

-EDITORIAL-

**Concepts: Fodor's little semantic BBs of thought
- A critical look at Fodor's theory of concepts -**

Eric Dietrich
Philosophy Department
Virginia Tech
Blacksburg, VA 24060

1. Concepts, AI, and the state of the art.

I find it interesting that AI researchers don't use concepts very often in their theorizing. No doubt they feel no pressure to. This is because most AI researchers do use *representations* which allow a system to chunk up its environment, and basically all we know about concepts is that they are representations which allow a system to chunk up its environment.

However, concepts are crucially important to other branches of cognitive science -- especially, to theorizing in cognitive psychology and philosophy. Which is a major embarrassment . . . for researchers don't know what concepts are. I don't mean they (I could say "we") are uncertain about what concepts are in detail; I don't mean that our theory of concepts is a little rough around the edges. I mean that we are just shy of completely clueless about what concepts are, and that we have no remotely agreed-upon theory of concepts at all. The situation we cognitive scientists are in here in the twenty-first century is *not* like the one the atomic physicists were in at the turn of the last century when they had the incorrect model of the atom -- what was called the plum-pudding model. The plum pudding model was so robust that it was open to experimental refutation, which was exactly what Ernest Rutherford did to it -- he experimentally refuted it, replacing it with the "solar system" model of the atom (which soon morphed into the Bohr model). No, our situation is not like this. There is *no* single theory of concepts that is refuted by the current data we know have. This is a rather amazing situation. (The one possible exception to my claim is the suite of theories known as the definitional theories or classical theories. These theories say that concepts are definitions, and definitions are necessary and sufficient conditions. Hence, concepts have necessary and sufficient conditions for their

application. But, no clever experiments refuted this theory, indeed, the classical theory has been known to be flawed since the ancient Greeks, rather clever arguments refuted it -- arguments supplied by armchair psychologists, aka. philosophers. Moreover, there are several neoclassical theories around now which seek to avoid problems that the classical theories had. See, Margolis and Laurence, 1999, for an introduction to these issues).

Our situation is more like that of the early Greek Atomists who believed that simple, minute, indivisible particles are fundamental constituents of the universe. This was right (more or less, except for the indivisible part), but just look at how thin this "theory" is. We know that concepts are representations of some sort, that they are the prime constituents of most thoughts and thinking, that they are what allow us to think about, e.g., dogs and cats (actually, all of this is contentious), but after this, agreement among cognitive researchers is hard to come by (or even harder to come by).

There is one major difference, however, between early atomism and our theories of concepts. We really do have theories of concepts. That's the problem: we have several theories of concepts, and we have data and arguments which support them all in various ways and we have data and arguments which call each theory into question.

One of the new leading views of concepts is due to Jerry Fodor (Fodor, 1998a and 1998b). In this editorial, I would like to explain it and then comment on it (well, ok, I'm going to criticize it).

2. Fodor's view of concepts.

Your contemporary cognitive psychologist, your thoroughly modern mentalist, is inclined to believe that humans have concepts and that they use their concepts for thinking. No doubt they are correct in this belief. Thinking, moreover, at least sometimes, involves recognizing, believing, doubting, suspecting, etc., and then reasoning from these to further recognitions, beliefs, doubts, and suspicions. Put succinctly, thinking involves epistemic processes. Indeed, thinking is *the* epistemic process. This much, too, is agreed to by all. The question arises, however, as to where

the epistemology resides. Is it in the concepts themselves, or is it somewhere else, using the concepts when needed as constituent parts involved in the epistemic processing? Here opinions diverge. Recent polls suggest that all cognitive scientists but one go with the former view -- epistemology resides in the concepts themselves. But Fodor, lonesome Fodor, the Don Quixote of cognitive science, goes with the latter -- concepts lack any epistemic component essentially, but nevertheless participate in epistemic processes in virtue of their semantic nature.

Here is Fodor's view. Concepts are first and foremost semantic things. They have a role to play in our thinking – our epistemology – in virtue of their semantics. They contribute their *meaning* to the epistemic processes responsible for our epistemic contact with the world and the resulting human epistemology.

