Block, N. and S. Bromberger (1980). "States' rights." Behavioral and Brain Sciences 3(1): 73-74.
This is a response to Jerry Fodor's article, Fodor, J. (1980). "Methodological solipsism as a research strategy in cognitive psychology." Behavioral and Brain Sciences 3: 63-109.

(refer to) the same thing, then they can be substituted in a context without changing its truth value.

transparent (context): to a first approximation, a sentential context which is *not* opaque (q.v.). However, see text.

type/token relation: the relation which holds, for example, between a word and an inscription (or utterance) of that word. By analogy, the relation which holds between a kind and anything which is *of* that kind.

Open Peer Commentary

Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article.

by **Ned Block** *and* **Sylvain Bromberger**

Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, Mass. 02139

States' rights

Fodor's "formality condition" is tied to a conception of mental processes which, we believe, requires careful attention. In particular, we believe that such attention may weaken two corollaries of his main thesis.

According to Fodor, mental processes have two essential characteristics: (a) they are "computational processes" which are "defined over representations," and (b) they are "processes in which representations have their causal consequences." Now cognitive psychology must unquestionably include a theory of mental operations on mental representations, for there is much good evidence that such operations exist. But the processes that have behavior and changes of psychological states as *causal* outcomes can't be these, that is, can't be processes simply "defined over representations." Rather, they must be operations "defined over" mental *states* and representations. The reason is obvious: the consequences of a belief that Marvin is melancholy are different from those of a hope that Marvin is melancholy, and from those of a doubt that Marvin is melancholy.

Clearly, Fodor knows this. In some places he says so explicitly. But in others he seems to equivocate. He writes, for instance, in a passage on mental processes as computational that "To a first approximation, we may thus construe mental operations as pretty directly analogous to those of a Turing machine." However, the outputs of a Turing machine are determined not only by *what is on the tape,* but by the states the machine is in and gets into as it peruses and modifies the symbols it scans. Or again – in the course of an argument – he writes "And the upshot of *this* is just the formality condition all over again. Given that mental operations have access to the fact that P is believed." But the belief that P is not a mental representation (though it may have mental representations as constituents). It is a mental state. Elsewhere, the distinction is still less clear. Over and over again, Fodor identifies mental causal operations with mental processes of which (a) above holds. He writes, for example, "It is what the agent has in mind that causes his behavior"; mental processes are computations. (Computations just are processes in which representations have their causal consequences in virtue of their form)"; or "By thus exploiting the notions of content and computation *together*, a cognitive theory seeks to connect the intensional properties of mental states with their causal properties vis-à-vis behavior." In short, Fodor apparently wants to ignore the distinction between two notions of mental operations: the notion of algorithmlike operations appropriate for theories of, for example, inference drawing and language processing (and which is a notion of operations that have representations as ranges and domains) and the notion of operations that account for the causal consequences of mental states (and which have mental states as domains.)

We think that proper attention to the difference between operations on representations and operations on mental states (with representations as constituents) would require the reformulation of a number of Fodor's arguments. This could perhaps be done without much damage to his most interesting conclusions (though it is not clear to us what will become of the formality condition when it is turned into a condition on mental state operations). We cannot go into this here. But the corollaries that "two thoughts can be different in content only if they can be identified with relations to formally distinct representations" and that "there can't be a psychology of knowledge" will have to be reconsidered.

The first of these doctrines amounts to the claim that mental representations must be unambiguous. For suppose that mental representations could be ambiguous, that is, that a representation with one form could have different contents on different occasions. Then there could be two different thoughts containing the same form of representation, contrary to the doctrine. (Fodor also endorses the "nonambiguity of mental representations" doctrine explicitly, saying "mental representations are distinct in content only if they are also distinct in form.")

