology showing that we have a tendency to misconstrue scientific findings as confirming our essentialist beliefs, she also points at literature challenging the science supposedly backing the idea that natural kinds are defined by their micro-essences. For instance, while for Kripke and Putnam biological species are prime examples of natural kinds and also have scientifically discovered essences (specified in terms of their DNA), Leslie points out that most biologists lean toward the idea that a species is defined by the boundaries of an ecological niche or by the boundaries of a reproductive community, where there can be significant genetic variation within those boundaries.⁷

Also, Leslie approvingly cites (van Brakel, 1986), who may have been the first to argue against the commonly held view that chemical kinds can be individuated in terms of micro-essences. As he notes, it would be a gross oversimplification to think that being water amounts to being the chemical substance consisting of H₂O molecules. For instance, in ice, it is not possible to identify individual H₂O molecules, and liquid water contains not only H₂O molecules but also H₄O₆ molecules as well as ions such as OH⁻ and H₃O⁺. Such complexities make it difficult to give a simple account of the micro-essence of water. In van Brakel's (1986) view, macroscopic or manifest properties and behavior determine, in the end, what "water" refers to, and in later research he argues that which of those properties are allowed to serve that purpose is determined by pragmatic factors (see, e.g., van Brakel 1992).

In response to these challenges, we would like to suggest an adaptation of the BSA that does without the naturalness constraint. While on Lewis' account natural kinds are taken for granted, the adaptation suggests a more dynamic interplay between best systems and natural kinds, where natural kinds are derivative from the best system rather than taken as primitive and serving as input. Specifically, the suggestion is that we (i) again begin with a raw mosaic of particular matters of fact, (ii) choose concepts which allow us to group particular facts together, (iii) compare systems of regularities

⁷ For further critique of the Kripke–Putnam view on biological kinds, see Sober (1980); Kitcher (1984), and Griffiths (1999), among many other publications.

stated in terms of these concepts in light of how well they reconcile simplicity and informativeness, and (iv) let the winner identify not only the laws of nature but also the natural concepts.

To further clarify the contrast, on the new proposal natural concepts are still privileged in that they feature in the laws of our best system, and in that sense natural kinds—their worldly correlates—can still be regarded as the fundamental building blocks of reality. But while Lewis postulated natural kinds, letting them serve as one constraint in a constrained optimization procedure, with as its objective finding the best summary of the totality of facts, the new proposal treats natural concepts as variables in the optimization procedure, next to the variables informativeness and simplicity, the objective remaining the same.

Is it not a leap of faith to suppose that this will yield a unique best system? First of all, as pointed out in Section 2.3, natural kinds can be sparse—as they should be, according to realist intuitions—without there being a unique best system. Naturally, it may be said that it is still a leap of faith that there will be at most few best systems—so that we will be able to maintain that natural kinds are sparse—but Lewis' assumption that, with some luck, the original BSA might yield a unique best system was a leap of faith, too, and we do not know how to compare these leaps of faith.

Also, the objectivity of natural kinds is not compromised on our proposal, for reasons pointed out previously: natural kinds result from an optimization process, and while what counts as optimal is related to our cognitive makeup—our cognitive capacities and limitations—nothing in the process is up to us in that we could steer it in this or that direction. Further, with regard to the matter of objectivity, it is to be noted that the position defended here is perfectly compatible with methodological fallibilism (according to which an empirically adequate theory that optimally balances the theoretical virtues may still be false), which is often seen as a hallmark of realism (Douven, 1999). That is because an empirically adequate theory is by definition consistent with all *observable* matters of fact but may be totally wrong about some or all *unobservable* matters of fact.

A number of authors have argued that the choice of primitive concepts must precede Lewis' optimization process (e.g., Loewer 1996; Cohen and Callender 2009). That is because—they hold—especially simplicity is a language-dependent property of theories. Loewer (p. 109) illustrates the point in terms of Goodman's (1955) predicates "grue" and "bleen." Assuming a language containing these predicates as primitives instead of "green" and "blue," we will deem "All emeralds are green" as more complex than "All emeralds are grue"; with "green" and "blue" as primitives, we would reach the opposite conclusion. While the idea that simplicity is an at least partly syntactical notion was favored by the logical empiricists, we do not see any reason to commit to it. Specifically, we would maintain that there are language-independent criteria for judging simplicity, and also informativeness and balance. For example, students may pick up one best system faster than a second, which could be an indication that the former is the simpler of the two. And groups of researchers taking best system T_1 as a basis for their work may, on average, make more progress than groups of researchers using best system T_2 as a basis. That could be an indication that T_1 is more informative, or simpler, or strikes a better balance between the two desiderata, than T_2 .

