



HAL
open science

Review of Jerry Fodor's The Elm and the expert

Pierre Jacob

► **To cite this version:**

Pierre Jacob. Review of Jerry Fodor's The Elm and the expert. European Journal of Philosophy, 1996, 4 (3), pp.373-378. ijn_00000365

HAL Id: ijn_00000365

https://hal.science/ijn_00000365v1

Submitted on 10 Jun 2003

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.

Review of Jerry Fodor's
The Elm and the Expert. Mentalese and its Semantics,
Cambridge, Mass: MIT Press,
1994
1995 (paperback)
128 pages
ISBN 0-262-06170-8 (HB)
0-262-56093-3 (PB)
£ 8.50 paperback

Pierre Jacob
CNRS/France

The goal of much of current philosophy of mind is to show that one can both be a physicalist and subscribe to intentional realism. Suppose token physicalism is true. Then an individual's belief state token is just one of the brain state tokens of the individual. If so, then its having content does not preclude its having physical, chemical and biological properties too. Assuming all of this, however, an intentional realist who is a token physicalist still faces two challenges.

First, physicists, astronomers, cosmologists, chemists and most biologists do not explain phenomena by ascribing content (or intentionality) to the entities with which they deal. So most physicalists fear that, by endowing a device's internal states with content, they, in Dennett's (1971) terms, "take out a loan" of intentionality which they won't be able to pay back. Unless the gap between the intentional and the non-intentional can be filled - unless one can tell which of its non-semantic properties confers onto a belief state token its semantic property -, one will be inescapably drawn to the view which Field (1972) has nicknamed "semanticalism", i.e., the disreputable view that semantic facts are primitive or surd facts.

Second, if intentional realism is right, then content must be a genuine property of some of an individual's mental states: it must make a causal difference. If it did not, then how could we come to know about content at all? The idea is that creatures

with minds - creatures capable of entering states with content - must be able to do things which creatures without are not so capable of doing.

The first challenge has been called the task of naturalizing intentionality. The second challenge is the problem of mental causation. Quite possibly, the two challenges pull in opposite directions: content must be one natural property among many and it must have special causal powers. Arguably too, it may be a constraint on an appropriate theory of content that content be both amenable to the naturalization task and that it be responsive to the demands of psychological explanation. Indeed this will be the main theme of my discussion.

Jerry Fodor is one of the leading philosophers of mind. He accepts both challenges. Ever since the publication of The Language of Thought, twenty odd years ago, he has relentlessly argued for his solution to the problem of the causal explanatory role of content: namely, a computational version of what he calls the representational theory of mind (RTM). In his latest book, The Elm and the Expert (E&E), he still fully subscribes to the computational RTM. In E&E, however, in reaction to the first task - the task of naturalizing intentionality - he now embraces pure informational semantics: all there is to content is information.[\[1\]](#)

According to the computational version of RTM, the job of psychology is to supply causal explanations of individuals' intentional behavior by subsumption under intentional laws. On this view, the causal explanation of an individual's intentional behavior is nomic: psychological laws are both intentional laws which refer to the contents of the individual's mental states and they are ceteris paribus causal laws. Whether intentional or not, what makes ceteris paribus laws of special sciences (which by assumption are not fundamental laws of basic physics) causal laws, is that they hold in virtue of some underlying causal mechanism. On the computational version of RTM, the underlying causal mechanisms responsible for the implementation of intentional psychological laws are computational processes. In fact, in Fodor's own version of computational RTM, the content of an individual's

mental state reduces to the semantic property of a formula of the individual's language of thought. Mental formulae have both semantic and syntactic properties. Mental processes are formal processes: they detect only the syntactic properties of mental symbols, not their semantic properties. So although psychological laws refer to the contents of mental symbols, mental processes which implement psychological laws are purely computational.

Given Fodor's commitment to intentional realism, the computational version of RTM raises the following question: Does it really secure a causal explanatory role for content? In a sense, it does since intentional psychological laws refer to content. This is the sense in which the computational version of RTM differs from Stich's (1983) syntactic theory of mind. But in another sense, it does not since content is not involved in the computational processes via which psychological laws are implemented. Because, as I said above, Fodor now embraces a purely informational approach to the task of naturalizing intentionality, the question of whether computational RTM can really distinguish itself from the purely syntactic theory of mind (as Fodor himself recognizes, E&E, p. 50) becomes more pressing than ever. Indeed, the main novelty is that in E&E, Fodor repudiates any distinction between broad and narrow content: more exactly, he gives up the notion of narrow content.

