Representations, Symbols, and Embodiment

Michael L. Anderson
Institute for Advanced Computer Studies
University of Maryland, College Park, MD 20742
(301) 405-1746
mikeoda@cs.umd.edu

February 19, 2003

I would like to begin by thanking Ron Chrisley for his interesting and detailed commentary.[1] There is much I find to agree with; indeed, it seems to me that his essay can largely be read, not as a *criticism* of the perspective offered in my own article[2], but rather as a useful and largely complementary alternate view. I am especially grateful for the insights expressed in the last section of the paper, to wit:

1 Embodied AI (EAI) can be understood, not just as a guide for building intelligent systems, but as providing a set of concepts and approaches which enrich our explanatory resources, thereby allowing us to better understand the systems we already build, and

2 If, as EAI holds, intelligence really *is* embodied, situated, and interactive in nature, then perhaps we need to take seriously the notion that understanding and evaluating these systems (and further developing the conceptual resources which will be necessary for scientific advances in this area) is likewise going to be a matter of actually interacting with them.

The first seems to me just right, while the second is extremely intriguing, and suggests important implications that ought to be spelled out. Perhaps Chrisley will do this for us in a later work; I would certainly welcome such a piece.

Still, there are a few places where we appear to disagree; some of these disagreements are merely apparent, but one or two are genuine. I will discuss each in turn.

[1] "Embodied Artificial Intelligence", this issue
[2] "Embodied Cognition: A Field Guide", this issue

1 Representations and embodiment

In an example of the first, merely apparent, sort of disagreement, Chrisley suggests that my (pragmatist) view of the meaning of symbols "seems to imply that if I act inappropriately towards something, I *ipso facto* cannot be thinking about that something, since appropriate behavior toward it is a requirement for representing it." (p 11) In defense of this interpretation, he cites my remarks to the effect that a commoner who sat on the King's throne must have misunderstood either the concept of "throne", or that of "chair". I suppose the example is misleading in various ways (after all, the person might understand these concepts perfectly well, and sit in the throne to make a subversive statement about monarchy), but I meant it to emphasize two things:

3 concepts have interrelations—hierarchical, inferential, and such—which matter to the ways and contexts in which they are deployed, and

4 a central role of concepts (reflected in their content) is in guiding behavior (this in addition to—not 'as opposed to'—organizing perceptual experience, although it should be noted that for the Embodied Cognition (EC) theorist, the role of a concept in organizing perceptual experience is best understood in terms of, or as a result of, its role in guiding action).

I don't see that these general observations, nor the specific remarks I make in "Embodied Cognition" commit me to a conceptual relativism on which "we could never be wrong" because we will always "either have said something true about something else, or . . .have said nothing at all" (p 11), nor to a theory of intentionality on which acting *appropriately* towards a thing is a condition of thinking about it. Chrisley has extrapolated to (and is disagreeing with) a position which is neither one I in fact hold, nor one to which an EC theorist is commited in virtue of accepting either or both of the points above. This is why I suggest the disagreement is merely apparent.

Still, it may be worth saying a little more about the shape an embodied theory of intentionality might take, and why such a theory wouldn't necessarily be vulnerable to criticisms of the sort Chrisley raises. To begin, we have to notice that there are two issues at play here, which need to be teased apart: the conditions which govern the acquisition of a concept (i.e. the same concept that the other members of one's linguistic community have) and the conditions which govern sucessful reference to (or the ability to think about) some given object.

Acquiring a concept—say, "chair", to stick with the familiar—is surely bound up with more than just being able to reliably detect chairs in the environment. The conditions for acquiring a particular concept may well involve putting it in the right inferential and hierarchical relations to other concepts, including hooking it up in the right ways with action and desire.[3]

[3] There are delicate issues here regarding how these conditions are treated. For the relevant debates about conceptual holism see, e.g. [5, 1, 2]

If we saw someone who was surprised to see chairs with tables, or who always sat on tables and ate off chairs, we might reasonably suspect that the person had not, in fact, aquired one (or both) of the relevant concepts.

This, in combination with (4), above, does indeed suggest that a condition for having acquired a concept might involve behaving appropriately; but it is better to say: behaving in certain rationally explicable ways towards instances of the objects falling under that concept. "Rationally explicable" will often (maybe even *usually*) be the same as "appropriate", but, as alluded to in the throne example above, not always. Evidence that a child has aquired the concept "breakable" might equally be gently handling a given vase (presumably what his parents hope) *or* smashing it on the ground (because watching it break would be fun). The reduction of "rationally explicable" to "appropriate" depends on the person in question sharing not just a conceptual repertoire, but also certain relevent desires.