A related concern that needs to be addressed stems from the fact that, to get the optimization process started, we must have a description of the totality of facts that we are trying to summarize. One may thus wonder whether there must not be a set of

primitive predicates—plausibly corresponding to natural concepts—already in place to get the optimization process going. First off, we already have our perceptual concepts in place, as per (Douven & Gärdenfors’ 2020) proposal. No doubt, we will need a far richer set of concepts to describe the totality of facts. Note, however, that while we must have *some* description of facts to get the optimization process going, in that process the facts themselves may get reconceptualized, if need be over and over again, until we arrive at a description of the Humean mosaic that allows for a systematization that is as simple as can be while also being informative and as informative as can be while also being simple. That old concepts get replaced by new ones, which allow us to formulate better theories, is an utterly common phenomenon in science (e.g., Nersessian, 2008).

To give a simple illustration, consider the Cheyenne concept VOVETAS under which fall most vultures, some hawks, swarms of two types of insects, and tornadoes (van Brakel, 1991). The grouping is obviously not arbitrary: all vovetas are naturally associated with some whirling type of movement. But the concept appears far from ideal if one’s task is to construct a simple yet informative theory. Supposing the Cheyenne have any interest in that, it would be reasonable for them to drop the predicate “vovetas” from their language and introduce words specifically for vultures, hawks, and so on. And—who knows?—later on, in the kind of process mentioned in the previous paragraph, they might see reason to replace these new words with newer ones still.

To recap so far, we have argued that an adaptation of Lewis’ BSA allows us to define natural scientific concepts in terms of optimality—natural scientific concepts are those that figure in a best systematization of the totality of facts—which extends (Douven & Gärdenfors’ 2020) optimality proposal for perceptual concepts. The question now is where this extension leaves *social* concepts. The short answer is that nothing in our new proposal privileges concepts from the natural sciences over those from the social sciences. Exactly how the latter are related to the former is neither here nor there, as far as our proposal goes. But if the best system must be one that is maximally cognitively useful for creatures like us, as is in line with our proposal and as other authors have argued on independent grounds (e.g., Hicks 2018; Dorst 2019; Jaag and Loew 2020), then we should expect social concepts to occur in the best system as much as we do physical concepts. After all, not even the logical empiricists (who were reductionists with regard to the special sciences) will have believed that all of sociology (or psychology) will one day follow from physics in a way that will be understandable for ordinary mortals. (That would not commit us to the view that, for instance, sociology postulates spooky entities: everything could still be said to supervene on the physical.) But perhaps it is more reasonable, here, to side with Schrenk (2008, 2017), Cohen and Callender (2009), and other proponents of the so-called Better Best Systems Account. These authors reject the idea of there being one big best system, covering the whole Humean mosaic, and instead argue that there may be a best system for each of the sciences separately; thus, there may be a best system optimally summarizing the physical facts, one optimally summarizing the chemical facts, and so on. While at least some of these authors were motivated by a skepticism precisely about Lewis’ appeal to natural kinds, there is nothing in their idea itself that should keep us from using it to *define* natural kinds (via our definition of natural concepts). Whichever option we choose—one overarching best system or separate best

systems for the various sciences—the important point is that social concepts come out as natural; indeed, they come out as being as natural as concepts from the natural sciences or as perceptual concepts.

Note that it is still impossible to say whether specific social concepts, like the ones discussed in Bourdieu (1979) or Bendifallah et al. (2023), turn out natural on our account. Equally, however, it is impossible to say for sure that the concepts ELECTRON and SPIN will turn out natural on this account. Our current theories, whether in the natural or in the social sciences, are our best guesses of what parts of the best system, or some best systems, look like. For reasons canvassed in the literature on scientific realism (e.g., Psillos 1999), some optimism is warranted that our guesses are not too far off the mark, and are generally getting closer to that mark as time progresses. But whether we will ever hit the mark, or how close we can get to it, are, at least currently, matters of pure speculation. Again, this is so as much for the natural sciences as for the social sciences, and is thus no reason to hold that social concepts are bound to be of a lesser status than the concepts central to the natural sciences.

5 Conclusion

We looked at the nature of natural concepts, taking as our starting point (Douven & Gärdenfors' 2020) proposal that natural concepts are represented by the cells of an optimally partitioned similarity space. We noted that while this proposal suggests that natural concepts are mind-dependent in *some* sense of that expression, mind-dependence in this sense does not compromise their objectivity. More concerning was the observation that Douven and Gärdenfors' proposal appears to have limited applicability. While it plausibly applies to perceptual concepts, the design criteria that are at the heart of the proposal do not make as much sense for more abstract, scientific concepts.