What is semantics? For Fodor, it is a relation between certain thoughts (beliefs, etc.) and the world, in virtue of which the thoughts are either true or false. Semantics, meaning, is about truth and falsity, period. Not, note, about *finding out* about which of your thoughts are true or false -- just about truth and falsity. This makes concepts fundamentally metaphysical: to understand the nature of concepts, we need only do metaphysics – ontology to be exact. What there is, what exists, is what causes your beliefs to be true or false.

Of course, you can find out (sometimes) which of your beliefs are true or false. That is because you are an epistemic creature: you can test, run experiments, ask questions, etc. But, for Fodor, your being an epistemic creature is not tied up in any essential way to your being a creature with concepts. You merely use your concepts in epistemic processes.

What are these epistemic processes? They are what Fodor calls "nondemonstrative inference," but you and I know them as induction. Induction contrasts with deduction, which is demonstrative inference.

So here's the picture. We humans have the ability to make inductions (statistical inferences, for example). This ability is implemented as processes that use concepts. Your concept of raven is used in inferring that all ravens are black from the fact that all the ravens you

have ever seen are black. But your concept for raven does not contribute any epistemic force to that inference. The epistemic force of that inference resides solely in the inference process itself.

While this might seem palatable so far, removing epistemological properties from concepts has some rather shocking consequences. For example, your ability to *recognize* ravens has nothing to do with your concept of ravens. Concepts, according to Fodor, are not for recognizing things in the world. You don't use concepts to know about the world. You use induction (nondemonstrative inference) to know about the world, and you use concepts in those inferences. What makes your inductive inference that all ravens are black about ravens is that your raven concept participates in it. Your raven concept contributes, via this participation, the semantics necessary to make your inference true or false. And this truth or falsity relation, metaphysical as it is, is what makes your concept about ravens. But, to repeat, your raven concept does *not* have anything to do with your ability to pick ravens out of the environment.

Fodor's argument for all of this can be stated quickly. Concepts compose. This we know for certain (Fodor says). "Brown" composes with "cow" to produce "brown cow." But if concepts are essentially epistemological, they can't compose. Why? Because being epistemological means having as an essential property the ability to pick out *good* instances. And picking out good instances *doesn't* compose. Proof: You can pick out good instances of "fish" (say, a trout); and you can pick out good instances of "pet" (say, a dog); it doesn't follow, however, that you can pick out good instances of "pet fish" (a goldfish). So the crucial epistemic property doesn't compose; yet, concepts *do* compose; ergo, concepts don't have the crucial (necessary) epistemic property. Hence, concepts are not part of our human epistemology. What properties, then, do concepts have? The only ones left are semantical properties. So it is these that they have.

Finally, according to Fodor, concepts have no constituent structure. Concepts combine to form complex concepts. But the basic atomic concepts have no structure. So for example (as we've seen), the concept "brown" and the concept "cow" can come together to form "brown cow," but "cow", for example, itself has no structure. If all of the forgoing wasn't unintuitive enough for you, here's how unintuitive this new claim is. Consider analogy making in humans. Consider one of science's most famous analogies: the atom is like the solar system. There is a fair

amount of evidence that this analogy was the result of mapping the high-level, relational structure of "atom" to the high-level, relational structure of "solar system" (see Gentner, 1983, 1989). And, at least *prima facie*, analogy looks like something that obtains between concepts. In particular, "atom" and "solar system" look like basic concepts. At least "atom" seems to be like "cow". If "cow" is basic (and Fodor assures us that it is), then I see no reason why "atom" couldn't be basic, too. But high-level, relational structure is a *fortiori* *structure*. So analogy happens when it does because basic concepts have structure. So Fodor is wrong. Unless the evidence for analogy is wrong.

Fodor's reply to this objection is that the basic, structureless concepts participate in theories, and that analogy is really a relational mapping between theories. Note how now we've got a new theoretical entity to deal with: mental theories. What are these things? Fodor hasn't said. Presumably these things *are* epistemic. So Fodor still has epistemic, structural things – viz., what he calls mental theories – which are still manipulable and mappable. All he's added is his BB model of concepts. So the question naturally arises as to how much of an advance Fodor's model of structureless, epistemology-free concepts really is. This is what we now turn to.

