

NOTIONS OF 'REPRESENTATION' AND THE DIVERGING INTERESTS OF PHILOSOPHY AND EMPIRICAL RESEARCH

Steven Horst
Wesleyan University

Many of you may have had a neighbour or a grandfather who was a fanatic about tools.¹ Grandpa owned everything Sears ever made, and took it as a sign of moral turpitude if you used, say, a flathead screwdriver to take out a Phillips-head screw. Now I confess that I still do not own a Phillips screwdriver, and also use my flathead screwdriver to open paint cans. But Grandpa did rub off on me a little bit: I eventually became a philosopher, and take it that an important part of what philosophers do consists in the examination of conceptual tools.

What I am going to do here is look at one of the main conceptual tools used in cognitive science: namely, the notion of representation. Due to limitations of time, the discussion will be very programmatic in nature, and much of the detail that might be supplied in a longer discussion is left out. I hope to persuade you of three main points: First, there is real unclarity and even ambiguity in the usage of the word 'representation'. Second, different notions of representation are suited to different tasks—in particular, that some philosophical projects require a stronger sense of the word than do most empirical theories. Third, it is possible to adopt a weak construal of the word 'representation' that lets us have a good scientific psychology and lets us have intentional realism too, but without providing a representational explanation of intentionality.

Now my interest in clarifying the very *notion* of representation may come as a bit of a surprise. There seems to be a widespread assumption in cognitive science circles that we know perfectly well what we mean when we say something "is a representation," and hence the only *interesting* questions are about what *sorts* of representations are used, how they are realized in the brain, and how they get this peculiar property of *representing something*. And the assumption that we know what it is to be a representation is quite understandable, since the notion of *representation* looks like an old friend. After all, we have known about representation as long as we have known about symbols and pictures and maps and *leitmotifs*. Anybody who knows English knows what we mean when we say that a picture or a symbol is a representation.

While this attitude is perfectly natural, it also involves some implicit (and arguably none-too-innocent) assumptions about the word ‘representation.’ I shall call this set of assumptions the “Received View.”

Received View:

- 1) There is a single, univocal and familiar notion called ‘representation.’
- 2) There is a corresponding *property* called “being a representation.”
- 3) There are a number of paradigm examples of things that possess this property, e.g.:
 - symbols
 - pictures
 - maps
 - schematic drawings
 - *leitmotifs*, etc.
- 4) Philosophers and psychologists have long known that mind/world relations are mediated by *mental* representations.
- 5) Until recently, it was impossible to hook up a representational theory of content with a causal account of mental processes.
- 6) The paradigm of machine computation has shown us how to make this connection.

1. PROBLEMS WITH THE RECEIVED VIEW

Once we spell things out this way, some of the assumptions begin to look a bit dubious. First, it is undoubtedly true that there is an abundance of paradigm examples to which the *word* ‘representation’ is applied. But it is not at all clear that the same *notion* is expressed by that word in different cases, nor that there is a *property* called “being a representation” that is shared in common by pictures, symbols, and the rest. It seems just as likely that the word ‘representation’ is homonymous, or that it signifies a family resemblance rather than a common property. So here is the first problem: it is not clear that there is *one* thing that we are saying even of all the *familiar* cases when we call them “representations.” And hence it is not clear what we are saying of the *new* cases postulated by cognitive science when we call *them* “representations.”

Second, some people *have* tried to articulate an analysis of ‘representation’ broad enough to cover the paradigm cases. I am thinking of a diverse group of writers that includes Thomas Reid [1983], Edmund Husserl [1913], A.J.Ayer [1968], Daniel Dennett [1969], Kieth Lehrer [1989] and Brenda Judge [1985]. According to these people, what it is to be a representation is to be used and interpreted in a particular way:

To say that **A** is a representation of **B** is (implicitly) to say that there is some *interpreter I* who *uses A to stand for B*.

As these writers have been quick to point out, there are disastrous consequences if you try to apply this interpretation-dependent notion of representation as an explanation of meaningful mental states, as it leads to a regress of homuncular interpreters, and requires you to assume the very things you are trying to explain—in particular, the mind’s access to extra-mental reality.² I shall call this line of criticism *the Reid/Husserl objection*. One might see the complaints about the interpretation-dependence and “derived intentionality” of mental *symbols* raised by John Searle [1980, 1984] and Kenneth Sayre [1986] as falling essentially along these same lines.

