Symbolism, Embodied Cognition, and the Broader Debate[*]

Lawrence Shapiro

University of Wisconsin – Madison

This actually happened at the dinner table recently.

Me, to wife Athena: How's that new bar on King St.?

Athena: Too much attitude.

Sophia, our eight-year old: What does that mean?

Pause.

Me: Means that they're kind of snobby.

Sophia: What does that mean?

Me: Means that that they have too much attitude.

Sophia: Oh, like that helps.

Sophia's disappointment in me, at least in this case, is well-deserved. Being told that the term 'X' means the same as the term 'Y' will not help anyone to understand what 'X' means unless they already understand what 'Y' means. I hope that this point is not controversial. Neither, I hope, is the following claim: Being told that the term 'X' means the same as the term 'Y', and that the term 'Y' means the same as the term 'Z', will not help anyone to understand what 'X' means if one does not understand the meanings of either 'Y' or 'Z'.

[*] For useful discussion I would like to thank the participants in the Garachico Workshop, as well as Ken Aizawa, Juan Comesaña, Dan Hausman, Carolina Sartorio, and Elliott Sober. Special thanks to Art Glenberg and Toni Gomila for comments that led to an extensive re-working of some of the ideas in this chapter.

Notice that the claims above need not be construed as *a priori*. Testing them would be easy enough. I could tell Sophia that being snobby means putting on airs. Though I did not actually say this to her, I hypothesize that she would then know nothing more about what it means to have too much attitude than she does presently. This makes the following induction very tempting: one cannot understand the meaning of a word or, more generally, a symbol, if one's only access to its meaning is via other symbols whose meanings are not understood.

Just to put a name on a target, I will call the view that a symbol can become meaningful to an interpreter solely in virtue of knowing its relations to other symbols whose meanings are not known, the *symbolist's folly*. Thus, if one were to assert that Sophia could eventually understand what "having too much attitude" means solely in virtue of her knowing that it means the same as other expressions she does not understand, or in virtue of her knowing that people to whom the predicate "has too much attitude" applies are often people to whom the predicate "is condescending" (a term with which she is unfamiliar) also applies, but to whom the predicate "is crunchy" (an expression with which she is unfamiliar) does not typically apply, and so on, then one would be engaged in the symbolist's folly.

But does anyone really believe in the symbolist's folly? Apparently some do, and at least one motivation among those preaching embodied cognition is to provide an account of meaning that avoids the symbolist's folly. For instance, Glenberg, De Vega, and Graesser, in their introduction to this volume, suggest that embodied approaches might provide the answer to the "what else?" question regarding how amodal symbols might gain meaning. Glenberg especially (e.g., Glenberg & Kaschak 2002) has been pursuing embodied approaches to cognition in an effort to circumvent the so-called grounding problem, i.e. the problem about how symbols

acquire their meaning. Barsalou too has looked to embodied cognition as a solution to the grounding problem, proposing perceptual symbols as a way out (1999, p. 614).

I think Searle's Chinese Room argument (1980), under a fairly standard interpretation, can be taken as directed against the symbolist's folly. The Chinese Room Argument makes acute the symbol grounding problem. However, I shall argue that the symbolist's folly is largely a red herring. There are reasons to pursue research in embodied cognition, but these are independent of the need to respond to the symbolist's folly. As for Searle's Chinese Room, there is not much beyond a trivial moral that we can draw from it.

Here's the plan. I will start with a discussion of Searle's Chinese Room, showing how it is supposed to render obvious the folly inherent in the symbolist's folly. However, I'll argue, the Chinese Room does *not* show that symbolic approaches, i.e. computational approaches, to cognition are deficient. That is, the Chinese Room contains nothing that need worry a good old-fashioned symbolist. But, I then argue, the same resources a symbolist can marshal in response to Searle work as well to deflect the charge that one must turn to embodied cognition to avoid the symbolist's folly. Here my focus will be on Barsalou's important work on perceptual symbol systems. In the final sections of the paper I suggest a positive motivation for embodied cognition. Embodied cognition proposes new ways to conceive of the role of representation in cognition, where these new ways promise to provide better and simpler accounts of cognition than those that insist on traditional conceptions of representation. In this sense, I suggest that the debate between symbolism and embodied cognition is an instance of a more general debate within philosophy of science over the justification for theoretical posits.