3. Concepts.

There's an important sense in which we could stop now, and re-label. What the rest of cognitive science means by "concept", Fodor means by "mental theory." (Of course, some psychologists have hypothesized that concepts *are* theories, see, e.g., Murphy and Medin, 1985.) Both the rest of cognitive science and Fodor think analogy is relational, structural mapping. We just disagree over the term we are going to use to describe what the mappings are between. Both Fodor and the rest of cognitive science think that concepts have meaning, and contribute this meaning to the representations they participate in. Each just puts the epistemology in a different place, places that, in the final analysis, aren't so different if you just relabel: cognitive scientists put the epistemology in concepts; Fodor puts the epistemology in mental theories. If we (everyone but Fodor) agree to use the term "mental theories" for what we call "concepts" when in the presence of Fodor, we will all get along and be able to have a nice dinner together. So basically, we are all in agreement here. All is well. Yawn.

But there is fly in the ointment. Three flies, rather. First, it greatly strains intuition to divorce our epistemology from our concepts. It seems overwhelmingly likely that our concepts *are* involved in our abilities to recognize objects. Overthrowing this intuition merely in favor of another intuition about the primacy of compositionality seems drastic. Secondly, the semantical relationship Fodor wants to leave in place is so metaphysical that seems incapable of doing any work for practicing psychologists or cognitive scientists. Thirdly, psychologists have an idea about what concepts are which both preserves their status as epistemic mental objects and enables compositionality. Perhaps we owe it to them to take their idea seriously. So as nice as things might be if we merely re-labeled, perhaps we ought to roll up our sleeves and come up with a new plan. Let's consider these three flies in reverse order.

3.1 Psychology.

Psychologists are well aware of the noncomposability of prototypes. But rather than retreating to antiseptic cleanliness for concepts, they have gone precisely the opposite direction: they have opted to make concepts very more complicated with lots of structure. We have here an instance of a general methodological principle which Fodor has missed. The principle is that the mind is *very* complicated. So it is not a good idea to postulate simple, well-behaved structures and processes as the implementations of thinking in the mind. Though Fodor was in at the kill of behaviorism, the complexity of what he and his cohorts wrought during those halcyon years of yore seems to have escaped him (and his cohorts).

All psychologists know that a theory of concepts must explain both their flexibility and their stability, so in psychology, concepts are hypothesized to have lots of different kinds of information in them at multiple levels of abstraction. Concepts are therefore much more than lists, and a fortiori, much more than semantic BBs. They contain, at a minimum, 1) prototypical information on the category in question, 2) detailed representations of the various properties that make up the category (e.g., shape, or having feathers, etc.), 3) semantic or lexical-entry information, and 4) theoretical information that informs the possessor of the concept of all kinds of complex relational and causal information (for example, psychologists hypothesize that our concept of "bird" contains such information as "born of bird parents," and "can produce bird offspring"; and as another example, Barsalou has suggested that concepts need to contain

context-independent information constituting conceptual cores, context-dependent information, and recent-context-dependent information (Barsalou, 1989)).

3.2 Semantics.

It seems like a bad idea to divorce epistemology from semantics. Semantics is much more than some gossamer metaphysical relation between thinker and thing thought about. Making thoughts merely true or false doesn't seem to be enough work for semantics, at least for cognitive science and understanding the mind. Propositions are true or false. Making the mind a lot like mere propositions (or a collection of them), while perhaps good for Fodor's language of thought hypothesis, is pretty obviously a dead-end. Things need to be more complicated than that.

Semantics is made up of two things. First, it is informational relations between one's concepts, percepts, and the like with the world, and second it is functional connections between all of one's concepts, and between one's concepts and lower-level perceptions and sensori-motor representations.

The informational relations produce things with truth-conditions (e.g., my belief that it is snowing now). But semantics is not just the informational relation that makes beliefs true or false. Rather, concepts also involve the ability to test and figure out, learn, what, e.g., snow is, what snowing is, and whether it is snowing now. So, the way we test truth or falsehood is by comparing concepts with each other and with other low-level sensori-motor perceptions.