It is important to emphasize that the point of the Reid/Husserl objection is *not* that pictures and symbols and the like have two separate features: *one* of being representations and *another* of being convention- and interpretation-dependent. The point, rather, is that this convention- and interpretation-dependence is *built right into the notion of ‘representation.’* On this view, someone who speaks of “non-interpretation-dependent representations” is merely showing that she does not know how the word ‘representation’ is used. It would be a mistake on the order of saying “tool without a purpose.” What it is to be a tool is precisely to be usable for some purpose, and what it is to be a representation is precisely to be used to stand for something else.³

So now we have two reasons to be interested in an analysis of the notion—or, better, of *various alternative* notions—of representation: (1) We may already have several usages of the word, and hence it is not clear which, if any, is being used when we speak of “mental representations.” (2) There is an argument of respectable lineage to the effect that a correct understanding of the notion of representation reveals a kind of dependence on interpretation that would render that notion useless for explanation in cognitive science.

To these two reasons I shall add a third: namely, (3) that there is some danger that we may argue for the need for *one* notion of representation, then invalidly draw conclusions based on a subtle elision to an alternative meaning of the word. We might, for example, argue that we need something like mental pictures, then argue as though we had proven that our “representations” had to have syntactic features, or argue that thought involves “representing the world as being thus” and then

proceed upon the stronger assumption that thought involves *entities* called “representations” having semantic properties of the same sort that are attributed to symbols.⁴ In these sorts of cases, there would be an ambiguity of the term ‘representation’ in the argument that would render the argument invalid and paralogistic.

2. THE SEARCH FOR A RULE FOR THE USE OF ‘REPRESENTATION’

The task, I take it, is to supply a rule for the usage of the word ‘representation’ such that: (a) it is adequate to the explanatory tasks to which the word is put in theories in cognitive science, and (b) it manages to avoid the Reid/Husserl objection. Now I am aware of four basic strategies for supplying a rule for the use of a word: continuity with ordinary usage, definition by ostension, stipulative definition, and theoretical definition. Some of these strategies seem to be present in the philosophical literature on representation, and others are not. I take it, for example, that Fodor’s language of thought hypothesis involves an extension of the familiar notion of *symbolic* representation to a new domain, and that Fodor [1975, 1981, 1987] intends us to apply what we already understand by the word ‘representation’ when it is applied to symbols when we conceive of “mental representations”—in particular, we are to think of them as entities having both syntactic and semantic properties. By contrast, Robert Cummins [1989] develops a technical notion of “S-representation” which could well serve as a stipulative definition of the word ‘representation’ as Cummins uses the term. It seems rather doubtful that the third method, the method of ostensive definition, would be used for mental representations, since these are generally regarded as *theoretical* entities, and theoretical entities are, in principle, difficult to point to. The very fact that mental representations are conceived as theoretical entities, however, suggests the fourth strategy: namely, that we regard the *expression* ‘mental representation’ as being an expression that is theoretically defined.

It is perhaps important to be clear about what I mean by a “theoretical definition” of a term, and how this might work with respect to the word ‘representation’. What I mean by “theoretical definition” is that sometimes the best way to determine what a technical expression *e* means in a theory *T* is to forget entirely about what it means in other contexts, and look exclusively at the

explanatory work it does in *T*. When a term's meaning is entirely determined by the explanatory work it does in a theory, I shall say that it is "theoretically defined." A basic form of such a definition for the word 'representation' might look something like this:

'representation' in theory *T* =*df* "whatever it is that does *x*"

where '*x*' names some subset of the things explained by *T*. For example, '*x*' might mean one or more of the following:

- accounts for the intentionality of mental states
- stands in appropriate informational relationships with environmental factors
- tracks environmental factors in a fashion that allows for adaptation, etc.

Of course, one should not assume from the outset that the word 'representation' does the *same* work in every theory in which it is used.