The argument for the nonambiguity of mental representations seems to run as follows: If two thoughts could differ in content without having formally different representations as objects, then they could have different causal consequences. This strikes Fodor as unlikely: "to put it mildly, it is hard to see how internal representations could differ in causal role *unless* they differed in form." But it is not hard to see at all! Fodor readily recognizes that formally equivalent representations can have different causal consequences when they are the objects of radically different states, for instance, belief and desire. So let us restrict ourselves to beliefs. Even then it is not very hard. Pretend that the language of thought is like English in having the form 'bank.' Here is how tokens of that single form could on some occasions have the causal role associated with *river bank,* and on other occasions that associated with *financial bank.* Take a silly possibility first: there could be a system which incorporates a timekeeping device that gives beliefs expressed using "bank" one causal role on Mondays, and the other causal role on Tuesdays. More seriously, there could be a system (there possibly is one) that applies one set of procedures to 'bank' in one type of context (e.g., the context of conversations about checks, money, or Bert Lance), and another set of procedures in another type of context (e.g., contexts of conversations about fishing or about drownings). A representational theory could deal with such possibilities (if they turned out to be actualities) by describing these systems as put into different (cognitive) states under different conditions, states involving – to use Fodor's terminology – different relations to the same representation. Such an account would, of course, require a notion of content as a function of both state and form of constituent representation (à la Turing machine). It would also require a catalogue of states more refined than the everyday one of beliefs *tout court,* hopes *tout court,* desires *tout court,* and so on. But there is no methodological or currently available empirical reason to reject any of this. On the contrary.

Fodor could respond by saying that what we have described as a single form with different contents is really a case of different *forms.* In the case where 'bank' has one causal role on Mondays and another on Tuesdays, he might say that the day of the week *as represented on the internal clock* should be deemed to be part of the orthography of the internal system. We have '$bank_M$', and '$bank_T$', depending on the clock reading. Similarly, in the most realistic case, he might say that since the difference in contexts would be internally reflected, this representational difference should also be taken as providing different subscripts on 'bank'.

However, if Fodor takes this line of reply, he will turn the claim that different content implies different form into a near triviality. If *any internal* difference that determines a difference in causal role *thereby* counts as defining a difference in form, the thesis that there is no difference in content without a difference in form loses the substance that Fodor imputes to it. He would also blur the distinction between state and representational content of state beyond acceptable limits.

The "No psychology of knowledge" doctrine is put by Fodor as follows: "Take, for example, knowing that such and such, and assume that you can't know what is not the case. Since, on this assumption, knowledge is involved with truth, and since truth is a semantic notion, it

Downloaded from https://www.cambridge.org/core. NYU Medical Center: Ehrman Medical Library, on 09 Feb 2018 at 22:22:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0140525X00002041

Text

is going to follow that there can't be a psychology of knowledge." It is clear from the context that when Fodor writes that truth is a "semantic" notion, he simply means that it is not a formal notion, that is, that truth and falsity are attributes that mental representations (or tokens thereof) do not owe to their form, but owe perhaps to their "corresponding" to something distinct from themselves. First note that this tenet, vague though it may be, does not hold in any obvious way for a number of representations, for example, truth functional tautologies, sentences like 'this sentence is in English,' 'I am here now,' and so on. Does Fodor believe that the language of thought does not have analogues of these? More to the point, think of arithmetical propositions. But first let us see how Fodor would argue for the "no psychology of knowledge" doctrine. He would presumably adapt his argument for the thesis that "it is mandatory to assume that mental processes have access only to formal (nonsemantic) properties of mental representations" – which currently is concerned with truth, and says nothing about knowledge – to cover the case of knowledge, by replacing rule *a* there ("Iff it's the case that P, do such and such.") with "If you *know* that P, do such and such," and making corresponding adjustments throughout. But the resulting argument carries no plausibility when P is thought of as an elementary arithmetical proposition. What are we to make of our knowledge of such propositions then? We, the commentators, of course need not answer. Nor would we. That question has been the bane of too many philosophers. But attention to states suggests a way out. A representational theory might well entertain the hypothesis that our cognitive relation to these propositions (as well, by the way, as to the rules that represent our linguistic competence) constitutes a specific type of cognitive state, a state appropriately called knowledge, and, by the way, a state whose causal antecedents and consequents we could study without being involved in the sort of vacuous "ideal of pure reason" that Fodor dismisses near the end of his paper.