1. Into the Room

There sits Searle in the Chinese Room. Slips of paper containing Chinese symbols enter through a slot in the door. Searle picks up the slips, consults a book of rules that specifies which Chinese symbols to write in response, and out goes his "answer" to the initial "question." The scare quotes are necessary because Searle does not know he has provided an answer. He does not know that the original slip of paper contained a question. In fact, Searle may not even know that the symbols he is manipulating are Chinese. As far as Searle is concerned, he's playing a rather tedious game: here are some marks on a piece of paper. Using your rule book, turn these marks into other marks. This, as far as Searle knows, is all that's going on.

To Tak Jun, who sits outside the room, something very different is happening. Tak Jun writes down questions, feeds them into the room, and a while later out come answers. Tak Jun writes, "Why did the Yangtze overflow its banks this morning?" From the box appears an answer, "Because of yesterday's heavy rains." Tak Jun assumes that the box understands Chinese. There's no distinguishing it from a person who understands Chinese, so this is not an unreasonable thing to believe.

Searle supposes that we will all agree that he doesn't understand Chinese. He's in worse shape than Sophia. Sophia's stuck trying to understand one expression in terms of another that she doesn't understand. But she at least knows that what she doesn't understand are words. Words, of course, are symbols. Sophia is trying to learn the meaning of symbols. But Searle may not even know that the marks on the slips of paper are symbols. He may not know that the marks have any meaning at all.

I think we should agree that Searle does not understand Chinese. If unsure, we could ask him. He would tell us that he doesn't. We could ask Tak Jun to insult Searle, who's unlikely to

be happy about this. If Searle remains calm, this is all the evidence we would need. Of course Searle doesn't understand Chinese.

2. Opening the Door

It is not my intention to evaluate the significance of the Chinese Room thought experiment for Strong AI. My concern instead is with diagnosing why Searle does not understand Chinese. I think the reasons are fairly obvious but I would like to place them in the context of a particular kind of theory of meaning. However, to do this, I am going to take Searle out of the Chinese Room and put him in the Greek Room. I do this because Chinese is Greek to me, but Greek is not.

The theory of meaning I have in mind is usually called a *causal theory of meaning*. Causal theories of meaning have been a mainstay in philosophy of mind since my colleague Dennis Stampe's seminal "Toward a Causal Theory of Linguistic Representation" (1977). Others, notably Fred Dretske (1981) and Jerry Fodor (1987), have developed causal theories of meaning in their own ways, but they follow Stampe in analyzing meaning in terms of a representation's causes. The basic idea is that a representation is of X (that is, it means X) when the representation, under certain conditions – what Stampe calls "fidelity conditions" – would be present *because* X is present. The fidelity conditions are those that, as Stampe puts it, govern the production of the representation. When appropriate fidelity conditions are present, it is "reasonable to accept" that the presence of a representation of X implies the presence of X. Of course, this is only the barest sketch of the account; Stampe and others have much more to say about how to understand fidelity conditions so that they do not result in the representation of the wrong objects, or so that representation can sometimes be vacuous. For my purposes, I mention the causal theory of meaning because it has a compelling story to tell about how symbols acquire

meaning. According to this story, my thought ATHENS means Athens because, in the context of certain conditions, it is reasonable to suppose that the presence of Athens would cause the thought ATHENS. Obviously, the existence of an external environment is necessary for the meaningfulness of many of our thoughts. Thoughts acquire their meanings not through their associations with uninterpreted symbols, as the symbolist's folly would have it, but through their connections to the world or, often, their connections to other symbols that are connected to the world.

Consider now Searle's thought that Athens is in Greece and Thalia's thought that Athens is in Greece. Searle and Thalia have the same thought, but they express it differently. Searle uses the word 'Athens' to express his thought. Thalia uses the word 'Αθηνα'. Searle might also think "'Athens' has six letters." Thalia would think, "'Αθηνα' has five letters." Searle and Thalia have different thoughts because their thoughts are no longer about the same thing. Searle's thought is about the English word 'Athens' and Thalia's thought is about the Greek word 'Αθηνα'.

Suppose now that Searle knows no Greek and we place him in the Greek Room. Thalia writes down a question on a slip of paper which, translated into English, reads "Is Athens the capital of Greece?" Searle receives the slip of paper, consults his English rules, writes down some Greek symbols that are as meaningless to him as the Chinese symbols are, and pokes the paper back through the hole in the door. Thalia reads the answer which, translated into English, is "Yes. Athens has been the capital of Greece for a very long time."

