

# *Undergraduate Review*

---

*Volume 6, Issue 1*

1993

*Article 7*

---

## Methodological Solipsism Reconsidered: Is There Anybody Out There?

Peter Asaro '94\*

\*Illinois Wesleyan University

This paper is posted at Digital Commons@ IWU. <http://digitalcommons.iwu.edu/rev/vol6/iss1/7>

© Copyright is owned by the author of this document.

## **Methodological Solipsism Reconsidered: Is There Anybody Out There?**

Peter Asaro

In his paper "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," Jerry Fodor distinguishes the Computational Theory of Mind as a stronger version of the Representational Theory of Mind. He goes on to argue that a Naturalistic Psychology which considers organism environment relations is impossible and unproductive. He concludes that only a computational view of the mind is relevant and therefore the task of psychology should be limited to a methodological solipsism. While I agree with most of Fodor's argumentation, he makes one mistake which causes him to miss an alternative. I argue that pure methodological solipsism, done to the exclusion of any content, would be extremely difficult, if not impossible. I propose a more moderate view of computational psychology which incorporates some of the intuitions of a semantic psychology, while still maintaining the conditions put forth by Fodor.

In drawing the distinction between the computational and representational theories of mind, Fodor develops what he calls the formality condition. Fodor describes the representational theory of mind as the broad view that organisms create mental representations and that propositional attitudes are the relations that organisms bear to these mental representations. The computational theory of mind is the narrower view that these representations are purely symbolic and formal-syntactic in nature. According to the representational theory, mental states can be distinguished by the content of the relevant mental representations and the relation that the subject bears to those representations (i.e. thinking, doubting, believing). It is the claim that mental processes are purely syntactic that makes the computational theory stronger than the representational theory. Simply stated, the formality condition is the notion that the formal aspects of thoughts are all that is essential to type distinguishing thoughts. Fodor construes formal to mean the precise opposite of semantic; syntactic is an imprecise opposite since it cannot apply to all cases, such as rotating an object.

What makes syntactic operations a species of formal operations is that being syntactic is a way of not being semantic. Formal operations are the ones that are specified without reference to such semantic properties as, for example, truth, reference and meaning. (309)

With his new-found formality condition in hand, Fodor then reexamines the controversy between Rational Psychology and Naturalism. Rational psychology, embraced by Rationalist and Empiricist alike, holds an introspectionist construal of type individuating mental states. That is, mental states are type identical if and only if they are introspectively indistinguishable, and since introspection cannot distinguish perception from hallucination, or knowledge from

belief, one's mental states might be as they are even if the world were radically different. Naturalism sees psychology as a branch of biology, and holds that one must see the organism as embedded in a physical environment. Naturalism is behavioristic in that it seeks to trace the organism/environment interactions which govern behavior. Fodor argues that the computational theory of mind shifts the debate from introspectionism vs. behaviorism and aids both sides.

If we embrace the computational theory of mind, it is easy to think of the mind/brain as a kind of computer. It has memory, capacities for scanning and altering its memory, and transducers of information from the outside world ("oracles" as Fodor calls them) which are roughly analogous to senses. Thus construed, the significance of "environmental information upon [mental] processes is exhausted by the formal character of whatever the oracles write [to the memory]"(314). Whether the information is true or not is insignificant: the oracles could be writing accurately about the environment or could be inputs for the typewriter of a Cartesian demon. He thereby arrives at his methodological solipsism.

I'm saying, in effect, that the formality condition, viewed in this context, is tantamount to a sort of methodological solipsism. If mental processes are formal, they have access only to the formal properties of such representations of the environment as the senses provide. Hence, they have no access to the *semantic* properties of such representations, including the property of being true, of having referents, or, indeed the property of being representations of *the environment*. (314)

The consequence of methodological solipsism is that researchers need not consider the semantic aspects of thought, which are taken to be misleading and fruitless. Examples are useful here. Take the computer program SHRDLU, developed by Terry Winograd. SHRDLU takes commands and accepts descriptions of a simple "block world". It can 'manipulate' geometric objects in its virtual world and issue 'perceptual' reports on the state of this virtual world, and of previous states from memory. All of the information it receives is false. It lives in a completely notional world, deceived by its inputs. All of SHRDLU's beliefs are false. It is effectively an organism being deceived by a malignant demon, a Cartesian nightmare. All that is necessary for a psychology of SHRDLU is its formal properties: this is methodological solipsism. Trying to do any other kind of psychology on SHRDLU, one which sought to find what made its statements about the "block world" true or the referents of SHRDLU's statements, would miss the point of why SHRDLU acts and believes the way that it does.