We would like to end with a question about Fodor's notion of formality. He warns us candidly that "the notion of formality will . . . have to remain intuitive and metaphoric, at least for present purposes" after telling us that "Formal operations are the ones that are specified without reference to such semantic properties as, for example, truth, reference, and meaning." Such vagueness is probably unavoidable at present. Still, is there any good reason for including "meaning" in the list? Are there any arguments against the adoption of a modified "formality condition" that would allow mental processes to exploit aspects of the meaning of representations? Fodor's arguments deal with referential attributes, that is, with extensional semantic attributes, and he writes as if they can be generalized to cover intensional semantic attributes. We don't see how this is to be accomplished.

by **Eugene Charniak**

Department of Computer Science, Brown University, Box 1910, Providence, R.I. 02912

Some aspirin for Dasein

As a believer in both the representational and the computational theories of mind I ought to be cheered by professor Fodor's article. He has, after all, told me that as far as my belief in these theories is concerned, I am on the winning side. Any joy, however, must be tempered by the fact that Fodor himself seems so unhappy with his conclusion that a theory of *Dasein* is not feasible. Perhaps I read too much into it, but to me his final lines bespeak a troubled man. I am afraid I can do nothing about the ultimate causes of his troubles, since his conclusions seem basically correct. My more modest goal is to do something about the pain.

The main thing to keep in mind before we grieve over the death of *Dasein* theory is that it could have no empirical consequences, or at least none which was not predicted by the formal theory. How could it? As Fodor has cleverly shown, one must reject the notion that "mental operations have access to the semantic properties of mental representations." It follows from this that all human (intellectual?) activities must be explicable in formal terms if they are to be explicable at all. This would not leave much room for *Dasein* theory.

In particular, when Fodor says that "there can't be a psychology of perception" he is using a "psychology of perception" in a particular way – one in which perception leads to truth, not mere belief. Indeed it is reasonably easy to outline a theory of the latter. Fodor could do it as well as I, but since I intend to build upon the example, let me elaborate.

The basic idea is simple: to get a computational theory of vision we need only to take a TV camera, or some other device which turns light into electrical signals, and hook it up to an appropriately programmed computer. Of course, if one just wants the theory, then the hardware is not necessary. However, if these theories are going to be as complicated as I believe, it will be handy to encase the theory in a program, and actually run the program, if only to determine what behavior is predicted.

To test such a program we might start by holding an object like a pencil in front of the camera and having the machine name it. (In saying that the object is a pencil I do not have in mind an equivalent to the NaCl description of salt. Rather we would certify it to be a pencil if, say, at least four out of five consumers thought it was a pencil.) If it worked for a large number of objects we would be encouraged. If the theory also accounted for reaction time experiments and so on, we might begin to think the theory was true. We would then want to try it out on more complex visual input, perhaps requiring the machine to negotiate Boston traffic.

This is all fantasy, of course, but while we are fantasizing, there is no reason to stop there. We might want to include a learning theory and "bring up" our machine to talk about, and recognize, Boston traffic, Robin Roberts, and Cincinnati. Perhaps we could even individualize these programs so that it would mimic particular individuals.

Now this is all pretty wild. Such ideas are closer to science fiction than cognitive psychology or artificial intelligence. But whatever the problems are in accomplishing such tasks, the lack of a theory of *Dasein* is not one of them. As we have already noted, if such things are possible at all, the theories and programs which accomplish them will not need truth, reference, and so on (or at least not the high standards of truth, reference, and so on that Fodor desires). To see this we need only elaborate our thought experiment. This time, when the program is not "looking" we will hook up the TV camera (or whatever it is) to a device which mimics the signals the real world sends, but which is not constrained by reality. So while we might hold up a picture of food lines from the great depression, our demonic device sends the camera signals corresponding to, say, a banana. We are back to Winograd's (1972) block world.