Why does Thalia take herself to be having a conversation whereas Searle does not? The answer, I submit, is that the conditions that must be in place for 'Αθηνα' to mean Athens are in place for Thalia, but are not in place for Searle. For instance, Thalia belongs to a community in

which utterances of 'Αθηνα' are caused by speech production mechanisms that in turn cause these utterances as a result of individuals' intentions to refer to Athens (Stampe 1977, pp. 52-3). Ignorant of these facts, Searle does not know that 'Αθηνα' means Athens. How could he? Consequently, Searle does not take himself to be having a conversation with the person slipping notes through the hole in the door.

The philosopher's distinction between use and mention provides another way to see how Searle and Thalia differ. Thalia can *use* the word 'Αθηνα', as she does when saying, in Greek, that Athens is in Greece. She can also *mention* the word, as she does when she says "'Αθηνα' has five letters." Searle, on the other hand, can only mention Greek words. He can say and think "'Αθηνα' has five letters", but he cannot use the word 'Αθηνα' to say anything about Athens. In order to use Greek words, Searle needs to know their meanings, but he does not know their meanings because he knows nothing about the connections that Greek words bear to the world. For Searle, these words bear connections only to each other. Thalia can see *through* the Greek words, but Searle can never see *past* them.

But, if this diagnosis of why Searle does not understand Greek is correct, there is an obvious solution to his problem. One simply has to open the door to the Greek Room and let Searle explore his Greek environment. In time, Searle will learn which words are appropriate to utter in which contexts. He'll learn to think ΣΚΥΛΟ when he sees a dog, to think ΣΠΙΤΙ when he sees a house, and so on. He'll understand Greek when he becomes attuned to the connections between Greek words, intentions, and objects.

3. Symbolism and Embodiment

The lesson to draw from the Chinese Room is, after all, not very interesting. Indeed, finding a non-trivial way to state the lesson is difficult. The lesson seems to be this: symbols do

not acquire meanings merely through formal transformations. The symbolist's folly is folly. *Of course* if one does not know the meanings of symbol X or symbol Y then knowing how X and Y are connected is not going to help. Sophia knew that. One needs to understand (tacitly or otherwise) how symbols are connected to the world in order to know what they mean.

Famously, Searle claims that thought cannot be just symbol manipulation, or symbol manipulation is, by itself, not sufficient for thought. This too seems right. If the operations of a computer involve only uninterpreted symbols, then a computer could never understand what it was "thinking." But this observation does not imply that the manipulation of *interpreted* symbols is not sufficient for thought. This point is significant for those who see embodied cognition as an answer to what ails symbolism, for consider two complaints that seem to motivate to some extent the embodied cognition turn:

1. Symbol manipulation on its own cannot produce understanding;
2. Symbols acquire their meaning only through embodiment.

My discussion of the Chinese Room, I hope, makes clear that one can accept the first complaint without giving up a symbolist conception of cognition. That is, a symbolist can agree that the manipulation of mere, uninterpreted symbols is not sufficient for understanding. In addition to symbol manipulation, there must as well be a semantics that assigns meanings to the symbols.[1]

---

[1] I do not wish to claim that the causal theory of meaning will, all by itself, suffice for understanding. Talk of understanding is vague. Understanding might involve a phenomenological component (e.g., the "Ahh!" feeling), and surely the causal theory of meaning has nothing to say about that. However, I would deny that understanding requires that one knows *that one knows* the meaning of the words she uses.

This easy answer to the first complaint might seem susceptible to the second complaint. Perhaps symbol manipulation does not create meaning because manipulation is in some sense insulated from the relations between mind, body, and world from which meaning emerges. Yet, from the perspective of the causal theory of meaning I described above, embodiment is irrelevant. Tree rings represent, or mean, a tree's age. Smoke means fire. The mercury's having a particular level means that it is 68 degrees. All of this is possible without it being true that trees, smoke, or mercury are parts of an embodied system. What matters is that symbols are connected in the right way to those things they represent. If embodiment provides such a way, all the better; but this would mean only that embodiment is one way to make the right connections.

It is worth exploring this idea a bit further in the context of Barsalou's theory of perceptual symbols, which, he believes, provides a solution to the grounding problem that is unavailable to the computational theorist. As we will see, the perceptual symbols that Barsalou introduces differ from traditional computational symbols less than he thinks. In essence, perceptual symbols are like traditional symbols except for having a grounding relation "built in".