Fodor then tries to defend definitively his requirement that any psychology must honor the formality condition and therefore methodological solipsism. To do this, he uses the distinction between opaque and transparent construals of meaning. Fodor argues that "when we articulate the generalizations in virtue of which behavior is contingent upon mental states, it is typically an opaque construal of the mental state attributions which does the work" (317). By such a construal, the belief that the Evening Star rises in the East is type distinct from the belief that the Morning Star does, whereas by a transparent construal these beliefs

would be type identical because they are coreferential. The belief about the Evening Star is formally distinct from the belief about the Morning Star: these beliefs must be distinct, since they result in different behaviors (an individual having a belief about the Evening Star could believe something quite different about the Morning Star). The only reason for saying they are type identical is that they are coreferential-- but reference is a semantic quality can not adhere to the formality condition. Thus only a fully opaque construal of beliefs respects the formality condition.

And where does all this get us? According to Fodor, it is possible that "the formality condition *can* be honored by a theory which taxonomizes mental states by their content." He explains, "it's because different content implies formally distinct internal representations and formally distinct internal representations can be functionally different-- can differ in their causal role" (324). In fact, this is the basic idea of modern cognitive psychology: to connect computation and content and thereby use the intensional properties of mental states to explain their causal properties. Yet Fodor pursues his intuition that mental states can only have access to the formal properties, and that "the contrary view [that is, any view which incorporates meaning, content or reference--the semantic properties] is not only empirically fruitless but also conceptually unsound." (325)

Having satisfied all of his intuitions regarding the computational theory, the formality condition, methodological solipsism, and rational psychology, he moves on to consider the possibility of naturalistic psychology. While he cannot argue against it theoretically, for it "*seems* very reasonable," he argues that it will be impossible to realize. First, he contends that the goal of a naturalistic psychology would be to "specify environmental objects in a vocabulary such that environment/organism relations are law-instantiating when so described" (334). But this necessarily requires that the vocabulary be "scientifically accurate". This is motivated by the Twin Earth examples of Putnam. He then invokes an argument from Bloomfield:

- (a) We don't know relevant nomologically necessary properties of most of the things we can think about and
- (b) it isn't the psychologists' job to find them out (334).

In fact, it is the physicists' job. In order to have a naturalistic psychology, we would have to wait for the completion of physics in order to know any of the terms of the objects we think about. Fodor doesn't believe that naturalism is necessarily wrong; he just thinks that it will take at least another 300 years of science before it can start.

Fodor then attacks the research strategies of the naturalists. The naturalists are very consistent and forcefully entrenched in their philosophy. But when it comes to actually doing their research, however, they completely ignore everything they have said. They have to ignore it: there just aren't any words in our language which are defined precisely enough to use in a naturalistic psychology. In reality they just fudge, saying that a term refers to 'just what it refers to'. They claim to look at the relations and interactions that result from the reference but fail to understand the precise nature of the referent. Fodor argues that because of this fudging, naturalism will never succeed.

Fodor concludes by saying that there are only two options in psychology. First, we could just fudge and try to do naturalism. Or second, we could try a computational psychology in which mental states are type individuated opaquely. But he has missed an alternative. Fodor doesn't argue that there is no naturalistic psychology: it's just not something we are capable of pursuing. Fodor also doesn't argue that there are no semantic qualities. He just contends that we don't really know what they all are, have no way of figuring them out, and don't really need them to do computational psychology anyway. Fodor chooses computationalism as a research strategy because it is productive, though only by default, and has a practical hope of succeeding when "for methodology, practical hope is everything."(337)

I argue that Fodor has missed the best alternative. I will first show how difficult a purely formal computational psychology will really be. I will then propose my own version of computational psychology, a weak computationalism, and show why it would work better, and how it could satisfy the opacity condition without being purely formal.

I maintain that conducting a computational psychology which attempts to type individuate our mental states in a purely formal manner will be virtually impossible. A computer analogy is in order. What better way to test computational theory?