Following Loewer & Rey (1991), we might call the new picture of mental content "the pure locking theory". According to the pure locking theory, for a mental symbol (or internal state of some physical device) to mean dog, as it might be, is for it to covary nomically with instantiations of doghood. Only informational relations, i.e., external nomic relations between tokenings of a symbol and instantiated properties in the world can confer content (or meaning) onto a symbol. Relations among symbols can't generate content. Pure locking theory is the view that semantics is in effect metaphysics: the job of semantics is to tell us how symbols hook up onto properties in the world. The slogan of pure locking theory goes like this: semantics isn't part of psychology. Nor is it part of epistemology. What Fodor has in mind is that

it is a mistake to derive semantic conclusions from either psychological or epistemological premisses.

The main question of E&E then is whether pure locking theory and the computational RTM can fit each other. In particular: Can purely computational mental processes implement intentional psychological laws which refer to the contents of individuals' mental states when content is purely informational? In effect, this is the question whether the notion of content captured by pure locking theory can adequately respond to the demands on content made by psychological explanation. Before I turn to this fundamental question, I'll clarify the claims that semantics is neither part of psychology nor part of epistemology.

Consider first the claim that one should not derive semantic conclusions from epistemological premisses. Pure locking theory (i.e., pure informational semantics) is radically externalist: meaning is constituted by external nomic relations between a symbol and properties instantiated in the environment. But it is radically anti-individualist. Typical anti-individualist accounts of so-called "deferential" concepts claim that an individual's concept of an elm is constituted by what botanists in his or her community think and say about elms. They argue for this view from the fact that an individual who is not a botanist can't distinguish elms from beeches. In order to block the anti-individualist conclusion, Fodor rejects the premiss according to which an individual who is not a botanist and who lives in a community including botanists cannot tell elms from beeches. The way an individual who lives in a community including botanists distinguishes elms from beeches is by asking a botanist. The informational move in epistemology is to see that experts in botany can be used by non-experts as reliable instruments for the detection of elmhood. According to Fodor, what the anti-individualist does then is commit a verificationist fallacy: botanists' thoughts about elms do not constitute the meaning of e.g., the English word "elm" any more than telescopes constitute the meaning of the English word "star". Telescopes contribute to the covariation between tokenings of individuals' thoughts

about stars and instantiations of starhood. Botanists' thoughts about elms contribute to the covariation between non-botanists and instantiations of elmhood.

This pure informational account of deferential concepts leads Fodor to some insightful remarks about the psychological equipment needed for the use of experts as reliable instruments (pp. 35-37; 91-101): only if a creature is capable of having "policies" with respect to both her thoughts and the thoughts of others will she be able to achieve the reliable correlation between tokenings of her thoughts and instantiations of property \underline{P} in the environment by means of coopting the reliable correlation between an expert's thoughts about \underline{P} and instantiations of \underline{P} . Only a creature capable of forming higher order thoughts about the mechanisms delivering reliable thoughts could coordinate her thoughts with the thoughts of reliable experts.

I now turn to the claim that semantics isn't part of psychology. The job of psychology is to account for the fact that intentional psychological laws are causal laws by supplying a theory of the underlying non-semantic computational mental processes. Given the language of thought hypothesis, mental processes relate mental symbols to one another. But informational semantics deals only with the relations between symbols and properties instantiated in the environment. And pure locking theory is pure informational semantics. Now, there are two ways in which informational semantics may reveal its purity: one way is by eschewing any collusion with evolutionarily based teleosemantic accounts of content; the other way is by not allowing any room for inferential (or conceptual) role semantics. Except for a cryptic remark on p. 20, Fodor doesn't devote much time in E&E to justify his rejection of a teleosemantic approach to content. Nor will I consider his reasons for doing so (which are presented in Fodor (1990)). But I do want to consider first why he so strenuously rejects the view he calls (pp. 86-87) "structuralism" or the "intrasymbolic view of content", namely inferential role semantics (IRS). In other words, I want to consider briefly his reasons for espousing semantic atomism. As he puts it (e.g., p. 37) about the concept of ELM,

there is nothing that you have to believe, there are no inferences that you have to accept, to have the concept ELM. According to externalism, having the concept ELM is having (or being disposed to have) thoughts that are causally (or nomologically) connected, in a certain way, to instantiated elmhood. Punkt.

It is of the essence of semantic atomism that you could have the concept ELM without having the concepts TREE, BRANCH, TWIG, ROOT, PLANT, PHYSICAL OBJECT and so on. But can you really?