3. ROBUST AND THEORETICAL CONSTRUALS CONSTRASTED

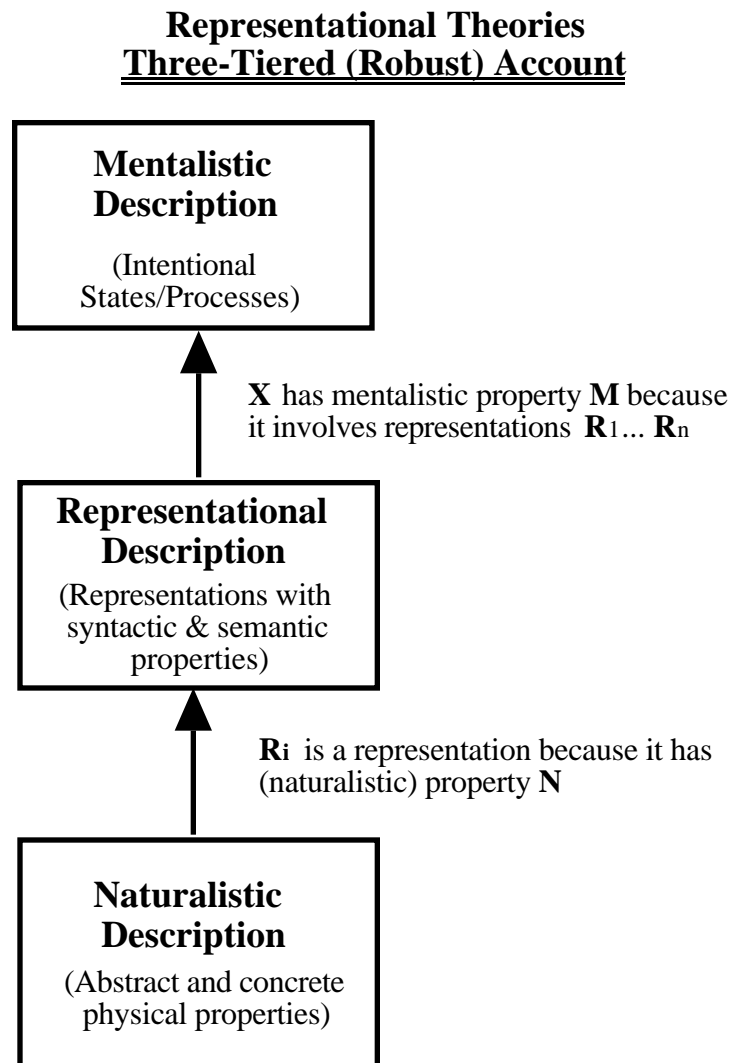
Granted that there is a dry philosophical question to be asked about different senses of the word 'representation,' we might well ask who ought to care and why. The first part of my answer is that I think the Reid/Husserl objection needs to be taken seriously. *If* theoretical work in cognitive science really depends on the notion of 'representation' that Reid and Husserl articulate, it is built on an untenable conceptual foundation. (I think some people have really steered clear of cognitive science because of just this kind of worry.) And the best way to show that it does *not* depend on *that* notion is to articulate an *alternative* sense for the word 'representation' that fits the theories and escapes the problems of conventionality.

But there is also another issue here. There are important differences between what you can *do with* a theory that takes ‘representation’ as a theoretical term and what you can do if you build some semantic presuppositions into your definition of ‘representation,’ either by continuity with familiar paradigms or by stipulative definition. There are six main points about these differences:

- 1) The structure of the explanations is different for the two cases.
- 2) The theoretical construal involves no commitments to dubious semantic properties.
- 3) The theoretical construal avoids the Reid/Husserl objection entirely.
- 4) The theoretical construal does not explain the intentionality of mental states.
- 5) The theoretical construal is very attractive as an account of the use of the word ‘representation’ in empirical theories in cognitive science.
- 6) The theoretical construal does not do all that is desired by some philosophers.