In computer science, the programs that actually control the computer are written in assembly code, also know as machine code, essentially a very cryptic binary language. The problem is that it is incredibly difficult and tedious to write even simple programs in assembly, because its ones and zeros don't mean much to even a skilled programmer. Programmers have therefore created high level programming languages and compilers, such as C and Pascal, which use English-based commands which make sense to a skilled programmer. The compiler then translates the high-level language into assembly code, since this is the only language the computer can really "speak". What computational psychology requires is to look at only the "machine code of the brain", if such a thing even exists. But the most difficult task in all of programming language research is reverse compiling, turning the compiled machine code back into an intelligible program. Reverse compiling is done only rarely, and usually only done for viruses: when it is necessary to know what the virus does in order to protect the system, yet there is only the compiled version of the program available. What makes it so very difficult is that assembly code is purely formal, free of any content: if it had content, they wouldn't need to reverse compile it. Assembly code just doesn't mean anything to a programmer, and *it* was at least developed by humans.

Will the computational psychologist be any better off than the reverse compilers? I contend that they will be far worse off. Computational psychologists don't even have an "assembly code" for the brain yet (there may not even be one), plus the fact that the brain is massively parallel and thousands of times as complex as any computer, and the certainty that the mind's "program" will be unimaginably complex. This all adds up to an impossible task for the strong computational psychologists. How can computational psychology work?

Fodor makes one fatal error in his argumentation. Fodor lumps content together with truth and reference as a semantic quality, and throws out semantic qualities as being transparent. Content would seem prima

facie to be a transparent semantic quality: it is commonly construed as a consequence of reference or "aboutness." Many may argue that content is external, and Fodor accepts this after some misguided consideration.

Fodor's primary concern is maintaining *nontransparency* in psychology. But Fodor believes, "The trouble is that nontransparency isn't quite the same notion as opacity." (320) While it is clear that transparently coextensional beliefs may be opaquely type distinct, according to Fodor,

there are nevertheless some semantic conditions on opaque type identification. In particular:

- (a) there are some cases of formally distinct but coextensive token thoughts which count as tokens of the same (opaque) type (and hence as identical in content at least on one way of individuating contents); and
- (b) *non*-coextensive thoughts are *ipso facto* type distinct (and differ in content at least on one way of individuating contents) (320).

Thankfully, Fodor gives examples to explain these conditions so I can show why he is wrong. Cases of type (a) consist of what he thinks are (opaquely) type identical, yet formally distinct thoughts. Cases of type (b) involve what he thinks are formally identical mental representations which characterize opaquely distinct mental states.

- (a) I think I'm sick and you think I'm sick. What's running through your head is 'I'm sick'; what's running through your head is 'he's sick'.
- (b) Sam feels faint and Misha knows he does. Then what's running through Misha's head may be 'he feels faint.' Misha feels faint and Alfred knows he does. Then what's running through Alfred's head, too, may be 'he feels faint.'

These are the obvious results of two mistakes on Fodor's part: an assumption that the only way to individuate content is by reference and extension, and a confusion between what's *in* the head with what's running through it. He addresses the latter in footnote 8:

One might try saying: what counts for opaque type individuation is what's in your head, not just what's running through it. So, for example, though Alfred and Misha are both thinking, 'he feels faint,' nevertheless different counterfactuals are true of them: Misha would cash his pronoun as: 'he, Sam,' whereas Alfred would cash *his* pronoun as: 'he, Misha.' The problem would then be to decide *which* such counterfactuals are relevant, since, if we count all of them, it's going to turn out that there are few, if any, cases of distinct organisms having type identical thoughts (321).

Fodor is wrong to ignore the counterfactuals as unimportant. If we want to explain behavior, it is important that we understand the world view of the subject, at least as it relates to the context of the situation. This is the

Produced by The Berkeley Electronic Press, 1993

only way we can come to understand the relations we stand in with our mental representations (believing, doubting, etc.). It seems obvious to me that distinct organisms will only very rarely have type identical thoughts, yet Fodor doesn't want to admit this. There are many cultural and personal differences in how we see the world, how we 'cut-up' the world, and what we find to be relevant to us. The point is that if we really want to know what is going on in someone's head to cause a behavioral response, you will have to know some of the relevant terms.