To address this concern, we introduced an adaptation of Lewis' Best Systems Account of laws of nature. This revised account, while echoing the spirit of Douven and Gärdenfors' emphasis on optimality, posits that natural scientific concepts are those integral to the best systematization of the totality of facts. Specifically, on the new proposal natural concepts are not primitives, but outcomes, of the best systematization of the facts.

It was seen that the new proposal has important implications for the ongoing debate about the status of social metaphysics. Our discussion underscored that the proposal treats concepts from the social sciences as on a par with those from the natural sciences, which suggests that concepts central to social metaphysics can be as inherently natural as those foundational to the natural sciences.⁸

References

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs. Computational Statistics*, 2, 433–459.

⁸ I am greatly indebted to Markus Schrenk, Christopher von Bülow, and an anonymous referee for valuable comments on a previous version of this paper.

- Alon, U. (2003). Biological networks: The tinkerer as an engineer. *Science*, *301*, 1866–1867.
- Ásta. (2018). *Categories we live by: The construction of sex, gender, race, and other social categories*. Oxford: Oxford University Press.
- Bendifallah, L., Abbou, J., Douven, I., & Burnett, H. (2023). *Conceptual spaces for conceptual engineering?* Review of Philosophy and Psychology: Feminism as a case study. in press.
- Borg, I., & Groenen, P. (2010). *Modern multidimensional scaling* (2nd ed.). New York: Springer.
- Bourdieu, P. (1979). *La distinction*. Paris: Minuit.
- Bourdieu, P. (1989). Social space and symbolic power. *Sociological Theory*, *7*, 14–25.
- Cohen, J., & Callender, C. (2009). A better best system account of lawhood. *Philosophical Studies*, *145*, 1–34.
- Cook, R. S., Kay, P., & Regier, T. (2005). The World Color Survey database: History and use. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 223–242). Amsterdam: Elsevier.
- Deauvieu, J Penissat, É, Brousse, C., & Jayet, C. (2014). Les catégorisations ordinaires de l'espace social français. *Revue Française de Sociologie*, *55*, 411–457.
- Decock, L., & Douven, I. (2012). Putnam's internal realism: A radical restatement. *Topoi*, *31*, 111–120.
- Dexter, E., Rollwagen-Bollens, G., & Bollens, S. M. (2018). The trouble with stress: A flexible method for the evaluation of nonmetric multidimensional scaling. *Limnology and Oceanography Methods*, *16*, 434–443.
- Dorst, C. (2019). Towards a best predictive system account of laws of nature. *British Journal for the Philosophy of Science*, *70*, 877–900.
- Douven, I. (1999). Putnam's model-theoretic argument reconstructed. *Journal of Philosophy*, *96*, 479–490.
- Douven, I. (2019). Putting prototypes in place. *Cognition*, *193*, 104007.
- Douven, I. (2021). Implicatures and naturalness. In S. Löbner, T. Gamerschlag, T. Kalenschner, M. Schrenk, & H. Zevat (Eds.), *Concepts, frames, and cascades in semantics, cognition and ontology* (pp. 143–163). Cham: Springer.
- Douven, I. (2022). *The art of abduction*. Cambridge MA: MIT Press.
- Douven, I. (2023). Best explanations, natural concepts, and optimal design. In D. Glass & J. Schubbach (Eds.), *Conjunctive explanations* (pp. 197–218). London: Routledge.
- Douven, I. (2024a) The role of naturalness in concept learning: A computational study. *Minds and Machines*, in press.
- Douven, I. (2024b) The learnability of natural concepts. *Mind & Language*, in press.
- Douven, I., & Gärdenfors, P. (2020). What are natural concepts? A design perspective. *Mind & Language*, *35*, 313–334.
- Galison, P. (1997). *Image and logic*. Chicago: University of Chicago Press.
- Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge MA: MIT Press.
- Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge MA: MIT Press.
- Gärdenfors, P., & Zenker, F. (2011). Using conceptual spaces to model the dynamics of empirical theories. In E. J. Olson & S. Enqvist (eds), *Belief revision meets philosophy of science* (pp. 137–153). Berlin: Springer.
- Gärdenfors, P., & Zenker, F. (2013). Theory change as dimensional change: Conceptual spaces applied to the dynamics of empirical theories. *Synthese*, *190*, 1039–1058.
- Goodman, N. (1955). *Fact, fiction and forecast*. Cambridge MA: Harvard University Press.
- Griffith, A. (2020). Realizing race. *Philosophical Studies*, *177*, 1919–1934.
- Griffiths, P. (1999). Squaring the circle: Natural kinds with historical essences. In R. Wilson (Ed.), *Species* (pp. 209–228). Cambridge MA: MIT Press.
- Hicks, M. (2018). Dynamic Humeanism. *British Journal for the Philosophy of Science*, *69*, 987–1007.
- Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2013). Multidimensional scaling. *WIREs. Cognitive Science*, *4*, 93–103.
- Jaag, S., & Loew, C. (2020). Making best systems best for us. *Synthese*, *197*, 2525–2550.
- Jäger, G. (2007). The evolution of convex categories. *Linguistics and Philosophy*, *30*, 551–564.
- Jäger, G. (2010). Natural color categories are convex sets. In M. Aloni, H. Bastiaanse, T. de Jager, & K. Schulz (Eds.), *Logic, language and meaning* (pp. 11–20). Berlin: Springer.
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data*. Hoboken NJ: Wiley.
- Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*, 1049–1054.

