**Against Informational Atomism**

Michael C. Frank

Cognitive science is in many ways preoccupied with concepts. One reason for this preoccupation stems from the role of concepts in what is commonly called the Classical view of cognition: the view that knowledge is represented and manipulated through a symbolic language of thought. The word-level constituents of this "language" are primitive concepts, which can be combined to produce an infinite variety of other complex, sentence-level concepts. Concepts are therefore mental states which have both representational properties that carry information about our environment and causal properties that govern our behavior. On this view, cognition is the process of manipulating both concepts and propositional knowledge about concepts—processing our knowledge about the world in order to produce behavior.

Jerry Fodor, in *Concepts: Where Cognitive Science Went Wrong* (1998), presents two options for defining the contents (or meanings) of primitive concepts, the building blocks of the Classical view. The first is Inferential Role Semantics. The primary thesis of Inferential Role Semantics is that having a primitive concept is partially a case of having the inferential relations that help to constitute that concept. So, having the concept BACHELOR is at least partially a question of knowing that bachelors are unmarried men, and hence that BACHELOR is inferentially related to UNMARRIED and MAN. The arguments that Fodor levels against Inferential Role Semantics target both definitional theories and prototype theories of concepts; Fodor's intent is to replace both with a different account of the content of primitive concepts. Informational Atomism, his alternative account, is defined by two theses: first, informational semantics, that for a mind to have a concept is for that mind to be in some sort of relationship to the world; and second, conceptual atomism, that primitive concepts—which make up the bulk of

Fodor's conceptual taxonomy—have no internal structure. In other words, on the Informational Atomist account, having the concept BACHELOR is simply a question of having your atomic, unstructured mental representation BACHELOR line up with an actual bachelor.

In this paper, I will argue that Informational Atomism, Fodor's theory of concepts, fails to live up to its promise as a generalized theory of representation. In particular, I will examine how it fails to account for data provided by a classical task in the psychological problem solving literature, the candle task. By showing how all plausible construals of the Fodor theory fail to account for laboratory data, I hope to suggest that the theory has serious shortcomings, especially in its treatment of inferential knowledge about concepts.

## I. Fodor's Theory

In this section, I will walk though a quick demonstration of how Fodor's theory of concepts functions. We will look at the decision to pet and then feed a gray cat. The account begins with the formation of a mental representation of the cat. According to Fodor's framework, concept individuation occurs purely through the representational properties of the concepts. In other words, a CAT token is created in the mental representation space only if the representational requirements of the CAT concept are satisfied. How can these requirements be satisfied? Only by the identification of an object in the environment that satisfies the stereotype associated with CAT. We might imagine this to be some kind of constraint-satisfaction process, but the details are neither clear nor necessary to the rest of the account.

Next, we notice that the same object satisfies the stereotype for the concept GRAY. We token a GRAY representation, and by compositionality we can compose these two representations into GRAY CAT. Now both GRAY and CAT are primitive concepts with no

internal structure, but they do have what Fodor calls "causal properties"—properties that govern the way they influence our behavior. So it could be the case that CAT has the causal property that petting the creatures represented by them gives me pleasure. So I can decide to pet the creature represented by GRAY CAT based on this causal property of CAT-hood.

Now imagine that the gray cat meows. I can now engage some of the knowledge about CATs from my knowledge store, maybe by putting it under GRAY CAT in the following way:

GRAY CAT
    NOISY(GRAY CAT)
    NOISY(…CAT) -> HUNGRY(…CAT)

and now if we treat …CAT for a schema indicating any kind of cat, then by modus ponens we know that

GRAY CAT
    NOISY(GRAY CAT)$_1$
    NOISY(…CAT) -> HUNGRY(…CAT)
    HUNGRY (GRAY CAT)

Finally, by the causal properties of HUNGRY, we know that hunger is undesirable, and that to relieve hunger, we need to give the hungry creature some food.

Of course, nothing in this scenario is as simple as it seems: I'm merely walking through it to provide an example of how Informational Atomism might plausibly function. Sometimes cats are noisy when they're not hungry, and there are some type of cats you shouldn't feed when they're noisy even if it means they're hungry (stray cats or cats on television, for example). But at least this should give an idea about how concept individuation (tokening) happens only by being in a nomic mind-world relationship: the only way you have a CAT token in your mental representation space is if it represents some kind of cat in the world. Likewise, the concept CAT is, by conceptual atomism, atomic: it is not constituted by any other concepts, nor does it have

any internal structure. So now that we have a picture about how Fodor's model of representation functions, we can move on to examine some psychological evidence that should bear on it.

