



**Review Essay: From Cognitive Science to Folk Psychology: Computation,  
Mental Representation, and Belief**

Terence Horgan

*Philosophy and Phenomenological Research*, Volume 52, Issue 2 (Jun., 1992), 449-484.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*Philosophy and Phenomenological Research* is published by International Phenomenological Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ips.html>.

---

*Philosophy and Phenomenological Research*  
©1992 International Phenomenological Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2001 JSTOR

# From Cognitive Science to Folk Psychology: Computation, Mental Representation, and Belief

TERENCE HORGAN  
*Memphis State University*

This is a review essay of three books: Baker (1987), Garfield (1988), and Cummins (1989). All three take up the philosophical foundations of computational cognitive science, an approach to mentality that was essentially the "only game in town" in cognitive science prior to the recent emergence of connectionism. And all three discuss interconnections between computational cognitive science and common-sense intentional psychology. The views of Jerry Fodor about these matters are of course extremely influential, and are discussed prominently in all three books. So I will begin by discussing some central themes from Fodor (1981, 1987), as groundwork for what I will say about Baker, Garfield, and Cummins.

1. **Fodor.** I will characterize a package of interrelated views held by Fodor as a list of numbered theses, prefixed with 'F' for 'Fodor'. In cases where a given thesis is based on preceding ones, this will be indicated in parentheses. (Such dependencies are not necessarily a matter of straightforward entailment, however.) Here and henceforth, references to states, representations, processes, and the like will refer to *types* rather than tokens; when reference is to tokens of such types, this will be indicated explicitly.

For a start, there is Fodor's firm allegiance to the computational conception of the mind:

(F.1) The computational conception of the mind is correct.

The computational approach, he holds, brings in its wake the following commitments:

- (F.2) Each mental state is identical to a complex monadic property, instantiable by a cognizer: the property of bearing a specific *relation* to a specific *mental representation*. (1)
- (F.3) Mental states and processes are type identical to computational states and processes.
- (F.4.a) Mental states and processes are symbolic; i.e., they involve mental representations with compositional syntax and semantics. (3)
- (F.4.b) Mental states and processes are formal; i.e., they apply to representations in terms of their syntactic features, independently of their semantic content (if any). (3)

Theses (F.4.a) and (F.4.b) are so labeled in order to emphasize their package-deal nature; in Fodor's view, the essence of computational states and processes is that they are both symbolic and formal. He does allow that *some* mental states and processes, such as mental images, might involve computational operations over representations with some kind of non-language-like structure. But for the most part, the relevant structure posited by computationalism is syntactic. Exceptions aside, then,

- (F.5) Each mental state is identical to a state consisting of a specific formal/computational relation to a specific syntactically structured mental representation. (2–4)

And this in turn leads to three principles about the semantics of mental representations, he thinks:

- (F.6) Mental processes have no access to the semantic properties of mental representations. (5)
- (F.7) For purposes of scientific taxonomy, the mental states of humans are identical in content to the mental states of their Twin Earth *doppelgangers*. (5)
- (F.8) The only kind of mental content that has causal/explanatory relevance in psychological explanation is narrow content. (5)

Thesis (F.6) is, in his view, essentially just a consequence of the fact that computational processes are formal: they are defined over the *syntactic* features of mental representations, irrespective of the semantic features:

I'm saying, in effect, that the formality condition...is tantamount to a sort of methodological solipsism. If mental processes are formal, then 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 ref-

erents, or indeed, the property of being representations *of the environment* (Fodor 1981, p. 231).<sup>1</sup>

Thesis (F.7) too he takes to be essentially a consequence of (F.5). For, since the corresponding Earthly and Twin Earthly states *believing that water is good to drink* and *believing that twater is good to drink* allegedly consist of the same computational relation to the same sententially structured mental representation, individuation of mental states by content will coincide with individuation as prescribed by (F.5) only if the relevant kind of content is narrow.

[I]f the *computational* theory of the mind is true (and if, as we may assume, content is a semantic notion par excellence) it follows that content alone cannot distinguish thoughts. More exactly, the computational theory of mind requires that two thoughts can be distinct in content only if they can be identified with relations to formally distinct representations (Fodor 1981, p. 227).

And if all this is right, then it becomes very hard to see how thesis (F.8) could fail to be true too.

Fodor also has views about the status of the propositional attitudes (henceforth, PA's) under the computational conception of the mind. Within computational cognitive science, he maintains, it is very common to posit states that qualify as PA's. Moreover,

(F.9) An adequate computational cognitive science would posit beliefs, desires, and other PA's.

Given (F.9) together with the preceding theses, we get the following derivative claims:

(F.10) Beliefs, desires, and other PA's are among the mental states that an adequate cognitive science would posit, and hence are among the mental states instantiated by humans. (1,9)

---

<sup>1</sup> Although semantic properties like being true, having referents, and being representations of the environment all involve relations between what's in the head and what's outside, it is worth noting that there may well be other kinds of semantic properties that do not involve such relations—viz., properties that are intentional in Brentano's sense, and thus whose instantiation does not require the existence of any system-external entities or states of affairs that answer to the relevant intentional contents. There might be properties of the latter kind that supervene upon what's in the head, and also are enormously rich and varied in content. Conceptions of intentionality like Brentano's have been largely overlooked in recent philosophy of mind. Some philosophers don't believe in "narrow content" at all; and many who do have a much thinner conception of it than did Brentano. For a useful antidote to these recent tendencies, see Tienson (forthcoming).

- (F.11) Each PA is identical to a complex monadic state consisting of a specific computational relation to a specific mental representation whose content is the content of the attitude itself. (5,10)
- (F.12) For purposes of scientific taxonomy, the PA's of humans are identical in content to those of their Twin Earth *doppelgangers*. (7,10)
- (F.13) The only kind of mental content that has causal/explanatory relevance in PA explanations is narrow content. (12)

Implicit in Fodor's treatment of the PA's is a general methodological presupposition about what is required to vindicate PA realism. Consider, for instance, this passage from Fodor (1987):

The main thesis of this book can now be put as follows: *We have no reason to doubt—indeed, we have substantial reason to believe—that it is possible to have a scientific psychology that vindicates commonsense belief/desire explanation.* But though that is my thesis, I don't propose to argue the case in quite so abstract a form. For there is already in the field a (more or less) empirical theory that is, in my view, reasonably construed as ontologically committed to the attitudes and that—again, in my view—is quite probably approximately true. If I'm right about this theory, it is a vindication of the attitudes. Since, moreover, it's the only thing of its kind around (it's the *only* proposal for a scientific belief/desire psychology in the field), defending the commonsense assumptions and defending the theory turn out to be much the same enterprise; extensionally, as one might say. (p. 16)

The theory he has in mind, of course, is the one embodied in the above theses. Now, what I want to stress here is an assumption which, though still in the subtext of this passage, is not far from the surface, viz.,

- (F.14) The PA's are vindicable only if they would be posited by an adequate cognitive science.

What Fodor actually says is that the PA's are vindicable *if* they would be posited by an adequate cognitive science; and he thinks there are good reasons to believe they would be. But what he evidently assumes is that they are not vindicable *unless* belief/desire psychology is destined to be part of mature science. And of course, eliminativists about PA's typically assume this too.

Another important aspect of Fodor's views about mentality is his conception of what is required in order to incorporate realism about intentional mental states into a naturalistic/physicalistic metaphysics. He writes:

I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin*, *charm*, and *charge* will perhaps appear upon their list. But *aboutness* surely won't; intentionality simply doesn't go that deep. It's hard to see, in face of this consideration, how one can be a Realist about intentionality without also being, to some extent or other, a Reductionist. If the semantic and the intentional are real properties of things, it must be in virtue of their identity with

(or maybe their supervenience on?) properties that are themselves *neither intentional nor semantic*. If aboutness is real, it must be really something else.

And, indeed, the deepest motivation for intentional irrealism derives...from a certain ontological intuition: that there is no place for intentional categories in a physicalistic view of the world; that the intentional can't be *naturalized*. (1987, p. 97)

For Fodor, as for other contemporary philosophers, the project of naturalization involves giving a tractable specification, in nonintentional and nonsemantic vocabulary, for a state's being an intentional state with a specific content. Tractability is my own term; and the demand for it is implicit in his and others' writings, rather than being on the surface. Roughly, a tractable specification is a relatively compact, relatively non-baroque, nondisjunctive, cognitively surveyable, formulation of *sufficient conditions* (for some philosophers, sufficient *and necessary* conditions). So we have the following theses about the place of the intentional in the physical world:

(F.15) If there is a place for intentional categories in a physicalistic view of the world, and if a physicalistic view of the world is correct, then the intentional can be "naturalized," in the sense that there are tractable sufficient conditions, formulable in nonintentional and nonsemantic vocabulary, for a physical system to have intentional states.

(F.16) There is a place for intentional categories in a physicalistic view of the world.

(F.17) A physicalistic view of the world is correct.

(F.18) The intentional can be naturalized; i.e., it is possible to tractably specify nonintentional and nonsemantic sufficient conditions for a physical system to have intentional states. (15–17)

On Fodor's own story, the most fundamental locus for intentional and semantic properties is the system of language-like mental representations posited by computational cognitive science:

[I]t's the interpretation of the primitive nonlogical vocabulary of Mentalese that's at the bottom of the pile according to the present view. Correspondingly, we would have largely solved the naturalization problem for a propositional-attitude psychology if we were to say, in nonintentional and nonsemantic idiom, what it is for a primitive symbol of Mentalese to have a certain interpretation in a certain context. (1987, p. 98)

Thus we can round out our list of theses with the following two:

(F.19) At bottom, the naturalization problem primarily involves saying, in a nonintentional and nonsemantic idiom, what it is for a

primitive symbol of Mentalese to have a certain interpretation in a certain context.

- (F.20) It is possible to tractably specify nonintentional and nonsemantic sufficient conditions for a primitive Mentalese symbol to have a certain interpretation in a certain context. (18,19)

Fodor has proposed a formulation of what he tentatively maintains are naturalistic sufficient conditions of the kind called for by (F.20); he calls this proposal the “Slightly Less Crude Causal Theory of Content.” Consideration of the specifics of the SLCCTC is beyond the scope of the present essay.