To see how the structure of the explanations is affected, consider as an example how theoretical definition differs from making wholesale use of the familiar notion of *symbolic* representation. The interpretation of cognitive science most familiar to philosophers treats intentional states as relations to symbolic representations that may literally be said to have both syntactic and semantic properties. Now part of what makes the notion of symbolic representation so interesting is that, in addition to having the syntactic overtones needed for computation, it also has *semantic* overtones built into it. To call something a “symbolic representation” is to impute to it semantic properties. *And semantic properties seem like the right kinds of things to explain semantic properties:* hence one can plausibly view the semantic properties of mental states as “inherited” from those of the symbolic representations they contain.

It is thus tempting to see the computer paradigm as leading to a three-tier account of intentionality:

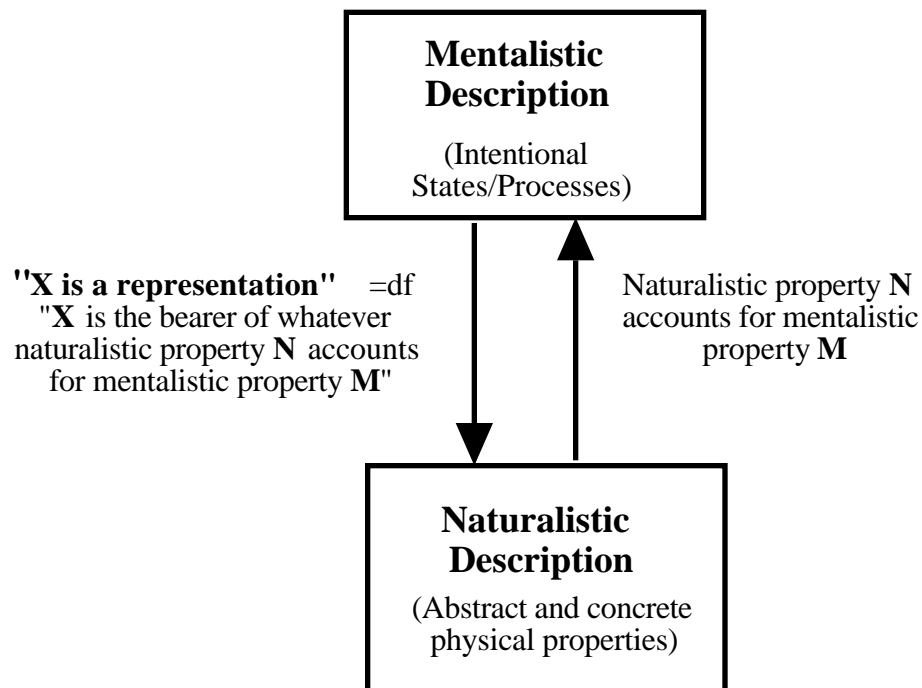


Mentalistic properties, lying at the highest level, are explained as relations to symbolic representations, lying at a second level of description. Here, what it is to be a symbolic representation is not defined in terms of the contribution to mental states, but has some independent meaning—presumably the same meaning that the expression “symbolic representation” usually has when applied to symbols. So there is *one* explanation that links the upper two levels: the intentionality of mental states is explained in terms of representation. There is then a further question of how to link up the representational level with more straightforwardly naturalistic properties. This would be a separate explanation linking the lower two levels of the hierarchy.

These two explanations are easily regarded as separable if there is some property of “being a representation.” In the Introduction to *RePresentations*, for example, Fodor seems to regard the one sort of explanation as having already been carried out, while the lower sort—a “theory of how representations represent”—might not realistically be available.

Constrast with this what happens if you take “representation” as a truly theoretical term.

Representational Theories Two-Tiered (Theoretical) Account



In this case, the term ‘representation’ cannot do any truly independent explanatory work. If you say, for example, that “representational properties are those properties of the objects of inner computation, whatever they turn out to be, that account for intentionality,” then you *don’t* have some property called “being a representation” that is offered as an explanation of the intentionality of the mental. (If you *do* try to do it that way, “representational properties” end up lacking explanatory power for the same reasons that *dormative virtues* lack explanatory power.) Instead, the word ‘representation’ functions as a kind of place-holder for whatever naturalistic properties might turn out to account for intentionality. There is no middle level of explanation, and indeed as yet there is

no account of intentionality: one is merely pointing to the possibility of a hoped-for explanation somewhere down the road.