It emerges early on (pp. 5-6) that two lines of thought conspire to make him want very much to resist IRS. The first line of thought is that Fodor assumes that informational semantics is the best approach to naturalizing intentionality. Informational semantics seems to capture meaning piece by piece: if a symbol nomically covaries with instantiations of cowhood, then it means cow. But in fact informational semantics is not bound to be atomistic. What if your mental symbol for ELM could not covary directly with instantiations of elmhood? What if the covariation between your concept ELM and instantiations of elmhood were mediated by your concept TREE? The second line of thought has itself two ingredients: the first ingredient is that there is no alternative to semantic atomism but semantic holism. The second ingredient is that semantic holism is inconsistent with there being intentional psychological laws which are required for the purpose of the psychological explanation of individual intentional behavior. The reason therefore Fodor wants at all costs to avoid IRS is that according to IRS, the content of a symbol depends upon the inferential relations between this symbol and other symbols. If the content of an individual's belief depends upon its inferential relations to the individual's other propositional attitudes, and if no two individuals have all the same propositional attitudes, then, as Fodor (1987, 1990) and Fodor & Lepore (1992) have emphasized, the risk is that no pair of individuals will ever instantiate any one intentional psychological law.

Whether or not IRS implies semantic holism, I think Fodor is wrong to think that the alternative between semantic atomism and semantic holism is exhaustive. But

since Fodor does not argue for this claim in E&E, I won't argue against it either. I also think Fodor is wrong to think that semantic holism threatens the possibility of intentional psychology because I think he is wrong to assume that intentional psychological laws which are requested for the causal explanation of an individual's intentional behavior must refer to the contents of the individual's mental states. What I think these laws require is quantification over particular contents such as "If X believes that not q unless p and X desires that q , then ceteris paribus X will try to bring it about that p . Reference to particular contents then yields instantiations of such general laws.

I now turn to the main question faced by E&E: is the purely informational view of content really consistent with the computational version of RTM? In chapter I of E&E, Fodor argues that the main obstacle to combining a computational view of mental processes with a pure informational view of content is the assumption that there must exist computationally sufficient conditions for the instantiation of semantic properties, i.e., of informational properties. This he concedes would be impossible. But something weaker will do, as the following analogy suggests: there is a mechanism which insures that the correlation between instantiations of the property of being a dollar bill and instantiations of the property of being a dollar-looking bill is reliable. This mechanism is, let us say, the action of the US police. On the basis of this analogy, Fodor proposes to replace the condition that there exist computationally sufficient conditions for the instantiation of informational properties by the weaker condition that the co-instantiation between implementing computational properties and implemented informational properties be reliable. He then faces two kinds of familiar challenges in which computational properties and informational properties are not in phase with one another: one standard example is Putnam's molecular twins where one and the same computational state can be ascribed distinct informational contents; the other is Fregean cases in which different computational states have one and the same informational content.[\[2\]](#)

Two tasks are therefore incumbent upon Fodor: one is to show that both Putnam cases and Frege cases are not really counter-examples to the implementation of informational properties by computational properties, but rather exceptions. Exceptions being, unlike counter-examples, compatible with a reliable contingent generalization. The other is to reveal the mechanism which is keeping implementing computational properties and implemented informational properties reliably in phase with each other.

As I already mentioned, Fodor rejects IRS (because he believes IRS entails semantic holism); and he rejects any distinction between broad and narrow content. So he cannot appeal to sameness of narrow content to handle Putnam's twin cases anymore: he cannot claim that when I and my twin on Twin Earth both think something we express by uttering "There is a lot of water here", we think a thought with different broad contents and the same narrow content. And what Fodor does argue is that Putnam's twins are indeed exceptions to chemical laws. What Putnam's twin cases show is "something about our concept of content", i.e. that "the supervenience of the broadly intentional upon the computational isn't conceptually necessary... But it doesn't argue against the nomological supervenience of broad content on computation since... XYZ is nomologically impossible" (E&E, p. 28).