No doubt there are delicate issues waiting in the details of any such account, but accepting the fact that exhibiting rationally explicable behavior might reasonably be a consideration in determining whether someone has acquired (or is currently deploying) a given concept (i.e. the same concept as the one you have in mind) surely commits one to neither an unacceptable conceptual holism, nor a pernicious cognitive relativism. Further—and this brings us to the separate issue of intentionality—none of this suggests that the manifestation of appropriate behavior, however carefully it is defined, is a condition on deploying the concept to sucessfully *refer to* or *think about* a given object.

For instance, although I myself believe that behavior is crucial to the ability to establish and maintain intentional connections—to the ability, that is, to think about individual objects [6]—the veridicality of those connections doesn't depend on the *appropriateness* of any behavioral interaction with the object(s) in question, but rather on the (continuing possibility of) behavioral interaction itself. Likewise, according to the Guidance Theory of Representation (introduced in [7], and to be worked out in detail in [8]) representational vehicles (such as concepts) are *representational* in virtue of the fact that they standardly provide guidance for taking action with respect to the entity represented.[4] The representational connection is dependent on the *guidance* of the behavior, not on the *appropriateness* of the behavior thereby guided—although, naturally, in a properly functioning representational system, *guidance* and *appropriate guidance* will generally coincide. Futhermore, on the Guidance Theory, representational error can be cashed out in terms of (and discovered and corrected in virtue of)

[4]Note this is not the same as making representation, reference, conceptual content, or grounding dependent on (a history of) causal interaction with the represented entity, as Chrisley suggests (p 10), although of course behavioral interaction and causal interaction go hand in hand in the normal case. But, as the Swamp Man example shows (the Swamp Man example, for those who are not familiar, goes roughly like this: imagine an *exact duplicate* of yourself forming by chance from a huge cloud of swamp gas. Will this being's thoughts about, say, Oxford University, actually *be about* Oxford University, as yours are? [3]) a history of actual causal or behavioral interaction with an entity is not necessary to make a mental token representational (at least, it shows this if one accepts, as I do, that except perhaps for some special cases Swamp Man's mental repertoire supports genuine intentional connections to the world). Rather, what makes a mental token representational is the fact that it would be used by Swamp Man to guide his behavior towards the entity in question.

failure of action. This notion that the world provides epistemic friction through the medium of action, sufficient to limit, guide, and correct our representations, is the centerpiece of [6], and will be worked out in much greater detail in [8].

2 Symbols and embodiment

Because Chrisley misconstrued the EC (or perhaps just my) view of the nature of representation, the above is perhaps no more than a distraction, although I hope an interesting one. Much more central to the topic of these essays—and a point of apparently genuine contention between Chrisley and I—is Chrisley's claim that I go wrong in identifying the issue of grounding as the central, defining, and *unifying* theme of EC research, for this "places an emphasis on representational content, as if EAI agreed with GOFAI that that is where the action is." (p 12) The easiest answer to this is to point out (as Chrisley himself allows) that what I claimed is that EC is organized around the *physical grounding project*, which is something much broader than symbol grounding.[5] But Chrisley is right to notice that I seem to come back to symbols often, perhaps more often than is seemly for someone commited to (explaining) the principles of EC.

The reason for this is simple: EC is committed, ultimately, to explaining human intelligence, to discovering the underlying mechanisms of complex, intelligent behavior, and I believe that this will not be possible in the end unless symbol use is part of that explanation. Symbol use is *far* from the whole story; there is a great deal that goes into intelligence, and some of what is now standardly explained in terms of the manipulation of symbols will probably turn out instead to be rooted in the operations of specialized sensory-motor systems. But the ability to use symbols—perhaps even the possession of a language of thought [4] and the mental flexibility this implies—is going to turn out to be a necessary part of the story, too. For an EC researcher who believes this (and surely I am not the only one) what emerges as crucial (and fabulously interesting) is understanding the relations between the lower level, older, specialized sensory-motor systems (of the sort by which many EC researchers are rightly enthused) and the structure, elements, and rules of operation of the more general, highly flexible, symbolic computational system we also seem to possess. My bet, for what it's worth, is that these are significantly intertwined, with bi-directional feedback and cooperation—that, for instance, some conceptual contents can be traced to specific sensory-motor systems, and some sensory-motor systems have been adapted to utilize some of the resources of (or at least be responsive to) more general conceptual systems. Section 3 of "Embodied Cognition" is meant, in part, to outline the kinds of explanations of intelligence (or its aspects) that result when one is attending to these interrelations, and to suggest that explanations like these are characteristic of EC and the physical grounding project.