At the core of Barsalou's theory is a distinction he draws between amodal and modal symbols. Words are paradigms of amodal symbols. Words are (typically) not similar to those things they represent. 'Car,' for instance, does not resemble a car: the word does not have four wheels, although the object does. Moreover, the connection between 'car' and cars is arbitrary in the sense that the vehicle might have been called by a different name. Indeed, names for a car differ across languages. In short, there is nothing like a law of nature that ties a car to a particular name.

On standard computational approaches to cognition, cognitive processes construct amodal symbols. Processing goes in stages, roughly as follows. First, stimulation hits the surfaces of sense organs. The perceptual system of which the sensory receptors are a part transduce this stimulation into a neural code. There is then a second stage of transduction in which a cognitive system converts perceptual neural symbols into non-perceptual, amodal symbols. Thus, for instance, perceiving a car requires first the stimulation of cells in the retina. The visual system then constructs a representation of the car. The properties of this visual representation correspond to information about the shape, size, and color of the car. This representation is then delivered to cognitive systems that utilize frames, or feature lists, or other representational structures, in order to categorize the visual representation as being of a car. Thus, the cognitive system takes as input a visual representation and produces as output the judgment CAR. Importantly, this judgment consists in the construction of a word-like symbol that bears no more similarity to the car, or to the visual representation of the car, then the word 'car' bears to an actual car. Because there is no similarity between this final neural symbol and the visual properties that are encoded, the tie between the two is arbitrary. There is nothing that requires that the car be represented with the symbol CAR rather than some other symbol (Barsalou, 1999: 578-9).

Reinforcing this amodal conception of mental representation is Fodor's well-received defense of a language of thought. Fodor (1975, 1987) has argued that mental representations are, in many important respects, language-like. The beliefs, desires, and other attitudes we have are relations to language-like representations. Thus, my belief that Madison sits on an isthmus consists of a believing relation toward a language-like representation of the proposition that Madison sits on an isthmus. Because the representation is linguistic, it is compositional and thus

capable of participating in the systematic and productive processes that are hallmarks of both thought and language. Crucially, adoption of the language of thought hypothesis seems to force a commitment to the same kind of amodality of mental representations that is true of linguistic representations (but see below).

Barsalou rejects this two-stage conception of cognition, proposing instead that there is no need for a separate processing stage in which perceptual representations are converted into amodal cognitive representations. "Cognition is inherently perceptual," he argues, "sharing systems with perception at both the cognitive and neural levels" (1999: 577). In place of the amodal renderings of cognitive systems, Barsalou posits *perceptual symbols*. These symbols are reconstructions of the representations that were present in perceptual systems when these systems were initially transducing stimulation into a neural code. Barsalou calls these symbols modal, for they "are represented in the same systems as the perceptual states that produced them" (1999: 578). For instance, the perception of a car involves the representation of visual, auditory, haptic, and proprioceptive information. Storage of these representations in the form of perceptual symbols underwrites future cognitive activities such as remembering a car, making inferences about cars, talking about cars, and so on. In the course of these activities, a representation of a car is produced anew in the very same perceptual systems that made possible the initial perception of cars.

Because my interest is in whether perceptual symbols offer a response to the grounding problem that is not available to, or is better than, the response I have already provided above to the computational theorist, I shall not discuss the evidence Barsalou collects in support of perceptual symbols. In short, my question is this: Are perceptual symbols grounded in a way not available to traditional symbols? Barsalou clearly thinks that the answer to this question is "yes".

However, I think Barsalou's insistence that perceptual symbols have an advantage over computational symbols rests on a confusion. The confusion resides in mistaking the modality of perceptual symbols with facts about their causal etiology. Having made this mistake, Barsalou then attributes the groundedness of perceptual symbols to their modality, when in fact perceptual symbols owe their grounding to the same causal facts in virtue of which computational symbols may be grounded. Let me explain.

Consider first that perceptual symbols are modal only in the sense that they are present in a perceptual system that is dedicated to processing information of a particular mode (e.g. visual information, or auditory, or haptic). They are not modal in the sense that a perceptual symbol that represents red is in fact red, or one that represents vibration is in fact vibrating. Barsalou is rightly quite careful about this: perceptual symbols "are modal because they are represented in the same systems as the perceptual states that produced them" (1999: 578). But Barsalou also tells us that perceptual symbols are analogical, because the "structure of a perceptual symbol corresponds, at least somewhat, to the perceptual state that produced it" (1999: 578). However, he elaborates in a footnote (1999: 608), the structure of a perceptual symbol does not, typically, correspond to the structure of an object in the world. To claim that it did, presumably, would be to commit the error implicit in the idea that these representations *are* red or *do* vibrate. Instead, the correspondence exists only between features of the perceptual symbol and the perceptual *representation* of which it is a record.