## II. Functional Fixedness Tasks

Functional fixedness tasks are classic tasks in the psychology literature in which subjects are presented with a situation that requires them to use a familiar object to perform a novel function in order to solve a particular task. In this article I will look at a specific task in the literature, known as the candle task, to illustrate differences in the process of representation between individuals.

In the candle task, participants are given a candle, a book of matches, and a box of tacks. They are asked to use these materials to mount the candle on a wall so that it burns normally and does not drip wax onto the floor. The optimal solution to the task is simple: empty the tacks from the box, tack the empty box to the wall, and then tack the candle to the box. However, many participants do not find this solution: they try to tack the candle to the wall, attempt to use the matchbook as a cradle, or explore many other creative solutions that are more difficult than the optimal solution. (Glucksberg & Weisberg 1966)

The candle task is known as a functional fixedness task because when subjects achieve the optimal solution, they do it by finding a novel function for the box: they use it as a support, rather than as a container. On the other hand, when they fail it is because they do not find this novel function, leading psychologists to speculate that its function is already "fixed" by its presentation with the tacks inside it. However, if the task is presented with the tacks piled next to the box, rather than inside it, then the task is trivial and nearly all participants achieve the

optimal solution. The candle task is a task about representation in that the mode of presentation of the materials is crucial to participants' success in the task.

## III. Applying Informational Atomism to the Candle Task

Given that Fodor's theory is a theory of mental representation, we should expect it to be able to account for data such as those provided by the candle task. The fundamental question that it should be able to help us answer regarding this task is how the mental representations of those participants who solved the task differ from the mental representations of those participants who failed to solve it. We can start by trying to draw up a potential map of both the solvers' and non-solvers' mental representation spaces. One possible hypothesis would state that in the representations of the candle task, non-solvers simply do not represent the box. We could describe this proposal as follows:

| | |
|---|---|
| CANDLE | CANDLE |
| MATCHBOOK | MATCHBOOK |
| TACKS | TACKS |
| BOX | |
| Solvers' Representation Space | Non-solvers' Representation Space |

However, there is a problem with this explanation. Non-solvers know that the box is there, in some form. If you ask them whether they were given a box, they will affirm that they were, even though they are not always able to spontaneously produce the box in their descriptions of the materials. Glucksberg and Weisberg note that "solvers report the box as a separate verbal unit, e.g., 'a box.' Non-solvers often must be prompted before they report the box, sometimes cannot report it at all, and when they do report it, refer to it in a verbally undifferentiated manner, such as 'a box full of tacks.'" (Glucksberg and Weisberg 1966)

The problem with the Fodor account is that there is no subtlety to the way it represents objects; there is only one way that an object can be represented, thus the theory has no way of accounting for the verbal undifferentiation of the box. On the conceptual atomist account, either the box is present or it is absent. In order to have any kind of knowledge attached to it, the box must be represented as an object, so we cannot even fall back on a picture that includes either (1) or (2):

| | |
|---|---|
| CANDLE | CANDLE |
| MATCHBOOK | MATCHBOOK |
| TACKS | TACKS |
| BOX | (1) IN(TACKS,BOX) |
| | (2) BOX OF(TACKS) |
| Solvers' Representation Space | Non-solvers' Representation Space |

Neither of the two options presented is viable. In the case of (1) IN(TACKS,BOX) refers to an object that is not represented, so the expression as a whole fails to refer. And in the case of (2), on the Fodor account, there are only two ways to interpret an expression like BOX-OF. Either it must be composed of BOX and OF or it must be its own primitive concept, and a wholly different entity. But it cannot be composed of BOX and OF because then it would be the same as representing something like PAPER BOX or RED BOX—it would still only be semantically viable if it represented a box. And likewise, it cannot be its own separate entity (on analogy with something like OUNCE OF ); it is very strange to think of a conceptual unit, BOX OF that has no inferential links to BOX. To give only one reason why this is a strange idea, making BOXOF a primitive concept would require a separate stereotype for what a BOX OF is, independent from what a BOX is. We can easily take this idea to absurdity and require separate stereotypes for nearly every complex concept. Fodor may believe that a large part of the conceptual lexicon is primitive, but surely this is not what he is referring to.