**2. Alternatives to Fodor.** I will now set forth some alternatives, frequently overlooked in recent philosophy, to various of Fodor’s theses about the computational conception of mind, about commonsense belief/desire psychology, and about the physicalist foundations of intentional realism. I will prefix them with ‘A’ for ‘alternative’, and I will give them numbers that correspond to the various Fodorian theses which they would respectively displace. Some of Fodor’s theses will not be called into question, and thus would be incorporated into the alternative package I will describe. Various kinds of mixing and matching are possible, in terms of philosophical positions one might construct consisting of certain Fodorian theses and certain alternative theses.

The Fodorian theses about the computational conception of the mind are (F.1)–(F.8). Of these, (F.1) is basic; Fodor regards the others as commitments that (F.1) brings with it. And the Fodorian theses about the PA’s *vis-à-vis* the computational conception of mind are (F.9)–(F.13). Of these, (F.9) is basic; he regards the others as commitments that (F.1) and (F.9) jointly bring with them. So suppose for now that both (F.1) and (F.9) are true; here is an alternative metaphysical story to go with them.

The leading idea is that the computational conception of mentality provides not an account of the type identity conditions of mental states, but rather an account of how these states happen to be *realized* in humans (and in other terrestrial cognizers, if any). The operative realization story concerns the functional architecture of terrestrial cognitive systems—roughly, the “engineering design” to which these systems conform, *qua* cognizers—rather than the neurobiological “wetware” that subserves this design. Borrowing from Colin McGinn (1989, p. 171) the term ‘psychotechnics’—an apt name for scientific theorizing about cognitive functional architecture—I will call the relevant relation *psychotechnic realization*. The core claim, then, is that mental states are psychotechnically realized by certain functional/computational states, which in turn are physically realized by certain

neurobiological states. (Thus, the former are physically realized by the latter; realization is transitive.)

For the past 20 years the dominant view in philosophy of mind has been that realization, rather than type identity, is the relation between mental states and *neurobiological* states.<sup>2</sup> But for some reason philosophers have generally failed to realize that realization might well be the relation between mental states and formal/computational states too. This is rather surprising, since one can give a multiple realization argument against the type identity of mental and functional/computational states that is quite similar to the familiar multiple realization argument against type-type *psychophysical* identifications.

The argument goes as follows. Even if (F.1) and (F.9) are both true of all creatures with mentality, the possibility remains open that in different kinds of creatures (say, humans and Martians), the same belief states, with the same contents, are differently realized psychotechnically.<sup>3</sup> Perhaps (i) Martian Mentalese is a system of language-like mental representations distinct from human Mentalese, even though pairs of sentences from these respective mental languages have identical contents. (For instance, maybe Martian Mentalese is in Polish notation, whereas human Mentalese is in standard *Principia Mathematica* logical notation.) Or, even if humans and Martians have a common internal language, perhaps (ii) mental states are psychotechnically realized in Martians via different *computational relations* than in humans. (After all, in general various different algorithms can compute a given (computable) function; accordingly, various different computational relations to internal representations could subservise the same transition function over these representations.) Moreover, again assuming that humans and Martians have a common mental language, it seems likely that there will be numerous *distinct* computable transition functions over this common system of mental representations, each of which conforms as well as any other such function with the *ceteris paribus* generalizations of common-sense intentional psychology. (As they say, one person's *modus ponens* is another's *modus tollens*.) So perhaps (iii) human mental processes conform to one such function and Martian mental processes conform to a different one; if so, then the computational processes that psychotechnically realize human mental states will inevitably differ somewhat from the

---

<sup>2</sup> Accordingly, the natural kind properties posited by the special sciences are generally considered naturalistically respectable despite not being identical to natural kind properties posited by physics; cf. Fodor (1974).

<sup>3</sup> For that matter, the possibility remains open that the same belief states are differently realizable in *humans*, even in *individual* humans. I shall ignore this possibility in presenting the multiple realization argument, both for simplicity and because it has frequently been ignored in standard presentations of the argument as directed against type-type psychophysical identity theories.



computational processes that psychotechnically realize Martian mental states (since different functions get computed). Furthermore, even if (F.1) and (F.9) are both true of *humans* (and any other cognizing terrestrials, and Martians), it remains an open epistemic and conceptual possibility that (iv) there are physically possible creatures (Venusians?) in whom beliefs and other mental states are psychotechnically realized otherwise than via computational relations to sententially structured mental representations. In short, just as it would be chauvinistic to identify mental states with the neurobiological states that happen to *physically* realize them in humans, it would also be chauvinistic, given (i)–(iv), to identify mental states with the syntactic/computational states that, according to computational cognitive science, happen to *psychotechnically* realize them in humans.

With these observations as motivation, we can now set forth some plausible alternatives to Fodor's (F.2)–(F.8) and (F.10)–(F.13). Leaving (F.1) and (F.9) in place, we have the following theses:

- (A.2) Each mental state is psychotechnically realized by a complex monadic property, instantiable by a cognizer: the property of bearing a specific *relation* to a specific *mental representation*.<sup>4</sup>
- (A.3) Mental states and processes are psychotechnically realized by computational states and processes.
- (A.4.a) The states and processes that psychotechnically realize mental states and processes are symbolic; i.e., they involve mental representations with compositional syntax and semantics.
- (A.4.b) The states and processes that psychotechnically realize mental states and processes are formal; i.e., they apply to representations in terms of their formal/syntactic features, independently of their semantic content (if any).
- (A.5) Each mental state is psychotechnically realized by a specific formal/computational relation to a specific syntactically structured mental representation.<sup>5</sup>

---

<sup>4</sup> If desired, one could instead say that each mental state is psychotechnically realized by a syntactically structured representation itself, provided that this representation plays a suitable functional/computational role in the system. Little hangs on which way one puts it, as far as I can see; this is basically just a terminological matter. (If one did put (A.2) this other way, then various theses below would also get altered accordingly.) Talking about computational relations was important to Fodor, of course, since he wanted to type identify believing with some such relation, and he wanted mental representations to function as the "objects" of the attitudes. But on the realization view, all that is mistaken anyway.

<sup>5</sup> Michael Tye has pointed out to me that there are passages in Fodor's writings from the early and middle 1980's that may indicate a tendency on Fodor's part to advocate (A.2)–(A.5), as against (F.2)–(F.5). Here is an example:

Once the type-type identity theory embodied in (F.2)–(F.5) is thus replaced by a type-type realization theory, one can hold that the semantic properties of mental representations are explanatorily relevant to mental processes *qua* mental, even while admitting that the computational processes that realize these mental properties, *qua* computational, “have no access” to those semantic properties:

- (A.6.i) The semantic properties of mental representations are explanatorily irrelevant to the computational processes that psychotechnically realize mental states.
- (A.6.ii) The semantic properties of mental representations are explanatorily relevant to mental processes, *qua* mental.

Moreover, individuation of mental states by content need not coincide with syntactic/computational individuation of realization states. Thus, one can individuate mental states by *broad* content for purposes of scientific taxonomy and causal explanation, thereby allowing that distinct mental states of the kind arising in Twin Earth scenarios can be psychotechnically realized by the same syntactic/computational state:

---

[I]t may be *empirically* possible that there should be creatures that have the same propositional attitudes we do (e.g., the same beliefs) but *not* the same system of internal representations; creatures that, as it were, share our epistemic states but not our psychology. Suppose, for example, it turns out that Martians, or porpoises, believe what we do but have a very different sort of cost accounting. We might then want to say that there are translation relations among systems of internal representation (viz., that formally distinct representations can express the same proposition)...Whether we can actually make sense of this sort of view remains to be seen; we can barely think about the question prior to the elaboration of theories about how representational systems are to be semantically interpreted; and as things now stand, we haven't got semantic theories for natural languages, to say nothing of languages of thought (1981, p. 202).

But despite occasional such passages that might be construed as gesturing in the direction of a psychotechnic realization story, the dominant tendency toward a type identity story is reflected in remarks like these:

[T]he least hypothesis that is remotely plausible is that a mental state is (type) individuated by specifying a relation to a representation such that the subject bears the one to the other (1981, p. 226).

[T]he computational theory of the mind requires that two thoughts can be distinct only if they can be identified with relations to formally distinct representations. More generally: fix the subject and the relation, and then mental states can be (type) distinct only if the representations which constitute their objects are formally distinct (1981, p. 227).

The type identity thesis figures less prominently in Fodor's writings after Fodor (1987), however. And I don't know whether he would currently assent to (F.2)–(F.6), rather than to (A.2)–(A.6).

- (A.7) For purposes of scientific taxonomy, the mental states of humans are distinct in content from the mental states of their Twin Earth doppelgangers.<sup>6</sup>
- (A.8) Wide content has causal/explanatory relevance in psychological explanation.

Admittedly, theses (A.6ii)–(A.8) remain controversial even if one replaces (F.2)–(F.5) by (A.2)–(A.6.i). There are various philosophical arguments abroad, some due to Fodor himself (1987, chapter 2; 1991), that purport to establish (F.6.ii)–(F.8) on independent grounds, for instance on the basis of considerations involving “causal powers.” I myself maintain that these arguments are mistaken, and that wide-content mental states as such can and do figure in genuine causal explanations (Horgan 1989, 1991, forthcoming). But those who find such arguments convincing could mix and match, combining (A.2)–(A.5) with (F.6)–(F.8) And in any case, the main point to appreciate here is that computational cognitive science *per se* is not committed to (F.6)–(F.8). Rather, independent philosophical arguments would be needed to defend them.

What about the status of the PA’s, given (F.1) and (F.9) plus the alternative theses just set forth? Thesis (F.10), asserting that PA’s would be posited by an adequate cognitive science and hence are instantiated by humans, remains in place, being a direct consequence of these two basic theses. We now have the following alternative theses:

- (A.11) Each propositional attitude is psychotechnically realized by a complex monadic state consisting of a specific computational relation to a specific mental representation whose content is the content of the attitude itself.
- (A.12) For purposes of scientific taxonomy, the PA’s of humans are distinct from those of their Twin Earth doppelgangers.
- (A.13) Wide mental content has causal/explanatory relevance in PA explanations.