4. THE THEORETICAL CONSTRUAL AND THE DIVERGENCE OF INTERESTS

This kind of theoretical construal of ‘representation’ has a significant virtue and a nontrivial price. The virtue is that it completely avoids the Reid/Husserl objection. That objection arises only when we try to transport some convention- or interpretation-dependent notions into our explanation of intentionality, and the theoretical construal does not do this. The price is that we lose the ability—or perhaps it was merely the appearance—of having a notion of representation that is itself robust enough to do some real explanatory work in the absence of a full-blown naturalized psychology. In particular, we do not have a notion of representation that can do any work in explaining the meaningfulness of mental states.

Here, I think, there could be a certain parting of the ways between empirical scientists and at least some philosophers. For the desire for a semantically robust notion of representation is driven in large measure, if not in full, by a distinctively *philosophical* wish list: in particular, by a desire to provide an analysis of necessary and sufficient conditions for meaningfulness for the mental, and to do so in accordance with particular ontological predilections. (That is, materialist ones.) But it is not at all clear that these are issues that the empirical psychologist needs to worry about. It is certainly my experience that scientists *do* not, by and large, get very excited about these issues, much to the frustration of some philosophers. And so, if the distinctively *scientific* requirements for a notion of representation do not themselves require a semantically robust construal, scientists might well have reason to prefer the unproblematic theoretical construal. And this, I think, is really the right line to take.

In my view, what makes cognitive science so attractive is that it seems to hold some promise of supplying psychological explanations that bear two crucial marks of scientific maturity. First, they might describe the systematic relations of the domain of inquiry in a fashion that is rigorously mathematizable. Second, they may describe connections between different levels of explanation—e.g., upwards to economics and downwards to neuroscience.⁵ And this project, I

think, is quite neutral with respect to ontological preference and to whether the relations between the levels is one of reduction or supervenience or mere correlation. From the standpoint of science, it is good enough to specify the mechanisms through which a psychological process is, to use an intentionally neutral term, *realized*. The question of whether realization involves supervenience or contingent identity or correlation or perhaps some relation that is *sui generis* in the case of psychophysical relations is really quite beside the point so long as we are just doing empirical science. And that, I suspect, makes up a good healthy part of the reason that it is so hard to get psychologists interested in Cartesian demons, brains in vats and counter-earthers. We might make an analogy with Newtonian mechanics. Newton's laws describe the *how*—the “in what fashion”—of gravitationally-induced motion without saying a word about *why* gravitational bodies attract. And while this may leave us, as it apparently left Newton, a bit dissatisfied, it is perfectly good science nonetheless. My suggestion is that psychology can be a perfectly respectable science by specifying the *how* of cognition—the functional relations and the mechanisms through which they are realized—even if it has nothing to say about the *why* of meaningfulness.

Moreover, I strongly suspect that most if not all cognitivist explanation in psychology could be accommodated by a theoretical construal of ‘representation’ along the following lines, which I like to describe as a Bowdlerized version of computational theory of mind.

Bowdlerized Computational Theory of Mind (BCTM):

- (B1) The mind's cognitive aspects are functionally describable in the form of something like a machine table.
- (B2) This functional description is such that
 - (a) attitudes are described by functions, and
 - (b) contents are associated with local machine states. Call these **cognitive counters**.
- (B3) These cognitive counters are physically instantiable.
- (B4) Intentional states are realized through relationships between the cognizer and cognitive counters. In particular, for every every attitude *A* and every content *C* of an organism *O*, there is a functional relation *R* and a cognitive counter type *T* such that *O* takes attitude *A*[*C*] just in case *O* is in relation *R* to a tokening of *T*.

Here “representations” = cognitive counters = “the things that (a) are the objects of computation and (b) are the things that covary with content in the realization of intentional states.”

I think that BCTM captures the heart of what cognitivist researchers are after with respect to both the *computational* and the *representational* sides of the computer paradigm. And it does so

without setting off any Reid/Husserl concerns by using words like ‘symbol’ or ‘meaning’ to describe cognitive counters. So if empirical researchers were to agree that all they mean by ‘representation’ is something like “cognitive counter”, they would be safe from attacks on the flank occupied by writers like Reid and Ayer and Searle and Sayre. This would, of course, leave plenty of room for discussion of what kinds of formal and causal properties cognitive counters would have to have to render them suitable for the realization of content. All it would prohibit would be invoking some additional property called “representing x ” that was supposed to do some additional work beyond what is done by BCTM supplemented by some further specification of cognitive counters.

