

Laurence, S., & Margolis, E. (1999). “Jerry A. Fodor: Concepts: Where Cognitive Science Went Wrong.” *The British Journal for the Philosophy of Science*, 50(3), 487-491.

The final version is available at The University of Chicago Press:

<https://doi.org/10.1093/bjps/50.3.487>

Review of Jerry A. Fodor’s *Concepts Where Cognitive Science Went Wrong*

Stephen Laurence and Eric Margolis¹

For a number of years, Jerry Fodor has been exploring the terrain of ‘atomistic’ theories of concepts. In *Concepts: Where Cognitive Science Went Wrong*, he continues in this vein, providing a rich account of what a science of the mind might look like if concepts had no structure, while simultaneously presenting a daunting critique of much of contemporary philosophy and cognitive science. Like most of Fodor’s work, this book is bound to infuriate some and delight others. Either way it is full of interesting arguments, and as Fodor’s most comprehensive overview of his philosophical position in years, his new book deserves to be—and surely will be—widely read.

Fodor’s discussion takes place within the context of a commitment to the Representational Theory of Mind (RTM). According to RTM, an agent who has a propositional attitude, such as a belief or a desire, stands in a specific computational relation to a mental representation from which the belief or desire inherits its content. For example, Amy’s believing that cats have whiskers is a matter of her standing in a particular computational relation to a mental representation that has the content *cats have whiskers*. Such representations are also supposed to have combinatorial structure (a syntax and a compositional semantics), and concepts are taken to be their constituents. Thus concepts are themselves mental representations.²

The thesis that concepts are mental representations is a minority view in philosophy, where concepts are typically viewed as abstract objects along the lines of Fregean senses or as psychological or behavioural abilities. And in cognitive science, where RTM is more of a common currency, the compositional nature of the representational system remains a hotly debated topic. For present purposes, however, let us grant the larger framework of Fodor’s book.³

As the title of the book indicates, Fodor is not happy with what he takes to be the received view of concepts in cognitive science. ‘Unlike practically everybody else [. . .] I think that the theory of

¹ This review was fully collaborative; the order of the authors’ names is arbitrary.

² Accordingly, concepts are formal objects, though they are at least partly individuated in terms of their semantic properties.

³ Fodor offers new arguments on all these topics—see especially Chapter 2 and the section titled ‘The Arguments for Compositionality’ in Chapter 5.

concepts ought to be atomistic' (p. vii). Everyone else, according to Fodor, is invested in some form or other of Informational Role Semantics (IRS), which takes concepts to be at least partly individuated by inferential relations that are mediated by the structure of concepts and serve to fix their contents. For instance, in linguistics and philosophy it is not uncommon to hold that lexical (that is, word-size) concepts have definitional structure. For philosophers, such structure is recruited to explain our intuitions of analyticity; for linguists, it is supposed to explain a variety of distributional patterns in language. Psychologists also hold that lexical concepts have structure, but they generally hold that lexical concepts have prototype structure (roughly, that their constituents express properties that items in their extension tend to have) or that they have 'theory' structure (roughly, that they are essentially embedded in a theory of the items in their extension). Fodor will have none of this. 'What is the structure of the mental representation DOG? [. . .] my answer will be that, on the evidence available, it's reasonable to suppose that such mental representations have no structure; it's reasonable to suppose that they are atoms' (p. 22).

Fodor's argument against these various alternative positions proceeds on a case-by-case basis, with each theory being rejected for its own reasons. Nevertheless, there is a central theme running through Fodor's treatment of these competing views. The problem with IRS theories is supposed to be that they are incompatible with the principle of compositionality (p. 107):

[H]ere's the impasse: prototypes are public (i.e. they are widely shared) and they are psychologically real, so they do meet two of the non-negotiable conditions that concepts are required to meet; but they aren't compositional. Definitions would be compositional if there were any, but there aren't, so they're not. As things stand, there is no version of the inferential role theory of conceptual content for which compositionality and psychological reality can both be claimed. I think there must be something wrong with inferential role theories of content.

In other words, for a theory of concepts to be a contender, it must be consistent with the constraints that compositionality imposes. The only IRS theory that respects compositionality is the definitional theory, but the definitional theory is a nonstarter since there aren't any definitions.

Certainly, if it is sound, this is an interesting argument with far-reaching implications for philosophy and cognitive science. But the situation is much more complicated than Fodor lets on, as we shall illustrate using the example of the Prototype Theory.

Fodor's case against prototypes is that they do not compose. He cannot see any way in which the prototype for PET FISH is a function of the prototypes for PET and FISH. Good examples of pet fish (e.g. goldfish) are poor examples of both pets and fish. How, then, Fodor asks, could the prototype for the complex concept be computed on the basis of the prototypes of its constituents?

There are several assumptions here that ought to be made explicit. One is that the theory that Fodor takes to be his target is one that *identifies* concepts with prototypes. It is not just that concepts have prototypes and that prototypes are part of the nature of concepts; prototypes are supposed to give an *exhaustive* account of the structure of concepts. Another assumption is that the type of semantic properties that a constituent has must also be exhibited by its hosts. So if the constituents of a complex concept have prototype structure, the complex concept itself must have prototype structure.

The latter assumption is somewhat ironic for Fodor to hold, since even his own theory of concepts does not abide by it. Fodor's account of the reference-determination of primitive concepts is broadly in terms of the causal-covariance of concepts and the properties they express. To a first approximation, DOG expresses doghood (and is thereby true of all and only dogs) because there is a causal law connecting dogs with DOGs. But notice, Fodor is not about to say that DOG OWNER expresses *dog owner* because of a causal-covariance between these two as well. No, for complex concepts Fodor relies upon a compositional principle that operates on the syntax of a complex concept and the referential properties that his covariational theory delivers for each of its constituents. The question, then, is what's to stop a prototype theorist from saying the same thing? A prototype theorist need not hold that PET FISH has to have a prototype that is computed on the basis of the prototypes of PET and FISH. She can adopt a theory much like Fodor's, only one that has a different account of how the constituents of a complex concept acquire their referential properties.