---

<sup>6</sup> Two interrelated points deserve mention concerning thesis (A.7). First, to accept (A.7) is to countenance not merely the multiple realizability of a single intentional mental property by distinct syntactic/computational properties, but also the realizability (in distinct environments) of two *distinct* intentional mental properties by a *single* syntactic/computational property. Second, to accept (A.7) is to drive a wedge between (i) a lower-level property that realizes a given intentional property (on a given occasion), and (ii) a lower-level supervenience base for that intentional property (on that occasion). In general, the supervenience base can be much wider, and much more inclusive, than the realizing property.

Again, although some philosophers have independent doubts about (A.12) and (A.13), it is crucial to appreciate that computational cognitive science *per se* is not committed to (F.12) and (F.13), over against the alternative theses (A.12) and (A.13).

So far in this section I have been adumbrating a non-Fodorian account of the metaphysics of the computational conception of the mind. Some philosophers and cognitive scientists, however, reject the computational conception; they would repudiate theses (F.1) and (F.9). It is worth noting that in doing so, they need not necessarily repudiate (F.10), the assertion that an adequate cognitive science would posit PA's. They could affirm that PA's are destined to be part of an adequate cognitive science, and then claim that a proper account of how PA's are psychotechnically realized would differ from computationalism's account. For instance, Colin McGinn (1989) advances the hypothesis that the psychotechnic realization of mental states generally, and of PA's in particular, involves mental representations that have *maplike* structure rather than *syntactic/sentential* structure. For another instance, John Tienson and I (Horgan and Tienson 1988, 1989, 1990, forthcoming) lately have been arguing that although an adequate cognitive science would posit PA's, and would posit psychotechnic realizations of PA's that involve syntactically structured mental representations, it would deny that PA's are psychotechnically realized via *computational* relations to mental representations. I.e., it would deny that cognitive processing conforms to programmable rules storable over mental representations themselves.

I turn now to Fodor's thesis (F.14). Some philosophers who are realists about PA's, myself included, do not agree that the fate of the PA's hangs upon whether or not they are destined to become part of mature science. The point can be put in terms of two potential kinds of psychotechnic realization. One kind would be *direct*: mature science would posit certain states which (i) are natural-kind states (within science itself), and (ii) themselves psychotechnically realize PA's. But another kind would be *indirect*: although PA's would indeed be psychotechnically realized by certain states countenanced by scientific theory, these realizing states would be quite baroque and complex, rather than being scientific natural kinds. Should it turn out that the correct scientific story about PA's involves highly indirect psychotechnic realizations, rather than reasonably direct ones, then so-called "folk psychology" could be true without becoming part of *scientific theory*.<sup>7</sup> Those who want to allow for this as a possible fate for the PA's will repudiate (F.14) in favor of the following alternative thesis:

---

<sup>7</sup> It is worth adding that even if folk psychology does end up being absorbed into science, the natural-kind states posited by scientific psychology still might be related to those of *physics* (or *neurobiology*) only via highly indirect realization, rather than being realized

- (A.14) The PA's are vindicable provided that they are psychotechnically realized, either directly or indirectly, by (possibly complex) states that would be posited by an adequate cognitive science or neuroscience.<sup>8</sup>

In order to appreciate the conceptual possibility of a scenario in which the PA's turn out to be indirectly psychotechnically realized and the generalizations of folk psychology turn out to be true without being absorbed into science, it helps to consider some set of systematically related, non-scientific, concepts central to everyday life—for instance, *clothing* concepts (cf. Cummins and Schwartz 1988, p. 49). The sortal properties posited by ordinary clothing talk (e.g., *being a hat*, *being a coat*, etc.) surely will not turn out to be scientific natural kinds; the body of mature scientific theories surely will not include a “science of clothing.” Yet items of clothing do exist nonetheless; they are *physical* objects, even though a raw physics-level description of a piece of wearing apparel would normally be enormously complex. In addition, linguistic competence over clothing terminology includes the (largely tacit) mastery of a host of reasonably systematic *ceteris paribus* generalizations about clothing. (Hats are normally worn on the head, for purposes such as warmth, shading from direct sunlight, or style. Coats are normally worn about the torso and arms, for purposes such as warmth, protection from inclement weather, or style. Etc.) And evidently one can give perfectly legitimate causal explanations, adverting to clothing *qua* clothing, that are largely (and often implicitly) based upon such generalizations. (For instance, I explain why Jane stayed warm outdoors today, even though Dick got the shivers, by pointing out that she was wearing a heavy coat and he wasn't.) The concepts of common sense belief/desire psychology might turn out to be analogous to clothing concepts, in all the respects just mentioned.

I turn, finally, to the Fodorian theses (F.15)–(F.20), concerning the putative implications of a metaphysical position that is realist about intentionality but denies that aboutness is destined to make the physicists' final catalogue of the ultimate and irreducible properties of things. The passage from

---

by physical (or neurobiological) natural-kind-states. I.e., the realization relation might be indirect not only for non-scientific kinds and properties *vis-à-vis* the scientific kinds and properties that realize them, but also for higher-level scientific kinds and properties *vis-à-vis* the lower-level kinds that realize them.

<sup>8</sup> The point of saying ‘cognitive science or neuroscience’ is to allow for the potential person who thinks that PA's are real but not destined to be part of mature science, and who also thinks that mature science will not include any discipline of “cognitive science” at all. For such a person, psychotechnic architecture would be just neurobiological architecture, rather than something characterized at a more abstract theoretical level.

Fodor I quoted shortly before setting out these five theses suggests that the following is an exhaustive pair of alternatives: either (i) there are tractably specifiable nonsemantic, nonintentional, sufficient conditions for intentional states; or else (ii) intentionality is among the fundamental and unexplainable properties of things. But there is a third possibility: viz., that although the intentional is supervenient upon the nonintentional, in general there is no way to *tractably specify* the nonsemantic, nonintentional, conditions that suffice for intentional phenomena. Although a physical supervenience base might always exist for any manifestation of aboutness, in general any adequate nonintentional, nonsemantic characterization of the supervenience base might be enormously baroque and complex. Perhaps, for instance, the supervenience base for the intentional content of a token thought (or token utterance, or token inscription) generally involves a good-sized chunk of spacetime extending well beyond the cognizer's own body and well beyond the time at which the token thought occurs; perhaps it involves a rather gargantuan number of physico-chemical goings-on within that extended spatio-temporal region; and perhaps there isn't any simple way to describe, in nonintentional and nonsemantic vocabulary, all the *relevant* aspects of this hugely complex supervenience base.<sup>9</sup> Perhaps, in addition, the supervenience of the intentional on the nonintentional is largely a holistic matter—with the intentionality of thoughts, utterances, and inscriptions supervening not individually (one token at a time), but rather collectively, as part of the correct global intentional interpretation of a cognizer—or perhaps of the cognizer's whole community or whole species.

In short, it might be that the search for tractably specifiable, cognitively surveyable, nonintentional and nonsemantic sufficient conditions for intentionality is utterly hopeless—and yet that the intentional supervenes upon the nonintentional nonetheless. Once one appreciates this possibility, one realizes that thesis (F.15) is really quite tendentious (and quite optimistic). Here is the weaker alternative:

---

<sup>9</sup> Moreover, the supervenience base for an intentional property of a specific token thought might be not merely enormously complex, but also highly *specific*. Perhaps there are no *general* sufficient conditions for being a thought that p, for arbitrary content p—not even complex and baroque ones. It could be that whenever an intentional mental property is instantiated, the supervenience base for that property (on that occasion of instantiation) is significantly different from the supervenience base on any other occasion when an intentional mental property is instantiated—so different that the best that could be done, by way of a *general* accounting of supervenience bases for intentional mental properties, would be an unsystematizable *list*. Each item on the list would consist of a specific supervenience base (instantiated on a particular occasion), together with a specific intentional mental property (also instantiated on that occasion).

(A.15) If there is a place for intentional categories in a physicalistic view of the world, and if a physicalistic view of the world is correct, then the intentional is supervenient upon the nonintentional.

One could subscribe to (A.15), and also to its antecedent clauses—(F.16) and (F.17)—and yet deny that there are tractably specifiable sufficient conditions for intentionality. I.e., one could also assert

(A.18.a) The intentional is supervenient upon the nonintentional.

(A.18.b) It is not possible to tractably specify nonintentional and non-semantic sufficient conditions for a physical system to have intentional states.

Accordingly, one could grant all of (F.16), (F.17), and (F.19) and yet still reject (F.20) in favor of

(A.20) It is not possible to tractably specify nonintentional and non-semantic sufficient conditions for a primitive Mentalese symbol to have a certain interpretation in a certain context.

(Could we paraphrase thesis (A.18.b) as “The intentional cannot be naturalized”? Perhaps, but I think an advocate of the thesis does better to resist conceding the term ‘naturalize’ to the opposition, and to deny that a naturalistic conception of intentionality requires the existence of tractably specifiable sufficient conditions.)

Having reviewed Fodor’s own views about computation, mental representation, and belief, and having surveyed some alternative options on the landscape of potentially viable philosophical positions, we are ready now to consider respectively the positions adopted by Baker, Garfield, and Cummins.

**3. Baker.** The fundamental question motivating Baker’s book is this: “What are the relations between emerging scientific concepts of the mind and the familiar, everyday concepts in terms of which we see ourselves and others as acting from beliefs, desires, and intentions?” (p. 3) The dominant philosophical approach to this question, she maintains, is characterized by a thoroughgoing commitment to a physicalism. (More below on how she construes this metaphysical position.) Her book is both a critique of physicalism and a defense of everyday intentional concepts.