5. LOCATING THE ISSUES

The same issues can be approached from the perspective of locating different positions on the notion of representation in relation to cognitive science. We might see a project in philosophical psychology like Fodor’s as standing at midstream. Fodor seems to want us to adopt a robust, semantically-pregnant notion of representation that is closely tied to, if not indeed identical with, the familiar notion of symbolic representation. There are at least three key reasons Fodor thinks this is the right way to look at representation:

- 1) He thinks it is forced upon us by the success of cognitivist theories in psychology.
- 2) He thinks it provides an explanation of the intentionality of mental states.
- 3) He thinks it provides a way of “vindicating” intentional psychology against charges of methodological and ontological impropriety.

To the left of this position are two groups of people who do not draw the same moral that Fodor draws from empirical research. The first group is made up of philosophers such as Stephen Stich, who hold to a “syntactic theory of mind.” Stich believes that recent cognitivist research in psychology may require the view that psychological processes are functionally describable and syntactically-driven, but not that they are manipulations of meaningful representations. Stich is also inclined to draw the stronger moral that we are entitled to dismiss *all* intentional states on the grounds that they are unnecessary theoretical posits. The second group to the left of Fodor’s

position is made up of empirical researchers such my one-time teacher Stephen Grossberg, who do not insist on anything like *symbolic* representation and find philosophical issues somewhat alien.

On the right hand side stands a different sort of philosopher, who is concerned with issues like the Reid/Husserl objection, or Searle's "derived intentionality," or homuncular regress arguments. This is really a variation on the Reid/Husserl concerns: the only semantically-rich notions of representation we are familiar with get us in a regressive tangle if we try to apply them to things in the mind. So if you want to account for intentionality by way of a rich notion of representation, you have to *articulate* a usage of that word that (a) avoids the regressive tangle, and (b) provides enough semantic "umph" to explain meaningfulness. I for one do not think that this has been done. Sometimes writers on the right, however, also tend to draw a stronger conclusion, to the effect that cognitivist/computational theories are somehow fundamentally wrong-headed, and this precisely because they depend upon the notions of representation and computation. But *this* conclusion *only* follows if you buy into the claim that theoretical research in cognitive science is committed to a semantically robust notion of representation.

Now I should very much like to get a more reliable canvass of how many researchers do, upon reflection, think they need something stronger than BCTM to provide theoretical underpinnings for their research. If I am right in thinking that, by and large, they will not, I think there is a way of playing both ends against the middle here and still being intentional realists. The essentials of my view are as follows: Psychology can attain mathematical and connective maturity without bothering about philosophical questions about the relationship between mentalistic and natural properties. Realization is good enough. Such a psychology would not need any vindication, thus avoiding eliminativist tendencies on the left. Psychology doesn't need a robust notion of representation to do this. BCTM is enough, and hence we avoid Searle's objections on the right. This doesn't give an account of intentionality, but it doesn't prohibit one either. The theoretical use of "representation" *might* point to natural properties that really explain meaningfulness. But if this is your view, you aren't entitled to say you have an account of intentionality *now*, and it's misleading at best to say that the account is "representational" if there

isn't a robust notion of representation doing any work in it. And *maybe* meaningfulness is just *fundamental*, and can't be explained at all. If you've got mathematical and connective maturity anyway, that wouldn't count against cognitivist psychology one bit.

6. CONCLUSION

Regardless of the significant heresies that I have expressed here, my conclusion is meant to be largely ecumenical and conciliatory: there really are different notions of representation in the air, and serious problems can arise if you fail to pay close attention to the differences. But there is a way of looking at representation as a fairly bland theoretical term. And this, I argue, is strong enough to serve for the empirical scientist while weak enough to avoid the Reid/Husserl objection. It is thoroughly consistent with intentional realism, but does not provide an account of intentionality for the mind. (Though it is in principle compatible with such an account as a supplement.) It is mainly certain philosophers who are concerned with the latter; and if they insist on having it out of a representational theory, it is here that the interests of philosophy and those of research diverge, as this project requires something out of a notion of representation that empirical researchers can do without.