Let me further explain with a Twin Earth example of my own. Bob1 and his Doppelganger, Bob2, are completely identical as far as their life experiences, memories and beliefs. Except, Bob2 recently developed a strange case of paranoia and thinks his boss is trying to kill him. The bosses' secretaries in the Bobs' respective worlds give the Bobs identical notes that their bosses want to see them. Both Bobs will be having the thought, 'My boss wants to see me,' but Bob1 may also be thinking, 'It's time for that big raise,' while Bob2 may also be thinking, 'My boss is going to kill me when I get to his office'. It seems that the context in which each Bob has their belief will effect greatly the consequence of that belief on subsequent beliefs, and therefore the behavioral responses. Bob1 will go beaming into his boss's office, while Bob2 may have a breakdown or go running home.

What Fodor fails to realize is the importance of context and the relations of the current belief to other related beliefs. Fodor is afraid that if we have to seek out relevant beliefs, there will be far too many and no way of telling which are relevant *enough*. But being able to type individuate a single belief in isolation from all other beliefs, and especially from the relevant beliefs, is useless and meaningless. Thought and computation are complex dynamic streams of beliefs and calculations. For Fodor to look at them in such a simplistic and static way is simply unrealistic. Only when thoughts are tied together with other thoughts into the complex web of beliefs that is one's world knowledge can we begin to understand what is going on 'in the head' o cause a behavior. And *this* is what a cognitive psychology should be trying to do. Besides, it is absurd to think that an organism could only have just one thing 'in mind' at a given time, remember Miller's "magical number seven."

As for the problem of deciding which related beliefs are relevant *enough*, this isn't as hopeless as Fodor might think. Certainly the organism knows which are relevant to it. The organism is weighing the justification of its beliefs and the outcomes from possible behaviors every time it makes a decision or exhibits a complex behavior. To understand these relations just is what the organism has 'in mind.' Another route might be to define massive parts of the web, or even divide the web into significant realms. Who knows what we will find when psychology starts to actually study beliefs.

If we can find the content relations of a belief token in such a way that when we find them we can come to understand how a subject 'thinks about' that mental representation, and how that token stands in relation to other tokens, we can come to understand what the subject is really thinking about<sup>1</sup>. We can do this if we construe content in a holistic

---

1. Do not confuse this 'about' with anything like reference or extension.

Asaro '94: Methodological Solipsism Reconsidered: Is There Anybody Out There sense: for any given token of the system, its meaning is a result of how it stands in relation to the other tokens in the system.

Content is not an external quality, whereby a token means something as a consequence of what it *refers* to in an external *an sich* reality. At most a token could refer to only the sensory data coming from our "oracles." This is as close as we get to having ideas 'about the world.' In this way, we are no better off than SHRDLU, for we may be deceived by a Cartesian demon as well, at least we can't tell from the 'inside'. I have to agree with Fodor that the external semantic qualities: reference, truth and Dasein, are out there (maybe) but we sure can't grasp them. It seems that our beliefs are meaningful introspectively and behaviorally at least most of the time, and we can only think and act meaningfully if we have some grasp of content.

Not only does an organism have access to holistic content, this seems the only way to explain how the organism can grasp the content of its own mental representations and the most viable explanation for the introspective capacities of an organism. Any knowledge we have about 'things in the world' seems to be based solely on our experience of them and our relations to them. There is a pedantic (and very annoying) way to argue against people which looks something like an epistemic regress, in which one simply keeps asking, "What does \_\_\_\_\_ mean?" One learns very quickly that people can't define their terms in a non-circular way. But I argue that content comes in just that circular, holistic way. The token *pencil\** has content because it has relations like: \* is yellow, \* is wooden, \* is used for writing, \* can be sharpened, \* little pink eraser on the end, \* is on my desk right now, and to stored sensory files of the visual and tactile sensations caused by pencil\* experiences. As for each of these tokens, their content comes in part from these same relations, i.e. wooden\* has relations like: \* is a quality of pencils, \* is a quality of some furniture, \* is something from a tree, and more sensory files of the look and feel caused by wooden\* experiences.