Physicalism, as she understands it, entails that common-sense intentional psychology is correct only if it will turn out to be vindicable by mature science, and in particular by a kind of scientific psychology (“physicalistic

psychology”) that itself meets certain physicalist strictures. She situates her own philosophical position, and the others against she will argue, by characterizing them relative to the following “argument from physicalism,” which purports to establish that the common-sense conception of the mental is bankrupt:

- (1) Either physicalistic psychology will vindicate (in a sense to be specified) the common-sense conception of the mental, or the common-sense conception is radically mistaken.
- (2) Physicalistic psychology will fail to vindicate (in the relevant sense) the common-sense conception of the mental.

Therefore,

- (3) The common-sense conception of the mental is radically mistaken.

“The first premise,” she says, “is simply a statement of physicalism” (p. 6). It is endorsed by physicalists like Fodor who are realists about belief, and also by those physicalists (viz., eliminativists) who think that the common sense conception of the mental is radically false. “The strongest sort of vindication of the common-sense conception,” she says, “would come from a theory whose generalizations apply to mental states by virtue of their contents; a minimal vindication would result even from a theory that held that...common-sense concepts...are extensionally equivalent to correct ones” (p. 6). The second premise, on the other hand, is endorsed by the eliminativists but repudiated by the Fodorian realists. Her own position is a non-physicalist form of realism: she accepts premise (2) but rejects premise (1). In the first half of the book she argues that the common-sense conception cannot be vindicated physicalistically; in the second half, she argues that belief is legitimate anyway because it is an indispensable feature of our conceptual scheme. I will discuss each half in turn.

3.i. Before considering Baker’s argument *against* physicalism, let us first consider her discussion of the position itself. She characterizes physicalism as having two components: a claim about science and particular conception of science.

The claim about science is that science is the exclusive arbiter of reality. This scientific realism is captured nicely by Wilfrid Sellars, who transforms the aphorism attributed to Protagoras to fit the current intellectual temper: “in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not.” On this view, scientific knowledge is exhaustive.

The particular conception of science embedded in physicalism is that physically indistinguishable individuals with physically indistinguishable histories are to be assigned the same states.... I shall use ‘physicalistic psychology’ to speak not of any particular psychological the-



ory but of this physicalistic interpretation of psychology—an interpretation overwhelmingly endorsed, explicitly or implicitly, by philosophers concerned with cognitive science. (p. 4)

I will call the two components of this position the *scientific hegemony* claim and the *supervenience* claim, respectively. Physicalism, she says, has both individualistic and nonindividualistic versions. “Roughly, if psychological states are specified without presupposing anything about the character of the external environment, then the physicalism is individualistic; otherwise it is not individualistic” (p. 5).

She takes it to be a corollary of the supervenience claim that “a physicalistic interpretation of psychology aims to provide nonintentional and nonsemantic sufficient conditions for psychological states, whether such conditions are construed individualistically or not” (p. 5). Thus, she takes physicalism’s supervenience claim to entail (F.18), the thesis that that there are *tractable* sufficient conditions, formulable in nonintentional and nonsemantic vocabulary, for intentional states. Like Fodor himself, and like numerous other philosophers, she fails to notice that (F.18) is actually a much stronger thesis than (A.18.a). She fails to notice that the supervenience thesis, as characterized in the above passage, actually entails only (A.18.a), not (F.18)—and hence that one can consistently assert both (A.18.a) and (A.18.b).

She uses the rubric ‘physicalistic psychology’ in such a way that anything falling under it would be both (i) a *theory*, and (ii) a *scientific* theory. Thus, in asserting that premise (1) of the argument from physicalism is simply a statement of physicalism, she evidently is construing physicalism’s scientific hegemony claim as entailing (F.14), the thesis that belief/desire psychology is vindicable only if it is destined to become a part of science. And certainly Fodor himself, and many other physicalists too—in particular, eliminativists like Stephen Stich, Patricia Churchland, and Paul Churchland—do indeed seem to regard (F.14) as a corollary of their own physicalistic positions. It should be noted, however, that the vague and pretheoretic Sellarsian idea that “science is the measure of all things” is susceptible to a range of alternative potential vagueness-resolutions. In particular, a philosopher who espouses (A.14) rather than (F.14) could still claim to be an advocate of a broadly physicalistic ontology. For, the “higher level” properties he countenances still would have to be physically realized by physical entities—even though these properties themselves would not necessarily be scientific natural kinds. (Think again, for instance, of the sortal properties that collectively constitute our common sense typology of clothing.)

In short, it is possible to espouse physicalism’s supervenience claim, and also to espouse a version of its hegemony claim, without committing one-

self to either (F.14) or (F.15); one would instead espouse (A.14) and (A.15). I.e., it is possible to *be* a physicalist, under Baker's own characterization of physicalism, without espousing either (F.14) or (F.15).

Her overall argument against physicalism goes as follows. According to individualistic versions of physicalistic psychology, attitudes would supervene on the creature's current intrinsic physical state. According to nonindividualistic versions, they would generally supervene on something broader: the creature's intrinsic physical state together with certain physical aspects of the creature's history and/or environment. Individualistic versions cannot succeed, at least insofar as we are attending to attitudes as ordinarily individuated (by ordinary *that*-clauses); for, that kind of individuation is incompatible with supervenience on the creature's intrinsic physical state. Nonindividualistic versions almost certainly cannot succeed either. For, Fodor's recent efforts (circa the mid-1980's) to give a satisfactory account of 'narrow content', and to construe 'wide content' as involving a function from contexts to truth conditions, are unsuccessful; Dretske's efforts to construe belief as reliable indication (Dretske 1981, 1985) are unsuccessful; and the problems these approaches encounter make it very unlikely that there are tractably specifiable, nonintentional and nonsemantic, sufficient conditions for intentionality.

The details of this overall argument involve various Twin Worldish thought experiments, plus a variety of considerations pointing to the likely hopelessness of trying to tractably specify, nonsemantically and nonintentionally, those aspects of a creature's environment and/or history that contribute to the wide content of its mental states.

She takes herself as having refuted, in the course of articulating this overall line of reasoning, certain more specific theses and doctrines in recent philosophy of mind, ones which are sometimes taken to undergird physicalistic psychology. These include (i) narrow functionalism, of the input-output variety; (ii) the general conception of mind as computer; (iii) computational cognitive science; and (iv) the notion of narrow content in philosophy of mind.

But although her argument does raise serious doubts about *individualistic* physicalism, it does not really call into question the nonindividualistic version. At most what it shows, concerning nonindividualistic physicalism, is that there probably do not exist *tractably specifiable* nonindividualistic sufficient conditions for mental states. This leaves open the possibility that nonindividualistic physical conditions always exist, even though they generally cannot be tractably specified. I.e., although the argument does make trouble for for thesis (F.18), it evidently leaves (A.18.a) unscathed.

What about the other negative implications she claims for her argument? The part of the argument directed at individualistic physicalism does, I

think, tell against narrow functionalism; but so does much recent work in philosophy of mind, in the Twin Earth tradition of Putnam and Burge.

Regarding the conception of mind as computer, and work in cognitive science that presupposes this view, she essentially accepts as untendentious Fodor's position that the computational conception of mentality is committed to all of theses (F.1)–(F.13) and thus to the narrow type-individuation of intentional mental states. On that basis, she takes her argument against individualistic physicalism to show that the mind is not correctly describable as a computer, thereby undermining the foundations of computational cognitive science. But the most the argument really shows, however, is that *if* theses (F.1)–(F.13) are an adequate philosophical reconstruction of the computational conception of mind (and of the foundations of computational cognitive science), *then* this conception of mentality is untenable, and so is the computational paradigm in cognitive science. But if one adopts (A.2)–(A.8), (A.12), and (A.13) in place of the corresponding Fodorian theses, then computationalism emerges essentially unscathed. (I again remind the reader that (A.2)–(A.8) could be adopted even without adopting all the remaining theses in section 2.)

Regarding the notion of narrow content, and the view that wide content is jointly determined by narrow content plus context, there are two things to say. First, the specific account of narrow content that Fodor was advocating in the mid-1980's does come in for a rather persuasive attack, pursuant to her overall argument. But second, the considerations she puts forth against the possibility of any viable notion of narrow content all presuppose that such a notion, in order to serve any useful philosophical purpose, would have to be specified by way of tractable sufficient conditions that are nonintentional and nonsemantic. To the extent that narrow content might have some legitimate theoretical role to play even in the absence of tractable sufficiency conditions, her general negative argument is inapplicable.

3.ii. In the second half of the book, Baker argues that belief and the other attitudes "are not dispensable for either everyday or scientific purposes" (p. 12). The core idea is that denial of the common-sense conception of the mental is pragmatically incoherent.

*[I]n the absence of a replacement, it is literally inconceivable that the common-sense conception is false.... I shall set out several ways in which denial of the common-sense conception may be self-defeating or otherwise pragmatically incoherent. If the thesis denying the common-sense conception is true, then the concepts of rational acceptability, of assertion, of cognitive error, even of truth and falsity are called into question. It remains to be seen whether or not such concepts (or suitable successors) can be reconstructed without presupposing the truth of attributions of content (p. 134, my italics).*

She proceeds to effectively lay bare how thoroughly the common-sense conception is woven into the fabric of our overall conceptual scheme, and consequently how extensive would be the pragmatic contradictions attendant to the denial of belief. To my mind, she makes a very convincing case.

Other philosophers too have charged eliminativism with pragmatic inconsistency, but Baker does more than merely repeat the charge. An important moral of her discussion is that the common-sense conception does substantially more work in our conceptual scheme than providing theoretical or quasi-theoretical explanations of behavior alongside the other potentially available kinds of explanation (e.g., neuroscientific, or physico-chemical). For instance, the *entire scientific enterprise*, irrespective of which kinds of explanation are involved, presupposes notions like rational acceptability, assertion, and truth. Thus this entire enterprise, as we ordinarily conceive it, presupposes the common-conception of the mental.