Fodor's other assumption is also problematic. Prototype theorists need not maintain that concepts *are* prototypes. A more natural view to hold is one in which a prototype comprises only one aspect of the structure of a concept. As such it may have nothing to do with the reference of the concept, which may well be determined by other aspects of its structure. In fact, many prototype theorists hold a view of this sort—what is often called a *dual theory*—where concepts are taken to have a ‘core’, which determines the reference of the concept, and an ‘identification procedure’, which is accessed in the sort of processing that is usually cited as evidence in favour of prototypes. For the purposes of semantic evaluation, it is only the conceptual cores that need to compose.

But why should one think that a prototype is part of a concept if it has nothing to do with the concept's semantics? The answer is clearly that prototypes figure prominently in numerous important psychological generalizations concerning categorization, learning, memory, and inductive inference, among others. Their effects are so widespread and so pronounced that no adequate theory of concepts could fail to take into account their prototype structure.

So Fodor's arguments against prototypes don't work. Prototype theorists can appeal to the same principles of semantic composition that non-prototype theorists use or they can adopt a dual theory of concepts while maintaining that prototypes need not be involved in reference determination at all.

Of course, a number important questions remain open, for example, what are conceptual cores if not prototypes? Psychologists have tended to hold that cores are definitions and, more recently, that some might be theories. But these accounts are not without their own problems. For this reason, we suspect that the best theory of conceptual cores is one that says they are atomic. The idea, in other words, is to integrate Fodor's own theory of concepts—or some other atomistic theory—with more standard theories in cognitive science by taking the atomistic theory to be an account of conceptual cores; see Laurence and Margolis ([1999]).

To make his conceptual atomism more attractive, Fodor devotes the final two chapters of the book to attempting to remove what many have taken to be the biggest obstacle to conceptual atomism—its perceived link to radical concept nativism, the view that virtually all lexical concepts are innate. These chapters constitute the book's most significant departure from Fodor's earlier work. Fodor

is famous—or rather, infamous—for having advocated the thesis that lexical concepts ranging from PINEAPPLE to PORCUPINE and MODEM to MOLECULE are all innate. Though the issues here are complicated, Fodor argued in earlier work that unstructured concepts have to be innate because they cannot be constructed by assembling their constituents, and if they cannot be constructed then they cannot be learned (Fodor [1981]). While we do not think that the connection between conceptual atomism and nativism is anywhere near as tight as has been supposed, many have taken Fodor’s earlier arguments to be a *reductio ad absurdum* of conceptual atomism.⁴ By recanting his earlier views, Fodor aims to address this worry.

Fodor’s discussion is centred around what he calls the doorknob/DOORKNOB Problem. This is the problem of accounting for why a concept is occasioned by its instances. His solution is to claim that unstructured concepts express mind-dependent properties and that, because these properties are actually constituted by how their paradigmatic members strike our kind of mind, it is no surprise that when we confront these instances we acquire the corresponding concepts.

While Fodor’s treatment of the doorknob/DOORKNOB problem is full of interesting insights and observations, ultimately it does not provide a satisfactory response to his earlier arguments for radical concept nativism. The problem is that it is completely silent on the process of concept acquisition. We are told that it makes sense that a concept’s instances will be among its original causes, but we are not told anything about the mechanism by virtue of which the concept is acquired. This being the case, Fodor’s account is simply neutral on the issue of nativism. To see the extent of the problem, consider that nothing in Fodor’s account prohibits there being a distinct innate mechanism for acquiring each lexical concept—one mechanism per concept. True, such a model would require only a commitment to innate mechanisms of acquisition, not innate contentful states (see p. 143), but the result would not be noticeably different from the nativist triggering model that Fodor originally advocated. That model, after all, does not say that concepts are literally present in the mind of an agent before they are triggered. Rather, it says that a concept is innate because it is the causal effect of an innately specified mechanism that is specifically responsible for its acquisition—a mechanism that is all wired up to occasion that very concept. The point is that such a story is entirely consistent with the new metaphysical machinery that Fodor introduces to get around radical concept nativism. So his new metaphysics does not provide a non-nativist alternative to triggering. For a real atomistic alternative to radical concept nativism, something has to be said about the process of concept acquisition itself.⁵

References

Churchland, P. [1986]: *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, Cambridge, MA: MIT Press/Bradford Books.

⁴ Or worse, a *reductio* of RTM. See e.g. Churchland ([1986]).

⁵ For an account that *does* explain concept acquisition within an atomistic framework and for a detailed response to Fodor’s earlier arguments for radical concept nativism, see Margolis ([1998]) and Laurence and Margolis ([MS]).

Fodor, J. [1981]: ‘The Present Status of the Innateness Controversy’, in his *Representations: Philosophical Essays on the Foundations of Cognitive Science*, Cambridge, MA: MIT Press, pp. 257–316.

Laurence, S. and Margolis, E. [1999]: ‘Concepts and Cognitive Science’, in E. Margolis and S. Laurence (eds) *Concepts: Core Readings*, Cambridge, MA: MIT Press/ Bradford Books, pp. 3–81.

Laurence, S. and Margolis, E. [MS]: ‘Concepts, Content, and the Innateness Controversy’ [revised title: ‘Radical Concept Nativism’].

Margolis, E. [1998]: ‘How to Acquire a Concept’, *Mind & Language*, 13, 3, pp. 347–69.