Consider now a typical Fregean case, e.g., the case of Oedipus who believed that Jocasta was attractive and who did not believe that his mother was attractive even though in fact Jocasta was his mother. Given his acceptance of the computational version of RTM, Fodor has the resources for distinguishing among some informationally equivalent concepts, e.g. my concept of WATER and my concept of H₂O, along differences among their syntactic and compositional properties: I can have the former, but I can't have the latter, without having the concepts of HYDROGEN, the concept of NUMBER 2 and the concept of OXYGEN. Even though Oedipus' belief that Jocasta was attractive has the same content (the same truth-condition) as his belief that his mother was attractive, Fodor might want to argue that the two beliefs are different belief states and that this difference is all that

is required for the purpose of psychological explanation: Oedipus' behavior based on his former belief differed from his behavior based on the latter belief. Of course, if he were to make this move, he would presumably have to recognize that the explanation of the difference between what Oedipus does when he believes one thing and what he does when he believes something else does not appeal to any semantic difference between the contents of Oedipus' beliefs. And he would presumably have then to face the objection that he cannot meet what Cummins (1991) and Kim (1991) call Stich's Syntactical challenge: on Stich's (1983) syntactic theory of mind, content is simply irrelevant to psychological explanation. If the contents of Oedipus' beliefs and desires are purely informational, then presumably he does not act out of the contents of his beliefs and desires.

But in any case, this is not the strategy Fodor follows. Instead, he proposes to argue that Frege cases are exceptions to reliable ceteris paribus psychological generalizations. Whereas, it seems to me, the claim that Putnam's twins are exceptions to chemical laws is not implausible, the claim that Frege cases are exceptions to psychological laws strikes me as wildly implausible. In fact, Fodor commits himself to the incredibly strong claim that "any intentional psychology... has to take for granted that identicals are generally de facto intersubstitutable in belief/desire contexts for those beliefs and desires that one acts on" (E&E, p. 40). We can see how strong this claim is when we realize that it amounts to a denial of the opacity of those beliefs and desires on which an agent acts. Suppose I do something which happens to be the best thing I could have done in the circumstances. Fodor's principle would require me to know that "what I just did" and "the best thing to do" refer to one and the same act! Fodor then tries to justify this amazingly strong claim by connecting knowledge of identities to rationality (by means of a Principle of Informational Equilibrium and two truisms, pp. 41-42). The obvious response to Fodor is that lack of knowledge of identities is not the same thing as irrationality. Here, we reach, I think, a tension in Fodor's recent work. On the one hand, Fodor (1990) recognizes that informational semantics must make room for the possibility of

misrepresentation. On the other hand, he now claims that cases of unknown coreference between two coreferential terms are mere "aberrations". If belief/desire psychology was, as Fodor writes (p. 42), "committed to treating Frege cases as aberrations", I wonder why he for one should think that a naturalistic semantics ought to account for the possibility of false beliefs and beliefs about non-existent objects.

[\[3\]](#)

References

- Cummins (1991) "The Role of Mental Meaning in Psychological Explanation", in B. McLaughlin (ed.) Dretske and his Critics, Oxford: Blackwell.
- Dennett, D. (1971) "Intentional Systems", in Brainstorms, Montgometry Vt: Bradford Books.
- Field, H. (1972) "Tarski's Theory of Truth", The Journal of Philosophy, LXIX, 13, 347-75.
- Fodor, J.A. (1975) The Language of Thought, New York: Crowell.
- Fodor, J.A. (1987) Psychosemantics, Cambridge, Mass.: MIT Press.
- Fodor, J.A. (1990) A Theory of Content, Cambridge, Mass.: MIT Press.
- Fodor, J.A.& E. Lepore (1992) Holism, a Shopper's Guide, Oxford: Blackwell.
- Kim, J. (1991) "Dretske on How Reasons Explain Behavior", in B. McLaughlin (ed.) Dretske and his Critics, Oxford: Blackwell.
- Loewer, B. & G. Rey (eds.)(1991) "Editors's Introduction" to Meaning in Mind, Fodor anf his Critics, Oxford: Blackwell.
- Stich, S. (1983) From Folk Psychology to Cognitive Science, Cambridge, Mass.: MIT Press.

[1] This is not strictly true for two reasons: First, Fodor (1990) already embraced a purely informational account of content, but then he didn't reject narrow content as explicitly as he does in E&E. Second, Fodor accepts something like a "use" theory (or an inferential role theory) for the logical vocabulary. But I will simply disregard the meaning of logical words here.

[2] This is not quite true: he in fact faces three potential counterexamples to his claim: Putnam's twin cases, Frege cases and Quinean cases in which informationally equivalent expressions refer to non co-extensive properties which are always co-instantiated, as "rabbit" and "undetached rabbit parts". But since his treatment of Quinean cases involves the logical vocabulary which I disregard here, I won't discuss his solution to Quinean cases either.

[3] Thanks to Ned Block and Mark Sacks for their comments.