Although it is possible to read Baker as propounding a full-fledged transcendental argument for the reality of beliefs, I think a closer look at the text supports a different interpretation. As the italicized parts of the above-quoted passage make fairly clear, she is prepared to acknowledge the epistemic *possibility* that our current notions of rational acceptance, assertion, truth, and so forth could be replaced by successor concepts that do not presuppose belief. The question is, how *good* is this epistemic possibility, given current evidence? The first sentence of the passage speaks to this question: since nobody, including any prominent eliminativist, has yet proposed even the barest sketch of what such successor concepts might be like, and since we currently lack any conception at all of what they might be like, there is presently only a *bare* epistemic possibility that such successor concepts could arise. Accordingly, the current *empirical* evidence for the indispensability of our common-sense conception of the mental is overwhelmingly strong, and the eliminativists bear a crushingly heavy burden of proof that they have not even begun to discharge. (I emphasize this empirical, nontranscendental, reading of her argument because it is all too easy to refute the full-fledged transcendental version: one merely points out the epistemic possibility of non-belief-tinged successor concepts, without having to worry about how good this possibility might be or what these concepts might be like.)

Even if humans cannot make do without the common-sense conception of the mental, however, it remains a further question whether belief attributions are ever *true*, and whether humans actually ever *instantiate* such states. Couldn't the common-sense conception be both indispensable and yet radically false? Baker does not address this question in its general form. She does spend a chapter developing a critique of Dennett's version of instrumentalism, and I think she gives Dennett a good raking over the coals. However, she

never directly addresses the question of whether, and why, the apparent indispensability of the common-sense conception constitutes evidence for realism about belief.

I myself think it does, because of what it suggests about the concept of belief itself and hence about the satisfaction conditions for belief attributions. The indispensability of the concept of belief constitutes powerful empirical evidence in favor of a fairly austere conception of the purely conceptual commitments of our concept of belief, and a correspondingly austere conception of the semantics of belief attributions (Horgan and Graham 1991). Under this austere conception, rather little is conceptually required for being a true believer, over and above satisfying the criteria that are ordinarily taken as behavioral evidence for belief attributions; consequently, there is no serious doubt that humans really have beliefs. Although it is an open empirical question *how* beliefs are psychotechnically realized in humans, there is no serious doubt that they are *somehow* realized. It might even be that belief properties will turn out to be realized by quite complex combinations of state-types that would be posited by a mature cognitive science, rather than being directly absorbed into science themselves.

**4. Garfield.** Garfield announces three principal, interrelated, aims of his investigation. First, he seeks to understand “the ontology of computational psychology, and the ontology of mind that would be entailed by its success” (p. 6). Second, he is concerned with “the nature of the connection between the ontology of science and the ontology of what Sellars...called the ‘manifest image of man-in-the-world’” (p. 6). And third, he seeks to make some progress toward solving “ontological problems in the philosophy of mind—specifically, to get an account of the ontological supervenience base...of psychological phenomena” (p. 7).

The global argument of the book runs as follows. Propositional attitudes, he maintains, pose a *prima facie* problem for computational cognitive science—a problem deriving from their relational and intentional character, together with the plausibility of the *autonomy principle* asserting that “the states and processes that ought to be of concern to a psychologist are those that supervene on the current, internal, physical state of the organism” (Stich 1983, p. 164). Garfield describes the problem this way:

The task of cognitive psychology is to characterize the structure of the human information processing system so as to provide an explanation of how we acquire, store, manipulate, and transform information, and of how we employ that information in guiding and producing behavior. It is, as such—*prima facie*, though perhaps not ultimately—a science of the internal workings of the individual organism. A way in which this point is frequently made involves what Stich calls the “autonomy principle”.... [T]he central problem of this investigation [is that] [i]t seems impossible that anything like the PA’s as construed by the manifest image—that is, as individuated as contentful—can meet the conditions for psychological phenomena, because no

phenomena individuated by content can satisfy the autonomy principle and no phenomenon that violate the autonomy principle are suitable objects for psychology. Hence, it appears either that psychology must give up on explaining PA's or that psychology cannot be a science. (pp. 56–58)

There are three broad philosophical strategies for surmounting these problems: (i) the reconciliationist strategy, which seeks to establish the compatibility of PA psychology with computationalism; (ii) the eliminativist strategy, which seeks to jettison PA psychology as incompatible with mature science in general and computationalism in particular; and (iii) the “binocular” strategy, which denies that the scientific image of the world and the manifest image of man-in-the-world (as Sellars called them) can be synthesized, and yet retains them both anyway (with PA's figuring as a central component of the manifest image). The reconciliationist strategy cannot work, he argues, because the prevailing computational theory of the mind is committed to an individualistic theory of meaning for mental states, and such a theory is untenable. Nor can the eliminativist strategy work, mainly because of pragmatic self-stultification problems of the kind also stressed by Baker. “[I]f the strategy were true, assertions of its truth would have no content” (p. 8). So the only viable strategy is the binocular one.

The final substantive chapter offers a positive proposal, supposedly falling under the binocular strategy, concerning the nature of the PA's, their place in psychology, and the implications of this view for the nature of psychology and the relationship between scientific psychology and the manifest image of man-in-the-world.

[T]he manifest image places constraints on the domain of psychology through its preoccupation (shared by other social and behavioral sciences) with phenomena constituted by persons embedded within cultures that construct the manifest image, and...these constraints issue in the incoherence of any account of intentional states that either denies their existence or does serious violence to their relational, intentional, character and to their kinship with linguistic states. (p. 8)

One important flaw in Garfield's global argument is its conflation of two kinds of reconciliationism: (i) a generic version, which would reconcile the manifest image in general (and PA's in particular) with the body of mature scientific theories; and (ii) a specific version, which would reconcile the manifest image (and the PA's) with *computational cognitive science*. If indeed the reality of PA's is incompatible with computationalism, then why isn't the appropriate conclusion—a conclusion in the spirit of generic reconciliationism—that computationalism is *mistaken* and hence is not destined to be a part of mature science? Such a solution to the putative incompatibility problem would surely be much less radical, and much less fraught with profound internal conceptual tensions, than opting for binocularism. Yet this

obvious-looking solution doesn't even make it onto Garfield's menu of options, because of the conflation just mentioned.<sup>10</sup>

A second, related, problem concerns his own positive proposal concerning the PA's *vis-à-vis* scientific psychology. As far as I can tell, the proposal really does not fall under the rubric of binocularism anyway. Instead, he stresses certain *constraints* that the manifest image supposedly places upon any acceptable scientific psychology, and which allegedly are violated by computationalism—constraints involving the incorporation of relational, environmental/historical, factors that allegedly don't supervene on the current, intrinsic, physical state of the organism. These look for all the world like *compatibility* constraints, the kind which (if respected) would subserve a reconciliationist perspective rather than a binocular one. Garfield thus seems to misconstrue the true structure and the overall import of his own global argument: he is evidently a closet reconciliationist, arguing against computationalism on the grounds of its alleged incompatibility with realism about PA's.

Garfield argues plausibly that PA state types, as ordinarily individuated, do not in general supervene upon the current, intrinsic, physical state of the organism (a point also argued by Baker, and widely accepted in the current era of Twin Earth scenarios). More problematic, however, is his contention that computationalism is committed to a narrow theory of meaning asserting that the content of mental states *does* supervene on the organism's current, intrinsic, physical state.

For Garfield, this contention is not (as it is for Baker) just a matter of essentially accepting as untendentious that computational cognitive science is committed to all of these (F.1)–(F.13). On the contrary, Garfield explicitly considers an alternative construal of the foundations of computationalism, inspired by the writings of Zenon Pylyshyn, under which computationalism seems *not* to be committed to a narrow theory of meaning. Garfield dubs the view “Naturalistic Individualism”; it is naturalistic in the sense that it allows a place for organism-environment interrelations as an aspect of content, and yet it is individualistic with respect to the nature of the token mental states that *have* content.

It may be somewhat unfair or inaccurate to attribute the Naturalistic Individualism I will characterize to Pylyshyn. His own statements of his position (Pylyshyn, 1984) are somewhat vague, and at times he seems to believe...that he is a Fodorian methodological solipsist. Nonetheless...the Naturalistic Individualism I will characterize is, I would argue, the dominant metatheoretic view of psychologists working in the cognitive paradigm.... In any real ex-

---

<sup>10</sup> Garfield also is assuming, in effect, that the PA's would need to be absorbed into cognitive science—directly realized by states posited by such a science—in order to be compatible with it. That too could be questioned; cf. my discussion in section 2 of thesis (A.14) as against (F.14).

planation of behavior...., Pylyshyn's argument goes, the content of the belief as determined by its causal semantic connections to [the organism's] environment plays a role.... The conclusion is that in cognitive theory, internal information-processing states have to be identified by their content...and that in order to do this one must examine their connections not only to other cognitive states and processes but also to the organism's distal environment (this is what makes Naturalistic Individualism naturalistic).... [H]owever...this strategy is thoroughly individualistic in its account of the nature of the states that "get interpreted." They are construed as purely internal. (pp. 64–66)

On the face of it, this is a construal of the foundations of computationalism that essentially adopts theses (A.2)–(A.8), (A.12), and (A.13), as against the corresponding Fodorian theses. It does not construe the computational state types and the intentional state types posited by cognitive science as *identical*, but instead construes the former as *psychotechnonic realizations* of the latter. It is individualistic only in the sense that it identifies *token* PA's with token computational states within the organism. If the view is correct, then presumably Stich's autonomy principle is just mistaken; the state *types* posited by computational cognitive science need not necessarily supervene on the current, intrinsic, physical states of the organism at all. On the face of it, then, this position can thoroughly reconcile computational cognitive science with the contention that PA's cannot be narrowly type-individuated.