[5] Rather than replay my definition and discussion of the physical grounding project, I refer the reader to "Embodied Cognition: A Field Guide", esp. sec 3

3 Embodied symbols

The final point of disagreement between Chrisley and I, somewhat related to the above, revolves around how to characterize the efforts by some researchers committed to traditional AI paradigms to claim EC concerns and methods as their own. In order to do this, I contend, they end up over-generalizing the term 'symbol' to cover all sorts of states: conscious, unconscious, conceptual, non-conceptual, neural, spinal, cerebral and bodily. I wrap up the relevant section by writing:

> How will the various items in their grab-bag labeled 'representations' be related to one another? How can a conscious, explicit representation encode information about or from an unconscious, inexplicit one? Under what circumstances might such a relation be needed? What does it mean for a representation—say, a motor representation of an action—to be encoded *in the body* rather than in the head? (p 33)

Chrisley rightly zings me:

> One cannot simultaneously lambaste GOFAI for being Cartesian in that it maintains a strong multiple-realisation thesis and an autonomy of the mental from the physical, while simultaneously claiming that its Cognitivism assumes that the symbols of mental processing are specifically cerebral, and not merely neural, bodily, or worldly. If... their view is the one that implies that properly strung together beer cans can realize thinking, then surely they have no strong claim as to what physical stuff underlies *human* mentality. They certainly won't be baffled by such questions as 'What does it mean for a representation... to be encoded *in the body* rather than in the head'. If one understands how bodystuff in the head can instantiate or implement ("encode") a representation, one *ipso facto* understands how bodystuff not in the head might do so. (p 14)

Of course, Chrisley is right: in the most general sense, it is easy to see how a representation can be encoded by a (any) body. But let me use this opportunity to emphasize that one of the problems that plagues discussion of these issues is inconsistency in the use of such terms as 'symbol' and 'representation'—inconsistency both between different researchers, and by individual researchers. I am both guilty of, and a victim of, such inconsistency here. In the discussion to which I was objecting, 'representation' was being used interchangeably with 'symbol', but in the spirit of the physical symbol system hypothesis, a symbol isn't just any representation (anything that designates or denotes) but canonically one that plays a very specific role, and is subject to a particular set of transformations/operations. While it may still be easy to see how a (any) given body *could* encode symbols in this sense, it seems clear that this is in fact the sort of thing that happens only in the heads of only some bodies. The contribution to intelligent behavior of these special bodies' other parts (aspects)—and EC

maintains that this contribution is significant—will, and *should*, be difficult to characterize in terms of the physical symbol system hypothesis, even when that contribution involves representation more generally, but especially when it does not. Not everything is a physical symbol system (PSS), nor is every part (or sub-part) of something which *is* a PSS *itself* a PSS. Likewise, and more to the current point, I think it is misleading to suggest that every representation is a symbol, and every system which employs representations is for that reason a PSS.

It is becoming ever clearer that not everything relevant to—nor even everything crucial to—intelligent behavior is captured in the PSS hypothesis. The mind is in, and of, the body, and this means far more than the simple fact that the mind is physically instantiated; the point instead is that mindedness is a property of whole organisms. This perspective—which is, I think the better one—can help open us to a broad re-consideration of what, in the organism, might constitute a contribution to its intelligence. 'Tis a consummation devoutly to be wished.

4 Acknowledgments

Thanks are due to Tony Cohn and Don Perlis, Review Editors for *Artificial Intelligence*, for arranging this helpful exchange of ideas about Embodied Cognition, and to Gregg Rosenberg for commenting on earlier drafts of the current essay. This work was supported in part by grants from AFOSR and ONR.

References

[1] Ned Block. An argument for holism. *Proceedings of the Aristotelian Society*, XCIV:151–169, 1994.

[2] Ned Block. Holism, mental and semantic. In *Routledge Encyclopedia of Philosophy*. Routledge Press, 2003.

[3] Donald Davidson. Knowing one's own mind. *Proceedings of the American Philosophical Association*, 60:441–58, 1987.

[4] Jerry Fodor. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press, Cambridge, MA, 1987.

[5] Jerry Fodor and Ernest LePore. *Holism: A Shoppers' Guide*. Oxford University Press, Oxford, 1992.

[6] Michael O'Donovan-Anderson. *Content and Comportment: On Embodiment and the Epistemic Availability of the World*. Rowman and Littlefield, Lanham, MD, 1997.

[7] Gregg Rosenberg. *A Place For Consciousness: The Theory of Natural Individuals*. Oxford University Press, Oxford, 2003.

[8] Gregg Rosenberg and Michael L. Anderson. Content and action, *in preparation*.