My point then is that, for this machine, at least, all ideas about the external world are only beliefs, not knowledge. The machine may very well not go along with us in this assessment, however. Nevertheless, we will be inclined to attribute the machine's insistence on the reality of what it sees to the way in which it was programmed rather than to anything intrinsic in the way in which it interacts with the world.

The extension of this thesis to people is neither difficult nor original.

One final thought experiment. Extending the idea of an individualized theory of learning we could imagine a "Fodor theory," which, embedded in a program, would act just like the real thing. Again, such a theory is not in the offing, but given the above discussion, it cannot be for the lack of a *Dasein* theory. One of the acts of the Fodor theory/program would be the creation of a paper on methodological solipsism, claiming, among other things, that "I don't know what *Dasein* is, but I'm sure that you and I and Cincinnati have all got it." The real Fodor presumably would know better. He would know that to the theory/program "Cincinnati" is just a formal object – no truth or references allowed. What I am trying to figure out is whether Fodor will convince the machine. If so, then maybe the pain can be relieved.

by **Paul M. Churchland**

Department of Philosophy, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

In defense of naturalism

Central among the aims of psychology must be to explain how the brain manages to evolve within itself an increasingly powerful and

Downloaded from https://www.cambridge.org/core. NYU Medical Center: Ehrman Medical Library, on 09 Feb 2018 at 22:22:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0140525X00002041

purely physiological terms, and so can any other pattern of retinal stimulation. But it may be useful for some psychologist's purpose to group together a set of physiologically distinct stimuli under the description 'visual stimulation of a kind normally caused by a human face,' Maybe that causal description, or some refinement of it, is the nearest he can get to defining a type of input state in terms of which he can then go on partially to define a type of psychological state.

A particular stimulus can be of that kind without being actually caused by a face in front of S. This mode of type-identification does not contravene methodological solipsism in its mild form. But it does involve in a more general way a claim about the kind of world that S is in. If there were no such things as human faces, there could be no such stimulus-type. The issues here connect up in interesting ways with the work of J. J. Gibson (1966), and ultimately with Kant.

The definition of output-types raises a further problem which goes to the very heart of solipsism. What, and where, is S? Suppose that Tom is in a state (it might be *wanting to wave*) which is a type that normally, under favourable conditions and given certain other states, causes a movement of one of Tom's hands. Suppose also that a functionalist wishes to partially define that psychological state in terms of that functional property. The response-type *moving a hand* cannot exist (nothing can count as a token of it) without there being a hand. But the desire could still exist, even if Tom had no hands. Provided that Tom believes he has hands and can wave, it is psychologically possible for him to be in the same state of desire even if the *world* were different in a relevant respect from what he believed it to be.

It may be replied that since Tom's hands are part of Tom, no presupposition is made about the existence of anything in the world outside Tom in calling his response a movement of a hand. However, a methodological solipsist can't duck the serious point here. Are we to take the 'subject' of psychological states to be, for the purposes of psychological theory, a disembodied mind? If not, then how much of S's body lies inside the boundary that is implied by thinking of mental states as internal states? Tom's narrow mental states could be what they are even if he had no motor nerves at all. Presumably he could have the visual experiences he has even if he lacked eyes. One is tempted to draw the boundary further and further in until the only physical entities whose existence is assumed are the neurons whose activities are the physical realizations of psychological processes [see Puccetti & Dykes: "Sensory Cortex and the Mind-Brain Problem" *BBS* 1(3) 1978]. But this kind of existence assumption is not grounded in the nature of the representational content of those processes (assuming that they are not irreducibly *de re* thoughts about those very neurons), but is forced upon the psychologist by his adherence to physicalism. If the semantic facts adduced by Fodor were the only ones deemed admissible to use in determining existential commitments, I see no reason why a rational psychologist should baulk at saying that S is an ideal entity, a representing device that has no essential links with a physical world at all.