But although Garfield is aware of Naturalistic Individualism as an alternative to Fodor's conception of the foundations of computationalism, he argues that Naturalistic Individualism is ultimately incoherent, since it is really committed to a narrow theory of meaning in spite of itself. The problem, he claims, stems from its commitment to the identity of token intentional states with token computational states:

The most important feature of an Individualistic Theory of Meaning is its commitment to a narrow supervenience base for all psychological phenomena.... An Individualistic Theory of Meaning is committed to the view that no two individuals could differ psychologically without differing in some individualistic physical respect as well. *Put another way*, according to an Individualistic Theory of Meaning, psychological events, states, and processes are supervenient upon individualistic states, events, and processes.... [I]f the PA's are to be counted among the phenomena posited on a realistic interpretation of a psychological theory, then they must be identified with particular tokens of physically—i.e., individualistically—characterized states of the organism.... But, as the Individualistic Theory of Meaning was characterized..., this commitment to an individualistic identification of token PAs with token physical states just is a commitment to the truth of that theory of the ontological status of meaningful phenomena. (pp. 90–92, emphasis mine)

Unfortunately, this argument is quite thoroughly confused. A token state internal to an organism can perfectly well be (identical to) a token belief, even if the belief *type* that is tokened does not supervene upon the organism's current, intrinsic, physical properties. (Similarly, a token sound-sequence emitted by an organism can perfectly well be identical to a token



assertion that snow is white, even though the property *having the content that snow is white* does not supervene upon the current, intrinsic, physical properties of the sound sequence or the organism.) Put another way, a token internal state *s* of an organism *O* can instantiate an intentional mental property *M* even if the supervenience base for *s*'s being an *M*-instantiation is vast and wide, and includes numerous phenomena that are not themselves internal to *O* at all. (Similarly, the supervenience base for a token sound sequence's having the content that snow is white is normally vast and wide, and includes much that is not intrinsic to the sound sequence itself—viz., the full range of phenomena in virtue of which the organism emitting the sequence counts as a member of a certain linguistic community, and in virtue of which numerous interrelated sound-sequence types have their specific contents within this community.)

Inspection of the wording in the above quoted passage suggests that Garfield's confusion on this point hinges largely on his all-too-casual talk, in the third sentence of the passage, of supervenience relations among "events, states, and processes." In context, it appears that he here allows himself to think of *token* events, states, and processes as being related by the supervenience relation. Evidently he construes this putative token-level supervenience relation as a necessary condition for the *identity* of token mental phenomena with token physical phenomena; and he thinks that a token mental phenomenon cannot token-supervene upon a token physical phenomenon unless the corresponding mental *type* supervenes upon the corresponding physical type. But this way of thinking about supervenience is just mistaken: in general, the supervenience base for a token phenomenon's intentional properties can be much broader than the supervenience base for its intrinsic, nonintentional, properties.<sup>11</sup> (Think again about token linguistic utterances and inscriptions.)

---

<sup>11</sup> This point holds even if one construes token events and states not as pure particulars, but instead as structured entities consisting of an object's exemplifying a property at a time (Kim 1966, 1973, 1976; Goldman 1970)—and even if one construes each token mental event/state as an organism's exemplifying a *mental* property at a time (Kim 1966), and each token action as an organism's exemplifying an *act type* at a time (Goldman 1970). On such a view, token events are individuated in a "fine-grained" way, and token identity claims are only true when the corresponding type identity claims are also true. Hence if two token events *e* and *f*, with the same respective constituent object and constituent time, are such that *e*'s constituent property realizes *f*'s constituent property, then *e* thereby realizes *f*. I.e., there is a (derivative) realization relation among token events, induced by realization among event types. But the key point is this: under the property-exemplification view, one token event or state can realize another even if the supervenience base for the higher-level constituent property is considerably broader than the supervenience base for the lower level constituent property. For instance, my token act of signaling a left turn at noon is realized by my leftwardly extending my left arm at noon—even though the physical supervenience base for my exemplifying the social/conventional property *signaling a left turn* extends spatio-temporally well beyond my own body at the time of the action,

The same deep-seated confusion suffuses Garfield's own proposed account, late in the book, of the ontology of the PA's. Making use of Sellars' device of dot quotation, he describes his view this way:

[T]he PAs are on this view relations of their bearers to linguistic types individuated by their content... [T]here is a striking parallel between the PA verbs and the verbs of saying, for to say that most Australian snakes are poisonous is to produce a public token of a •Most Australian snakes are poisonous•, whereas to believe that most Australian snakes are poisonous is to produce, or be disposed to produce, an internal token of a •Most Australian snakes are poisonous•. PAs...are, from the standpoint of ontology, states that...supervene on a motley plethora of relations their bearers stand in with respect to their environments. (pp. 134–35)

One would think, reading these remarks, that the view in question is much like Pylyshyn's own token identity theory concerning beliefs; indeed, Garfield seems to be saying that an *occurrent* token belief that p is identical to a token internal state that instantiates an intentional state type with the content that p. But Garfield goes on:

My account diverges from Pylyshyn's...with respect to what it is to represent the state of affairs that the building is on fire. Whereas for Pylyshyn that is an individualistic fact about Mary, for me it is a relational fact; whereas for Pylyshyn some computational state of Mary's central nervous system is to be identified with the belief in question, for me her belief consists in her psychological state's being a •The building is on fire•, a state with an indefinitely broad supervenience base. (p. 151)

He evidently takes the breadth of the supervenience base undergirding the token state's being a •The building is on fire• as grounds for concluding that the token state *itself* must be a "broad" phenomenon—too broad to be some internal state of Mary. Conversely, he evidently thinks that since the token state Pylyshyn has in mind is indeed some internal state of Mary, Pylyshyn is committed to saying that the supervenience base for this token state's being a •The building on fire• is *itself* "an individualistic fact about Mary." The underlying mistake, made twice in this passage, is the same one as before: viz., supposing that if a token state internal to an organism has some intentional property, then the supervenience base for the state's having that property must itself be some phenomenon (or collection of phenomena) that is wholly internal to the organism.

Clear away this confusion, and it appears that Garfield's ontological approach late in the book is probably best reconstructed as a version of the view he attributes to Pylyshyn.<sup>12</sup> Thus, there is ultimately a double irony in

---

whereas the physical supervenience base for my exemplifying the arm-extending property is intrinsic to my own body at noon. Likewise, *mutatis mutandis*, for the realization of token mental events and states by token physical events and states.

<sup>12</sup> By this I mean the view attributed in the above quotation from pp. 64–66, where he acknowledges that for Pylyshyn, content doesn't supervene on the intrinsic. This does not

the discrepancy between the actual import of his discussion and his own understanding of its import. In the end, and after the confusion is straightened out, the discussion is best viewed not only as pointing toward *generic* reconciliationism over against binocularism (and hence toward the repudiation of computationalism, should that approach to scientific psychology prove incompatible with realism about PAs); it is also best viewed as pointing toward the specific form of reconciliationism asserting the compatibility of computational cognitive science with the conception of PA's embodied in the manifest image of man-in-the-world.

Moreover, his discussion late in the book, arguing that various current theories and models in cognitive psychology actually conform well with his own recommended conception of the foundations of scientific psychology, only reinforces the impression that his own positive position is best reconstructed in a way that renders PA realism directly compatible with computationalism. For, the theories and models he discusses are situated well within the computationalist paradigm. In arguing that these theories need not be tethered to a narrow theory of meaning for the mental states they posit, he himself—in spite of himself—further underscores the case against saying that computationalism is incompatible with PA realism.

##### 5. Cummins. Cummins's book is directed at what he calls The Problem of Representation, which he characterizes this way:

[T]he Problem of Representation...is, at least as I understand it, a paradigmatic problem in the philosophy of science. To a large extent, empirical theories of cognition can and do take a notion of mental content as an explanatory primitive. But this is a kind of explanatory loan (Dennett, 1978): If it turns out that the notion of mental representation cannot be given a satisfactory explication—if, in particular, no account of the nature of the (mental) representation relation can be given that is consistent with the empirical theory that assumes it—then, at least in this respect, that theory must be regarded as ill founded, and hence as a less than adequate response to the drive for the kind of thorough intellectual understanding that motivates the scientific theory in the first place. (pp. 1–2)

---

include the Individualistic Theory of Meaning, which Garfield wrongly thinks is entailed by Pylyshyn's token physicalism.

Another possible reconstruction of Garfield's position, in view of the above quotation from p. 151, is that he is identifying a token belief not with a token representation which *is* a •The building is on fire•, but rather with a token state consisting of that token representation's *being* a •The building is on fire•. This construal better respects the last sentence of the quotation. But once his underlying confusion is cleared away, I don't see why he *should* adopt the view he evidently espouses in that final sentence. It is more natural to adopt the Pylyshyn view—just as it is more natural to identify a token utterance with a token sound sequence which has the content that p, rather than construing a token utterance as a (putative) entity consisting of the token sound sequence's *having* the content that p.

Cummins does not say explicitly what would count, by his lights, as a “satisfactory explication” of mental representation. But presumably he has in mind something like tractable conditions, formulable in nonintentional and nonsemantic vocabulary, that are either *necessary and sufficient* for something’s being a mental representation with a specific representational content, or anyway are at least *sufficient*. I.e., presumably the problem he is addressing is essentially the same one that Fodor thinks must have a solution since aboutness will not figure in the physicists’ final catalogue of the ultimate and irreducible properties of things, and that Baker thinks almost certainly does *not* have a solution.

His principal concern is the notion of mental representation employed in computational theories of cognition. Thus he seeks an account, within the context of such theories, of what it is for a state of system to have a representational content, and of what determines the specific content it has. He holds that the idea of mental representation plays somewhat different explanatory roles in ordinary belief/desire intentional psychology, in orthodox computational cognitive science, in connectionist cognitive science, and in neuroscience. So we should not assume, he cautions, that the relevant notion of mental representation for the computational theory of cognition (the CTC) is the same one required by ordinary belief/desire psychology.

Nor should we assume that a mature computational account of cognition would traffic directly in states that play the role of the propositional attitudes of common sense, and that involve computational relations to representations whose contents are those of the attitudes themselves. “We need to keep open the possibility that, e.g., belief attribution, though a legitimate case of semantic characterization, is not a semantic characterization of any representation in the believer” (p. 15). (In the terminology I introduced in section 2, Cummins is evidently broaching the possibility that propositional attitude state-types will be realized not *directly* by the natural kinds posited by computational cognitive science, but only *indirectly* via states whose complete description at the level of cognitive science might be quite baroque and complex.)