- Khalidi, M. A. (2015). Three kinds of social kinds. *Philosophy and Phenomenological Research*, 90, 96–112.
- Kitcher, P. (1984). Species. *Philosophy of Science*, 51, 308–333.
- Kripke, S. A. (1980). *Naming and necessity*. Cambridge MA: Harvard University Press.
- Leslie, S.-J. (2013). Essence and natural kinds: When science meets preschooler intuition. *Oxford Studies in Epistemology*, 4, 108–166.
- Lewis, D. K. (1973). *Counterfactuals*. Malden MA: Blackwell.
- Lewis, D. K. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61, 343–377.
- Lewis, D. K. (1986). *On the plurality of worlds*. Oxford: Blackwell.
- Loewer, B. (1996). Humean supervenience. *Philosophical Topics*, 24, 101–127.
- Nersessian, N. J. (2008). *Creating scientific concepts*. Cambridge MA: MIT Press.
- Nowak, M. A. (2006). *Evolutionary dynamics: Exploring the equations of life*. Cambridge MA: Harvard University Press.
- Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (2000). *Spatial tessellations* (2nd ed.). New York: Wiley.
- Psillos, S. (1999). *Scientific realism: How science tracks truth*. New York: Routledge.
- Putnam, H. (1975). The meaning of “meaning”. *Minnesota Studies in the Philosophy of Science*, 7, 131–193.
- Putnam, H. (1981). *Reason, truth and history*. Cambridge: Cambridge University Press.
- Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences USA*, 104, 1436–1441.
- Regier, T., Kay, P., & Khetarpal, N. (2009). Color naming and the shape of color space. *Language*, 85, 884–892.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4, 328–350.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale NJ: Erlbaum.
- Schrenk, M. (2008). A theory for special science laws. In H. Bohse & S. Walter (eds.), *Selected papers contributed to the sections of GAP.6, 6th International Congress of the Society for Analytical Philosophy* (pp. 121–131). Paderborn: mentis.
- Schrenk, M. (2017). The emergence of better best system laws. *Journal for General Philosophy of Science*, 48, 469–483.
- Sober, E. (1980). Evolution, population thinking and essentialism. *Philosophy of Science*, 47, 350–383.
- Taylor, E. (2020). Social categories in context. *Journal of the American Philosophical Association*, 6, 171–187.
- Taylor, E. (2023). Substantive social metaphysics. *Philosophers’ Imprint* 23: 18. <https://doi.org/10.3998/phimp.1972>
- van Brakel, J. (1986). The chemistry of substances and the philosophy of mass terms. *Synthese*, 69, 291–324.
- van Brakel, J. (1991). Meaning, prototypes and the future of cognitive science. *Minds and Machine*, 1, 233–257.
- van Brakel, J. (1992). Natural kinds and manifest forms of life. *Dialectica*, 46, 243–261.
- Verheyen, S., & Peterson, M. (2021). Can we use conceptual spaces to model moral principles? *Review of Philosophy and Psychology*, 12, 373–395.
- Xu, Y., & Regier, T. (2014). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. In: P. Bello, M. Guarini, M. McShane, & B. Scassellati (eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 1802–1807). Austin TX: Cognitive Science Society.
- Xu, Y., Regier, T., & Malt, B. C. (2016). Historical semantic chaining and efficient communication: The case of container names. *Cognitive Science*, 40, 2081–2094.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com