The ability to figure out what your concepts pick out does *not* entail, require, or use any checking with the world, for that is impossible. Thinking systems like us cannot check their concepts with the world, rather, we check our concepts with other concepts. All knowledge in humans, machines, and everything else we can think of is mediated by processes between us and the world (Markman and Dietrich, 2000, and in press). The best we can do, logically, is to check these mediated states with each other. This is all right, though. We have five senses. And even in a given modality, we have lots of perspectives on the given category. So we can *triangulate* and usually get it right. "Triangulate" means "hold two or more concepts fixed (i.e., assume

you've got them right) and then figure out what a third has to be." If you make a mistake "That's not a tiger;" Oops...Chomp. Then you lose. But often enough, we don't lose.

And that's the way semantics works. (Admittedly, this is light-years from a theory of semantics, but it does seem pointed in the right direction.)

3.3 Epistemology.

Epistemology is also more than Fodor says it is. Fodor says that epistemology is primarily picking out good instances. If you have a concept, C , you can pick out good instances of concept C , or good members of the category of C -things. But epistemology, is also the ability to figure out when you are in *error* (epistemology is also more than picking out good instances and error, but nevermind that for now). Error is very important. When designing a thinking thing (whether you are Mother Nature or an AI researcher), error is your friend. You can use it to triangulate and gauge how good your concepts are (relative to each other, of course). And doing this produces concepts that have meaning both relative to each other (which is all that we can actually do) and with the world (which is a side-effect of keeping the concepts in synchrony with themselves – isn't it nice things turned out that way?).

4. Conclusion.

So how do concepts compose? In the usual case, they don't. Fodor is right: We can't know what a pet fish is merely by combining "pet" and "fish." But the inference to be drawn from this is that concepts don't usually compose, not that concepts aren't epistemic capacities. They clearly are. To learn what a pet fish is, you have to commune with known pet fish.

Ok, so how do concepts compose, when they do? Ahhh . . . this is a very interesting question. The answer, I think, is that when concepts compose *they form abstractions of each other*. This is true even of "pet fish." I discussed this in my last editorial (Dietrich, in press). It is because they are abstractions -- that is, it is because they combine the way that they do -- that concepts have the power that they do. Abstractions leave out information, yet the abstractions

produced have a wider applicability than the concepts or representations they were abstracted from. The reason for the wider applicability is that concepts are abstract data types that are *first-class objects*. Hence, concepts encapsulate the functional information about their construction and use. Concepts also encapsulate functional information about what other concepts they are linked to and how they are linked. With all this encapsulated information, concepts have, in a sense, more information than the percepts and low-level information they were constructed from. Hence, they have more power. And this conceptual power is what gives our epistemic capacities the power they have.

This isn't a theory of concepts. But perhaps there are glimmerings of one, here. And it looks nothing like Fodor's. Now all is really well. Even with Fodor, for he wouldn't be happy if anyone agreed with him.

5. References.

- Barsalou, L. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (eds.) *Similarity and analogical reasoning*. Cambridge, UK: Cambridge Univ. Press, 76-121.
- Dietrich, E. (in press). AI, Concepts, and the Paradox of Mental Representation, with a brief discussion of psychological essentialism. *J. Exper. and Theor AI*.
- Fodor, J. (1998a). *Concepts: Where cognitive science went wrong*. Oxford, UK: Oxford University Press.
- Fodor, J. (1998b) *In Critical Condition*. Cambridge, MA: MIT Press.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7, 155-170.
- Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (eds.) *Similarity and analogical reasoning*. Cambridge, UK: Cambridge Univ. Press, 199-241.

Margolis, E. and Laurence, S. (1999). *Concepts: Core readings*. Cambridge, MA: MIT Press.

Markman, A. and E. Dietrich (in press). Extending the classical view of representation. *Trends in Cognitive Science*.

Markman, A. and E. Dietrich (2000). In defense of representations. *Cognitive Psychology* 40, pp. 138-171.

Murphy, G. and Medin, D. (1985) The role of theories in conceptual coherence. *Psychological Review*, 92.