Assuming that this is the extent to which perceptual symbols are modal and analogical, the question arises whether perceptual symbols differ from computational symbols in ways that are relevant to the symbol grounding problem. Like computational symbols, perceptual symbols are not similar to those things they represent. They may be structurally similar to the perceptual

representations from which they are descended, but they do not *represent* these perceptual representations. Thoughts, inferences, and knowledge about cars are supposed to be about *cars*, not about representations of cars. And, as Barsalou cautions in the footnote I cited, perceptual symbols are not structurally similar to objects in the world.

But if perceptual symbols and computational symbols do not differ with respect to being analogical, perhaps their difference rests on the fact that computational symbols are arbitrary. Here too, however, amodal symbols of the sort that computational theories of cognition posit seem no more arbitrary than perceptual symbols. "Let this salt shaker stand for the quarterback," one might say when describing an interesting play observed on the grid iron earlier in the day. The salt shaker represents the quarterback, but this is just stipulative or accidental. The pepper shaker would have done as well. Barsalou is right that perceptual symbols are not arbitrary in this way. Given how the brain works, presumably there is a causal story to tell about why a given perceptual representation will leave the record that it does, will result in the particular perceptual symbol that it does. However, and this is the crucial point, the computational symbols in a language of thought are not arbitrary for the same reason. There is a causal process that leads from an object in the world to its neural representation. This is true whether the representations over which cognitive processes range are perceptual, as Barsalou thinks, or are derived from perceptual representations through further processes of transduction. Because there is a law-governed process leading from objects to their neural representations, it makes no sense to suggest that some other neural representation of, e.g., a car, would have "done as well." Some other neural representation would not bear the correct causal relationship to cars, and would thus constitute a delusion of some sort ("you tell me that it's a car in front of me, but I'm seeing a giraffe!").

In light of this point, an objection that Barsalou raises to amodal symbols seems curious. Amodal symbols, but not perceptual symbols, confront the symbol grounding problem because we have no account of "how amodal symbols become mapped back to perceptual states and entities in the world"(1999: 580). However, we do have an account, and it is one that Barsalou himself describes. Amodal symbols are caused by perceptual representations which are in turn caused by stimulation of the sensory organs. If Barsalou's account of how perceptual symbols are grounded is correct, then this account of grounding amodal symbols, which appears simply to involve an additional causal step, ought to be correct as well. If this account of how amodal symbols are grounded is *incorrect*, then it is hard to see why Barsalou's account of grounding perceptual symbols should not also be incorrect.

I have argued that perceptual symbols and computational symbols are not distinct with respect to being analogical or arbitrary. Neither resemble the objects that they represent. Because both are determined by causal processes that begin with stimulation, neither are accidentally connected to their objects. Despite these similarities, however, perceptual symbols may differ from traditional symbols with regard to the amount of processing they require. In particular, the processing of amodal symbols reveals a kind of *indifference* to etiology that the processing of modal symbols does not. There is nothing in the syntax of an amodal symbol to mark its content as visual, or auditory, or haptic, and so on. Consequently, theories of cognition that restrict themselves to amodal symbol manipulation might require more sophisticated and complex layers of processing.

In contrast, consider that because modal perceptual symbols declare their content as being of a particular perceptual kind, their processing is constrained insofar as visual content will be processed by the visual system, auditory content by the auditory system, and so on. This

feature of perceptual symbols presumably brings with it gains in processing efficiency. Processing of amodal symbols would seem to require an extra step, *viz*., the identification of the symbol as having the mode of perceptual content that it does. In contrast, the fact that perceptual symbols are "built" for particular kinds of perceptual processing eliminates the need for this additional step. We will see below that this tendency to conceive of representational features of the mind as built to reduce processing demands is a hallmark of embodied cognition.

If my comments so far are on track, then there is clearly a *wrong* reason to reject computational theories of cognition, and so to motivate the development of alternative theories of cognition. The following argument exemplifies this wrong way:

1. Computational theories of cognition assume that symbols acquire their meanings by their associations only with other symbols.
2. Symbols cannot acquire meaning through association only with other symbols.
3. Therefore, computational theories of cognition are inadequate.