He discusses four kinds of answer that have been proposed to the problem of mental representation, both as this problem was posed in the history of Modern philosophy and as it is posed nowadays: answers that appeal respectively to similarity, to covariance, to adaptational role, and to functional role.

After a brief discussion of why it became clear in the 17th century that similarity theories, appealing to the alleged resemblance between mental representations and the things they represent, are not tenable, he turns to Locke’s version of the theory that the essence of mental representation is covariance between the occurrence of the mental representation and the pres-

ence of the thing represented. He argues that Locke's account faces insurmountable problems, and that these same problems also arise for contemporary covariance accounts—notably Fodor's and Dretske's. The most fundamental difficulty is that such theories cannot satisfactorily accommodate the phenomenon of *mis*-representation. He maintains that the standard way of trying to handle misrepresentation—viz., via appeal to covariation in an *idealized* cognitive system—is not tenable. Among other reasons, idealizing away error must inevitably involve envisioning a cognitive system whose psychotechnic design is fundamentally different from that of systems describable by the CTC. "Error," he argues, "is the inevitable price of computational tractability" (p. 54).

Adaptational role accounts of mental representation, such as Millikan's, get dismissed rather quickly as not applicable to the notion of representation assumed by the CTC. For, these accounts are committed to a "historical" notion of representation—a notion under which a state's being a representation with a specific content depends upon its ontogenetic history, in the individual organism and/or the organism's species. But the CTC, Cummins claims, assumes an ahistorical notion of representation.

He then proposes an account of mental representation he calls "interpretational semantics." Although this position is ontologically equivalent to a kind of functional role semantics, he says, it is not really an alternative to traditional theories of mental representation, because it is specifically tailored to the CTC's version of this notion; it is *not* an account of the kind of representation presupposed by common sense belief/desire psychology.

The account goes as follows. For a physical system that counts as a computational system, such as an adding machine or a digital computer, there will be some function *g* that is literally *satisfied* by the system: the entities in the domain and range of the function will be entities physically tokened in the machine, and temporal sequences of tokenings of these entities in the machine will accord with the function itself. For instance,

The input to a typical adding machine is a sequence of button pressings: <C,M,+,N,+,>, i.e., <clear, first addend, plus, second addend, equals>. The output is a display state, D, which is a numeral representing the sum of the two addends. We may think of the button-pressing sequences as arguments to a function *g* that gives display states as values. An adding machine *satisfies* *g*; that is, the arguments and values of *g* are literally states of the physical system. (p. 89)

The function *g* is automatically computed by the physical system, by virtue of the system's causal architecture—computation being program execution, and program execution being "disciplined step satisfaction" (p. 92). Representation enters the scene when there is some further function *I*, involving

entities in its domain and range that are not literally tokened in the physical system itself, that is suitably related to the function  $g$ . In the case of a typical adding machine, for instance, this is the *addition* function, whose domain and range traffic in numbers rather than in states literally tokened in the physical device:

Addition...relates numbers, not physical states of some machine, so a physical system cannot literally satisfy the plus function. What an adding machine does is *instantiate* the plus function. It *instantiates* addition by *satisfying* the function  $g$  whose arguments and values represent the arguments and values of the addition function, or in other words, have those arguments and values as interpretations. (p. 89)

Representation is just the pairwise relation that holds between entities in the domain and range of  $g$  and associated entities in the domain and range of  $I$ , when  $g$  and  $I$  are themselves suitably related. For this interpretation relation between  $g$  and  $I$  to obtain, there must be a structure preserving 1–1 mapping from  $g$  to  $I$  (i.e.,  $g$  and  $I$  must be isomorphic); and certain further conditions must also be met—more about these presently. When all the conditions are met, the function  $g$  *simulates* the function  $I$ . Accordingly, Cummins calls pairing relation between  $g$  entities and  $I$  entities, *s-representation* (for “simulation representation”).

Three features of Cummins’s account deserve special emphasis. First is his way with *mis*-representation:

[I]n the world of *s*-representation, misrepresentation differs from failure to represent only in degree; failed representation becomes misrepresentation when the failure isn’t too bad.... [W]e cannot suppose...that first there is the issue of what (if anything) is represented and then there is the issue of what processes act on the representations. What is *s*-represented is essentially a matter of the processes, for it is essentially a matter of simulation. And simulation is essentially a matter of degree.

Thus, in the end he does not really hold that *s*-representation requires a full-fledged isomorphism between  $g$  and  $I$ ; evidently it is enough that there be something like an *approximate* isomorphism (the sort of relation that is easy to understand but hard to “explicate”).

Second, he regards *s*-representation as non-unique: normally a variety of different functions will be simulated, either perfectly or imperfectly, by a given function  $g$  satisfied by a computational system. Thus, there is no such thing as *s*-representation *simpliciter*, but only *s*-representation relative to a specific function, a “target of simulation.” For him, this is a virtue:

We needn’t *worry* that we can always trade misrepresentation of  $x$  for accurate representation of Something Else; we *can* do that, but it doesn’t *matter*.... No doubt adding machines simulate functions other than  $+$ , but that does not compromise the standard explanation of addition in adding machines. (p. 101)

Third, he never explicitly spells out the additional conditions for s-representation, over and above the requirement that the simulating function  $g$  should be (approximately) isomorphic to the simulated function  $I$ . He admits that the isomorphism requirement alone is much too weak. (If structure preservation is all that's required, then any symbol in any physical computational system will represent virtually anything at all—will have virtually any content—relative to *some* function simulated by the system's  $g$ -function.) And he does describe intuitively, via examples, the sorts of further restrictions he thinks are needed. One is that the simulated function should provide a *direct* interpretation of the physical system's symbolic data structures—so that the adding machine computes  $x + y$  but not  $2\pi(x + y)$ . “I must confess that I don't know how to define *directness*” (p. 104). Another is that the simulated function should not be a “degenerate” target for simulation—as is the function  $f$  that is identical to the addition function except that the number 5 is everywhere replaced, in the function's domain and range, by Richard Nixon. Intuitively, the trouble is that the numeral ‘5’ in the adding machine “doesn't track Richard Nixon as he is buffeted around by any natural discipline” (p. 105). “Once again, I am embarrassed by the fact that I have no general account of what makes [this function] a degenerate target for simulation” (p. 105).

*Mental* representation, as construed by the CTC, is just a special case of s-representation, claims Cummins.

[H]aving a cognitive capacity is instantiating a function that relates propositional contents, i.e., a function that takes propositional contents as arguments and values and relates them as premises to conclusion. A cognitive system, in short, is an *inference engine*—a system that merits an inferential characterization.... Under this conception, the problem of cognition becomes the problem of explaining the fact that the system is described by a *cognitive function*, or, for AI, of building a system that is described by a cognitive function (p.109).

Since, on this account, the notion of representation presupposed by the CTC is nonunique, it is not the same as the notion presupposed by common-sense belief/desire psychology. Underlying our ascriptions of beliefs and other propositional attitudes to one another, he recognizes, is the assumption that the contents of these states are (by and large) unique. Accordingly, he thinks, belief/desire psychology will not, and cannot, emerge as a simply a component of a developed computational theory of cognition; the intentionality of beliefs cannot turn out to be just a byproduct of the account of content that is appropriate for the CTC. He makes a few brief and speculative remarks about how common sense psychology might actually be related to the CTC, but leaves this matter largely open.

I turn now from summary to commentary. Let me raise three concerns about Cummins' positive proposal. First is a worry about the explanatory

relevance of content, given his (cheerful) admission that s-representational content generally will be nonunique. To the extent that any of various different, mutually incompatible, contents are all correctly assignable to the symbolic data structures that count (under the CTC) as mental representations, there arises the *prima facie* presumption that none of these contents really have any explanatory relevance to the system's performance. If the distinct content assignments to the data structures are all on a par *vis-à-vis* explanation of why the system does what it does, then it appears that they cancel each other out—i.e., none of them has any real explanatory relevance at all. Thus, to the extent that the CTC assumes the explanatory relevance of mental representation (and I agree with Cummins that it does), his account seems inadequate.

Of course the explanatory relevance of content is itself a highly vexed issue in contemporary philosophy of mind. Cummins does have some things to say about it. He lays down some plausible looking conditions on explanatory relevance; and he argues that s-representational contents, despite their nonuniqueness, satisfy these conditions. But the trouble is that the conditions do not seem *sufficient* for explanatory relevance, because they do not rule out cases of properties that seem intuitively to be mere epiphenomenal correlates of the properties that are genuinely explanatorily relevant to a system's behavior. If the representational content of symbolic data structures is nonunique, and if the various distinct but incompatible contents correctly assignable to data structures are on a par explanatorily, then it is hard not to conclude that these contents are all mere epiphenomenal correlates of those properties of the data structures that do real explanatory work.

My second concern involves Cummins' contention that there is a generic notion of representation that is both (i) nonunique, and (ii) applicable to computational systems at all levels of complexity—from lowly adding machines, to microcomputers and mainframes, to (if the CTC is right) human beings. There is a very different way of viewing representation in this complexity hierarchy, which many of us find more plausible than Cummins's. To wit: Symbolic data structures in adding machines are not representations *for the system* at all, but are only representations for the community of people who design and/or use these artifacts. Likewise, *mutatis mutandis*, for the microcomputers and mainframes that currently exist. Although it is a highly vexed question what it would be for data structures to be representations “for a system,” presumably the answer has something to do with how those structures would contribute to that system's capacity to successfully navigate a sufficiently complex body through the vicissitudes of embodied existence in the world. (The embodied system would successfully get around in the world, despite all these vicissitudes, because—



and only because—of systematically content-appropriate ways that these structures figure in its behavioral control systems.)<sup>13</sup>

Those, like me, who find this alternative perspective more plausible than Cummins's are not likely find persuasive his argument that the notion of representation presupposed by the CTC is a nonunique one. For, much of the dialectical force of his argument rests on the fact that in general there need not be any unique function that is mirrored by a computational system's inner symbol-manipulating activity. But this fact becomes relevant to the computational theory of *cognition* only if one accepts the contention that data structures in artifacts like adding machines are representations *for the system itself*. Only then does it begin to seem natural to say that for computational systems generically, for-the-system representation is mere simulation.