Descartes held it to be conceivable that his experience of the world was illusory, including his impression that he had a body. Most modern philosophers would argue that epistemological solipsism is inferior as a theory to realism, and many would argue this partly on methodological grounds. Methodological solipsism, likewise, should be evaluated on the basis of its usefulness for some purpose. Fodor has once again pursued a certain line of argument further than anyone else dares. The paper demands that we think hard about what kind of science a psychology of intentional states can and should be.

Author's Response

by **J.A. Fodor**
Department of Psychology, Massachusetts Institute of Technology, Cambridge, Mass. 02139

Methodological solipsism: replies to commentators

My critics disagree with one another about as much as they disagree with me, and I admit to the temptation to retire to my tent and let them fight it out among themselves; all the more so since limitations of time and space make it next to impossible to do justice to some of the points that they have raised. However:

Block & Bromberger. 1. I have indeed played fast and loose with the question of whether mental operations apply to representations or to states; the reason being that I had hoped to spare myself (and the reader) the excruciating details. Probably the canonical formulation ought to be that mental operations apply *only* to states (it is states that are causally interrelated) but that mental states are relational (with mental representations figuring among the relata) and, in the relevant cases, mental operations apply to mental states in virtue of the form of the representations. So, for example, to say (short form) that a mental operation permutes symbols would be to say (long form) that a state which consists of a relation to a representation in which the symbols occur in the order AB is causally sufficient to bring about a state which consists of a relation to a representation in which the symbols occur in the order BA. It is, on balance, easier just to say that there is a mental operation which permutes a pair of symbols.

2. I'd reply to the "bank" point just the way that Block & Bromberger expect me to: the formality condition is to be made compatible with the no ambiguity condition by the (boring) trick of gerrymandering the identity conditions on representations (or states, or both). If this is uninteresting, that's because the formality condition wasn't supposed to be about how you count states or representations, given that the latter *are assumed to be specified nonsemantically;* on the contrary, the formality condition is precisely *the requirement of nonsemantic specification*. What determines the syntactic conditions upon identity and difference of mental states is not the formality condition but the argument-to-the-best-explanation. What looks like gerrymandering from an a priori point of view is just the usual process of individuating theoretical entities in whatever way is required by the a posteriori demands of theory construction.

3. If logical truths seem to be an exception to the formality condition, that is presumably because (and to the extent that) there is some formal property of representations which is coextensive with logical truth (in something like the way that derivability is coextensive with validity). And, of course, the formality condition does *not* preclude access to semantic relations when they are coextensive with formal ones, so long as the access is achieved, as it were, via the latter. Beyond that, I have nothing at all to say about the problem that Block & Bromberger raise except that it is very hard.

4. I knew I'd get into trouble about meaning. What I *hope* (see the discussion of **Rey** below) is that it can be eliminated from the list: that a notion of content-as-inferential-role, together with a theory of truth and reference, will give us everything we need to specify the various notions of content that we require. But I haven't any argument for this, and I can't say anything sensible about whether the formality condition allows access to the meaning of mental representations until somebody obliges by saying something sensible about meaning.

Charniak. I'm not at all sure that I understand what Charniak's position is, or even how he's carving up the options. I suspect – though it would take lots of discussion to convince either of us that this is so – that there is a very radical difference between the ways we look at things, and that it comes out in what we would take to be the criteria for successful theory construction. Charniak says he would be happy with a device which takes something to be a pencil if and only if "four out of five consumers thought it was a pencil." But this criterion is seriously underdescribed, since

Downloaded from https://www.cambridge.org/core. NYU Medical Center: Ehrman Medical Library, on 09 Feb 2018 at 22:22:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0140525X00002041