The effectiveness of such a notion has been demonstrated in John Anderson's *ACT* model for the mental representation of knowledge. *ACT\** is the computer program which uses the *ACT* model. Its knowledge about its world is provided by the researcher, its representations interconnected in certain relations. The contents of *ACT\**'s beliefs are just like those relations between tokens in the previous example, and appear very tree-like when diagrammed. *ACT\**'s responses are very 'intelligent' in that it only speaks about what it 'knows' and can draw new, nonexplicitly stated relations between indirectly related representations. It is a major accomplishment in AI, but what keeps it from passing a Turing conversation? Probably only the sophistication and hardware. If *ACT\** had sensory inputs, a means to interact with the world, and a means to manage its copious incoming and stored memories, it would seem very intelligent indeed. Again, what is important is not what makes computers intelligent, but that a sufficient and productive psychology of *ACT\** would have to explain a great deal of these internal semantic relations between the tokens of its mental language.

With a holistic account such as this, content can be construed completely nontransparently. I think that if Fodor had considered this possibility, he would have agreed. His main concern is preserving opacity and allowing the researcher to study only what the organism has

'access' to opaquely. Using a holistic account of content is the only way to achieve this. It would also seem that if one can define or assign the meaning of one or several tokens, one would be able to translate the entire mental system, or at least large tracts of it.

By disallowing the researcher any use of content, the strong computationalism of Fodor is like being locked in a Chinese Room (Fodor might prefer Mentalese for this example) without any English rules. The psychologist has to look for the English rules to answer questions in Chinese about stories in Chinese. Imagine trying to rebuild the entire Chinese vocabulary and grammatical rules necessary to answer any question in Chinese without ever translating a single word to English or knowing what a single character means. It would be impossible. Yet this is the requirement of strong computationalism. On the other hand, if we were to translate a few of the Chinese symbols and words for the psychologist, the task is feasible (as evidenced by Chinese-English dictionaries).

What weak computationalism calls for is the "translation" of a few of the subject's mental representations into a contentful medium: English or approximations of the researcher's mental representations. In so doing, one may begin the mammoth task of computational psychology. In order to "translate" the first terms, the researcher must use intuition and look at causal relations. The idea is that most, or at least some, of our mental states will be roughly as we expect them to be. When two people look at a tree, it seems reasonable to assume that they are both having similar, though certainly not identical, mental states. Only by finding all of the relevant relations of an individual's belief state will we find its precise content.

The relativists may argue that no direct translations between languages are possible or that beginning with imprecise translations will result in the entire system being incorrectly translated. What is important to remember is that the translations are only a tentative starting point. When the translations are imprecise or incorrect, we will eventually run into problems translating the system. When translators first began translating Greek, they certainly translated "love" imprecisely. Only after having translated enough Greek did they realize there are three different kinds of "love" in Greek, and revise their translation. It would seem impossible to translate an entire system incorrectly (allowing something like the inverted-color spectrum possibilities which seem inconsequential to such a project). Any computational psychology will only give us a view into what is going on 'in the head'; it will never really tell us what it is like to be in the world view of the subject. It may be impossible to translate the sensory representations, but we can look at the relations to see how they are 'cut-up'; sensory representations also seem prime candidates for applying intuitions about causal relations.

The strong computationalist may argue that the use of imprecise definitions is really no different or better than the fudging of the naturalists. It is important to remember that the naturalists are only fudging because they can't get the precise terms that they need. Their real problem is that they are claiming to study the nature of the subjects' thoughts as a consequence of the essential character of an external object and causal relations between the two. They fail because they don't know this character nor do they understand the causal relations. The 8

**weak computationalist isn't trying to make any monumental claims based on their imprecise definitions. They are merely using these as tools: tools to find more definitions and more precise definitions. The weak computationalist is also going to be far more productive than either the naturalist or the strong computationalist, where productivity is all important to Fodor.**

**Fodor has shown that computationalism is necessary for a cognitive psychology. I have demonstrated a need for a more moderate version of computationalism than Fodor was willing to provide. I have also shown how content can be utilized in a nontransparent way. And I have answered the objections as to how a weak computationalism can begin, and why it doesn't fall into the same Stygian Abyss that claimed naturalism. Not only do I think that weak computationalism is the proper way to conduct cognitive psychology, but this is the way that it is already being conducted, at least in the fields of knowledge representation and linguistics.**

**Bibliography**

Fodor, Jerry A.; "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology"; in Mind Design; ed. John Haugeland; MIT Press; Cambridge, MA; 1981.