A third concern about Cummins's proposal is the size of the gap between what he seeks—viz., a "satisfactory explication," in nonintentional, nonsemantic terms, of mental representation as this notion figures in the CTC, and what he actually delivers. As noted already, he leaves the explication unfinished: he does not cash the notions of *direct interpretation* and *degenerate target of simulation*, to which he resorts in his account; and he admits that he has no idea *how* to cash them. This admission lends credence to the suspicion that a tractable naturalistic "explication" of the concept of mental representation is not to be had—not even for the relatively weak, nonunique, notion of s-representation which Cummins claims is the only one presupposed by the CTC. Still less, then, is a such an explication likely to be possible for the stronger kind of intentionality that Cummins himself concedes is presupposed by common-sense belief/desire psychology.

Concerning the large promissory note he takes out for 'direct interpretation', he says:

*Something* must account for the fact that instantiating *f* [a function] isn't enough to instantiate every function isomorphic to *f*. I am inclined to accept a kind of transcendental argument for the solvability of the directness problem: The standard...[simulational] explanation of addition is *correct*, after all, and it presupposes a nontrivial concept of interpretation; therefore, such a concept of interpretation exists.

When one ties these remarks back to the book's overall concern to find a "satisfactory explication" of the concept of representation, one sees that this transcendental argument evidently assumes, in effect, that if the CTC's

---

<sup>13</sup> It is necessary for the system actually to be navigationally tethered to such a body, or for the system's representations of its environment actually to be caused by an environment of the kind represented? I am inclined to say no; the Brentano intentionality of the system's representations could be radically at odds with its actual environment, and with the actual etiology and actual effects of those representations. The system could be, and could have always been, a brain in a vat. Cf. note 1, and Tienson (forthcoming).

notion of representation is not to end up in the catalogue the physicists are compiling of the ultimate and irreducible properties of things, then there must be *tractably specifiable* conditions, formulable in nonsemantic and nonintentional terms, for directness—and for being a nondegenerate target of simulation, and thus for s-representation itself. But, as I have stressed already in discussing Fodor and Baker, this kind of assumption is quite dubious. Even if there are no *tractably formulable* sufficient conditions for mental representation that are stable in nonintentional, nonsemantic vocabulary, it just does not follow that mental representation thereby acquires the same ontologically and explanatorily basic status as those properties, whatever they are, that the physicists are cataloguing.

**6. Explaining Supervenience.** One common thread running throughout this essay has been the serious possibility that thesis (F.15) is false, even if a physicalist/naturalist metaphysics espousing (A.15) is true. Let me close by tying this issue to a related one involving explanation.

Consider once again Fodor's remarks, quoted above just prior to (F.15), about aboutness not being in the physicists' catalogue of the ultimate properties of things. In some sense of 'ultimate', he is surely right. Moreover, for anyone who professes to hold what Fodor calls a "physicalistic view of the world," the non-ultimacy of intentionality should not be construed as merely a matter of supervenience upon the nonintentional. For, if certain inter-level supervenience facts are themselves *sui generis* and unexplainable, then the supervening properties will thereby qualify for inclusion on the list of ultimate and irreducible properties of things—supervenience notwithstanding. (From a physicalist/naturalist perspective, one reason to reject G. E. Moore's meta-ethical position is his claim that there are certain synthetic necessary truths, of the form 'Anything with natural property N is intrinsically valuable', that are utterly unexplainable and thus are metaphysically rock-bottom.) So metaphysical physicalism/naturalism should not merely assert thesis (A.18.a), which says that the intentional supervenes upon the nonintentional; it should also assert that inter-level supervenience facts are (at least in principle) *explainable*, rather than being themselves included among the fundamental, unexplainable, facts about the world.

Too little philosophical attention, it seems to me, has been directed at exploring what kinds of explanations might be possible for inter-level supervenience relations, and what kinds of criteria such explanations should meet.<sup>14</sup> Philosophical views (particularly those that are purportedly physicalistic) that invoke supervenience without explaining it are thereby deficient—which makes for a lot of deficiency in recent philosophy.

---

<sup>14</sup> For some relevant discussions, see Lewis (1973), Horgan (1984), and Horgan and Timmons (1992).

One general format for explaining supervenience relations would go roughly as follows: Certain families of higher-level properties (e.g., intentional mental properties) are jointly subject, as families, to certain semantic/conceptual constraints on their instantiability. Such constraints might include, for instance, the requirement that certain *ceteris paribus* generalizations (e.g., the most platitudinous *ceteris paribus* generalizations of "folk psychology") must come out true—or by and large true, anyway—whenever those properties are instantiated. Specific supervenience facts would be explainable if it would be possible in principle to show why all the semantic/conceptual constraints governing a family of higher-level properties are collectively satisfied by a family of lower-level realizing properties. (For further elaboration of this explanatory format, see Horgan and Timmons 1992.)

In sum, recent naturalistic programs in the philosophy of mind have been too stringent in one respect, and too lax in another. On one hand, it is excessive to insist on tractably specifiable sufficient conditions for intentionality; but on the other hand, it is not kosher to invoke supervenience relations unless they are subject to naturalistically acceptable modes of explanation. So some rethinking of programs is called for, especially for those of us who seek a philosophical account that accommodates intentionality within the natural order described by physical science. It seems entirely possible that things could turn out this way: although (i) there are no *tractably specifiable* nonsemantic and nonintentional sufficient conditions for intentional mental states (or for intentionality of public languages), nevertheless (ii) the supervenience of the intentional on the physical is indeed susceptible, in principle, to naturalistically acceptable modes of explanation.<sup>15</sup>

#### REFERENCES

- Baker, Lynne Rudder (1987): *Saving Belief: A Critique of Physicalism*. Princeton: Princeton University Press.
- Cummins, Robert (1989): *Meaning and Mental Representation*. Cambridge: Bradford Books/MIT Press.
- Cummins, Robert and Schwarz, Georg (1988): "Radical Connectionism," in T. Horgan and J. Tienson (eds.), *Spindel Conference 1987: Connectionism and the Philosophy of Mind*, Supplement, *Southern Journal of Philosophy* 26.
- Dretske, Fred (1981): *Knowledge and the Flow of Information*. Cambridge: Bradford Books/MIT Press.

---

<sup>15</sup> I thank David Henderson, John Tienson, Mark Timmons, and Michael Tye for helpful comments and discussion.

- Dretske, Fred (1985): *Machines and the Mental*, *Proceedings and Addresses of the American Philosophical Association* 59, pp. 23–33.
- Fodor, Jerry (1974): “Special Sciences,” *Synthese* 28, pp. pp. 77–115. Reprinted in Fodor (1981).
- Fodor, Jerry (1981): *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge: Bradford Books/MIT Press.
- Fodor, Jerry (1987): *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge: Bradford Books/MIT Press.
- Fodor, Jerry (1991): “A Modal Argument for Narrow Content,” *Journal of Philosophy* 88, pp. 5–26.
- Garfield, Jay (1988): *Belief in Psychology: A Study in the Ontology of Mind*. Cambridge: Bradford Books/MIT Press.
- Goldman, Alvin (1970): *A Theory of Human Action*. Englewood Cliffs: Prentice-Hall.
- Horgan, Terence (1984): “Supervenience and Cosmic Hermeneutics,” in T. Horgan (ed.), *Spindel Conference 1983: Supervenience*, Supplement, *Southern Journal of Philosophy* 22.
- Horgan, Terence (1989): “Mental Quasation,” *Philosophical Perspectives* 3, pp. 47–76.
- Horgan, Terence (1991): “Actions, Reasons, and the Explanatory Role of Content,” in B. McLaughlin (ed.), *Dretske and His Critics*. Cambridge: Basil Blackwell.
- Horgan, Terence (forthcoming): “Nonreductive Materialism and the Explanatory Autonomy of Psychology,” in S. Wagner and R. Warner (eds.), *Beyond Naturalism and Physicalism: Notes Toward a Program*. South Bend: University of Notre Dame Press.
- Horgan, Terence and Tienson, John (1988): “Settling into a New Paradigm,” in T. Horgan and J. Tienson (eds.), *Spindel Conference 1987: Connectionism and the Philosophy of Mind*, Supplement, *Southern Journal of Philosophy* 26.
- Horgan, Terence and Tienson, John (1989): “Representations without Rules,” *Philosophical Topics* 17, pp. 147–74.
- Horgan, Terence and Tienson, John (1990): “Soft Laws,” *Midwest Studies in Philosophy* 15, pp. 256–79.
- Horgan, Terence and Tienson, John (forthcoming): *Connectionism and the Philosophy of Psychology: Representational Realism without Rules*. Cambridge: Bradford Books/MIT Press.
- Horgan, Terence and Graham, George (1991): “In Defense of Southern Fundamentalism,” *Philosophical Studies* 62, pp. 107–34.
- Horgan, Terence and Timmons, Mark (1992): “Troubles on Moral Twin Earth: Moral Queerness Revived,” *Synthese*.

- Kim, Jaegwon (1966): "On the Psycho-Physical Identity Theory," *American Philosophical Quarterly* 3 (1966), pp. 227–35.
- Kim, Jaegwon (1970): "Causation, Nomic Subsumption, and the Concept of Event," *Journal of Philosophy* 70, pp. 217–36.
- Kim, Jaegwon (1976): "Events as Property Exemplifications," in M. Brand and D. Walton (eds.), *Action Theory*. Dordrecht: D. Reidel.
- Lewis, David (1973): "Radical Interpretation," *Synthese* 23, pp.331–44.
- McGinn, Colin (1989): *Mental Content*. Oxford: Basil Blackwell.
- Stich, Stephen (1983): *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge: Bradford Books/MIT Press.
- Tienson, John (forthcoming): "The Private Life of the Brain in the Vat."