We have only one more option to distinguish solvers from non-solvers, according to the model of representation that we are currently working under. They can have different knowledge about the relevant objects. But there are still two ways that this proposal could be implemented: solvers could have knowledge in their knowledge store that non-solvers do not, or they could have the same knowledge but access it in a way that non-solvers do not. However, there are several reasons for ruling out the possibility that solvers know something that non-solvers do not. First, by varying the presentation of the candle task, we can change drastically the percentage of participants that achieve a solution. In a written version of the task, simply underlining relevant materials in the stimuli ("on the table there is a candle, a box of tacks, and a book of matches…") increases the percentage of solvers from 25% to 50% (Frank and Ramscar 2003). This kind of experimental manipulation changes representation, but not by changing subjects' knowledge. Second, and more importantly, assuming that finding a solution revolves around some kind of prior knowledge that participants are assumed to have badly misconstrues the process of problem solving. There are an infinite number of functional fixedness tasks that one could dream up, using a wide variety of objects, some familiar and some novel, and there is no reason to suppose that every solver of each of these would have knowledge about the constituents of the problem that non-solvers did not.

Therefore, we need to explore the possibility that solvers have greater access to their knowledge about boxes or hammers than non-solvers do. Just how would this proposal work within Fodor's model? Well, we can imagine that in this model there is some mechanism that adds knowledge from the knowledge store to the representation space, much the way NOISY(…CAT) -> HUNGRY(…CAT) was added in our earlier example. Given this tool, we can draw up the following proposal for differences in solvers' and non-solvers' representations:

| | |
|---|---|
| CANDLE | CANDLE |
| MATCHBOOK | MATCHBOOK |
| TACKS | TACKS |
| BOX | BOX |
| IN(TACKS, BOX) | IN(TACKS, BOX) |
| CONTAINS(BOX) | CONTAINS(BOX) |
| SUPPORTS(BOX) | |
| Solvers' Representation Space | Non-solvers' Representation Space |

where SUPPORTS(X) or CONTAINS(X) means that X can function as a support or a container.

Unfortunately, this account suffers from many of the same problems as earlier proposals. First, it lacks a principled way to distinguish solvers' representations from non-solvers' representations: it simply pushes off the task of problem solving to some kind of unspecified machinery that retrieves inferential knowledge about the represented objects. So this move comes close to violating one of the dictates that we started with: that functional fixedness problems were interesting because they were problems about representation. This complaint does not, of course, rule out the account as true; it simply takes away much of the interest in testing Informational Atomism if the empirical complexities have to be passed off to another part of the system.

Second, the account fails to deal with the issue of what inferential knowledge comes along naturally with the tokening of a representation of an object. Even if BACHELOR is not semantically defined as UNMARRIED MAN, that knowledge must have a privileged inferential link to the concept in some way. In other words, it must be true that someone who not only possesses a BACHELOR stereotype but also knows how to properly manipulate the concept BACHELOR in inferential settings must represent UNMARRIED and MAN almost automatically when BACHELOR is tokened. Perhaps there are some exceptions to this, but it must be true on the whole. We can pass this job on again to our unspecified machinery for

pulling inferential knowledge in the representation space, but then we have again failed to address the issue at hand. For the question of whether CONTAINS has a privileged inferential link to BOX, and whether SUPPORT does as well, will be crucial to determining why non-solvers have one and solvers have both.

Third, it still fails to deal with the data from the Glucksberg and Weisberg paper. On this account there is still no reason why some participants might fail to name the box as one of the materials for solution: if the box is represented, then it must be represented based on some objects' satisfaction of the box stereotype. Therefore, there is no reason to predict any of the problems in memory or in verbal differentiation that we see in the actual laboratory task, and in turn no way to explain them.

## IV. Conclusion

The problem here is deep: the representational machinery provided to us by a Classical account such as Fodor's Informational Atomism is not delicate enough to provide an accurate model of the phenomena found in real problem solving data. This point is not a falsification of the Fodor account. Rather it is an identification of a problem, and in fact a domain of problems, in which the Fodor account has little to say about real psychological data. Any explanation of the phenomena observed in the laboratory must come either from ad hoc additions to the account or from mechanisms entirely outside of it.

## References


Fodor, Jerry A., and Z. W. Pylyshyn. "Connectionism and cognitive architecture, a critical analysis." Cognition 28 (1988): 3-71.

Fodor, Jerry A. Concepts: A Potboiler. Cognition 50 (1994): 95-113.

Fodor, Jerry A. Concepts: Where Cognitive Science Went Wrong. Oxford, UK: Oxford UP, 1998.

Frank, Michael C., and Michael Ramscar. "How do Presentation and Context Influence Representation for Functional Fixedness Tasks?" Proceedings of the 25th Annual Meeting of the Cognitive Science Society, 2003.

Glucksberg, S., and R. W. Weisberg. "Verbal behavior and problem solving: Some effects of labeling in a functional fixedness problem." Journal of Experimental Psychology 71 (1966): 659-664.