LIST OF REFERENCES CITED

- Ayer, A.J. 1968. *The Origins of Pragmatism*. London: Macmillan.
- Cummins, Robert. 1989. *Meaning and Mental Representation*. Cambridge, Massachusetts: Bradford Books.
- Dennett, Daniel C. 1969. *Content and Consciousness*. London: Routledge & Kegan Paul.
- Fodor, Jerrold. 1975. *The Language of Thought*. New York: Thomas Crowell.
- Fodor, Jerry. 1981. *Representations*. Cambridge, Massachusetts: Bradford Books/MIT Press.
- Fodor, Jerry. 1987. *Psychosemantics*. Cambridge, Massachusetts: Bradford Books.
- Horst, Steven. 1990. *Symbols, Computation and Intentionality: A Critique of the Computational Theory of Mind*. Doctoral Dissertation.
- Husserl, Edmund. 1913. *Ideen au einer reinen Phänomenologie und phänomenologischen Philosophie*. The Hague: Nijhoff. English edition, *Ideas: General introduction to pure phenomenology*. Trans. W.R. Boyce Gibson. Collier Books.
- Judge, Brenda. 1985. *Thinking About Things*. Edingburgh: Scottish Academic Press.
- Lehrer, Kieth. 1989. "Conception without Representation—Justification without Inference: Reid's Theory," *Noûs* XXIII, number 2 (April, 1989), pages 145—154.
- Reid, Thomas. 1983. *Thomas Reid's Inquiry and Essays*. Edited by Ronald E. Beanblossom and Keith Lehrer. Indianapolis: Bobbs-Merrill.
- Sayre, Kenneth. 1986. "Intentionality and Information Processing: An Alternative Model for Cognitive Science" *Behavioral and Brain Sciences*, Volume 9, no. 1 (March, 1986):121–138.
- Searle, John. 1980. "Minds, Brains and Programs." *Behavioral and Brain Sciences* 3:417-424.
- Searle, John. 1984. *Minds, Brains and Science*. Cambridge, Massachusetts: Harvard University Press.

¹ This paper was originally presented at the Conference on Cognition and Representation held at SUNY Buffalo on April 3—5, 1992. This written version is substantially the same as what was presented there, except that certain materials that were deleted for presentation due to considerations of length have been restored, and the style has been changed in a few places where it seemed too conspicuously "oral." An earlier and longer version of this material was presented at the NEH Summer Seminar on Mental Representation held at the University of Arizona at Tucson, in July 1991. That earlier version was developed under the support of an NEH fellowship for that seminar.

² More recent writers such as Dennett have tended to emphasize the problem of homuncular regress. Reid and Husserl emphasize a different problem: the whole point of representational theories is to say how our access to extramental reality is mediated by an inner entity, the representation. But if

“representation” requires an interpreter to interpret A as standing for B, the interpreter must already have some cognitive access to B (the extramental object) in order to interpret A in this way. So the theory is an attempt to explain our access to extramental objects, yet we can only make sense of the theory on the assumption that we already such access before representation takes place; hence the theory is self-defeating since it presupposes what it sets out to explain.

³ I think that this is fairly clear, at least, in the case of the criticisms by Reid, Husserl, Ayer, Dennett and Judge. It is not so clear that this is how Searle and Sayre intended their more specific criticisms of symbolic representation in computers to be taken. A fuller discussion of this may be found in Horst [1990].

⁴ It is very interesting, for example, that Fodor [1981: pp. 26—27] seems to try to defend the need for embracing realism about symbolic representations by appealing to the reader’s intuitions about the need for “representations” in the sense of images.

⁵ There are, of course, very serious empirical questions about (a) whether the computational approach to the mind will succeed in either of these endeavors and (b) whether it will do so better than any competitors, such as the information-theoretic approach favored by Sayre, the PDP approach, or the eclectic mathematical approach of Grossberg. My point is merely that the computer paradigm seems to hold some promise for such maturation, not that this promise is likely to pan out.