This argument is not sound because the first premise is false. The question whether cognition is computational is orthogonal to the question about how symbols acquire their meaning. For this reason, one can believe that cognition involves symbol manipulation but also believe that the second premise is true, i.e. that symbol manipulation all by itself does not generate meaning. One can be a computationalist without committing the symbolist's folly.

4. What's Embodiment About?

So, it is possible to be a computational symbolist without falling victim to Searle's Chinese Room. The Chinese Room is little more than a dressed up tautology: meaningless symbols don't mean anything. I conclude that if one pursues embodied cognition because one thinks that computationalism entails the symbolist's folly, then one has made a mistake. But if it

is not because of the symbolist's folly that one should seek a motivation for embodied cognition, then what? The motivation, I submit, becomes clearer on recognition of the distinction between two claims. The first is that an association between symbols is by itself sufficient to provide them with meaning. The second is that cognition is symbolic in the traditional computational sense. As I argued in my discussion of Searle, one can accept the second of these claims without accepting the first. There are various theories of meaning, and a commitment to a symbolic view of cognition does not entail a commitment to the symbolist's folly.

As I see the significance of embodied cognition, among its important aims is to provide an alternative to the second claim – that cognition is symbolic in the traditional computational sense. We saw this aim in Barsalou's work, which attempts to build a case in favor of perceptual symbols in contrast to traditional computational symbols. Perceptual symbols differ from traditional symbols in announcing their etiology, i.e. in declaring themselves to have a particular sensory origin. This feature of perceptual symbols short cuts the extra steps involved in processing amodal symbols. Moreover, perceptual symbols incorporate bodily facts into their syntax. Their syntax marks them as visual, or auditory, or haptic. But researchers in embodied cognition have offered approaches to cognition more radical than Barsalou's – approaches that attempt to eschew altogether the need for symbolic representation. These approaches, I shall now argue, do not succeed in showing that cognition can do without symbolic representation. However, like Barsalou's more tempered view, they do suggest an account of representation that, in virtue of embodiment, requires less or different kinds of processing than traditional symbolists have supposed.

The prominence of work by J.J. Gibson, Rodney Brooks, and Esther Thelen in introductory discussions of embodied cognition suggest that embodied cognition has invested

considerable interest in seeking to account for cognition in a way that does way with symbols altogether. Gibson, Brooks, and Thelen especially encourage a view I shall call *symbol antagonism*. Let's first see how symbol antagonism receives expression in Gibson.

A traditional approach to the problem of vision, such as David Marr's (1982), describes vision as a kind of "inverse optics." The three dimensional world causes on the retina a two-dimensional representation. Vision is then a matter of converting this two-dimensional representation into a perception of a three-dimensional world. On this conception, vision is a process that starts and ends with representation. It is a process acting on symbols for the purpose of producing new symbols. This is the conception of vision that Gibson rejects.

In its place, Gibson (1979) contends that representations of the world are not required for vision because the ambient light bouncing back and forth between surfaces in our environment has an information-laden structure. This structure becomes apparent from the point of view of a perceiver, at which time the ambient light becomes an ambient optic array for this perceiver. Merely by moving through the environment, the perceiver is able to sample information contained in the ambient optic array. Information comes in the form of invariants in the optic array. Thus, for instance, surfaces in the environment will create discontinuities in the optic array that specify edges. These discontinuities remain invariant as the observer moves around the surface, providing information about the shape and composition of objects in the environment.

The novelty in Gibson's theory of perception lies in this shift from inverse optics. No longer is the task of the perceiver one of extracting a three-dimensional representation of the world from a two-dimensional one. Rather, the perceiver's task is to explore his environment, uncovering invariants in the ambient optic array that specify features of the world. The

information is in the world and ready to be "picked up." There is, therefore, no need to construct a representation to play the role that the information in the optic ambient array already plays. For Gibson, perception is direct. It involves no symbolic middlemen.

One way to approach Rodney Brooks' (1991) work in robotics is to take it as implementing the Gibsonian theory of perception. Brooks' "creatures" rely on what Brooks calls a subsumption architecture. This architecture comprises "layers" that perform very simple tasks. Many of these tasks involve little more than immediate reactions to the detection of simple features of the environment. Thus, there might be a layer whose job is to stop the creature when it senses an object in front of it. Another layer might have the job of maintaining sensory contact with an object to the creature's left or right side. Another layer might simply direct the creature to move straight ahead. Making Brooks' work exciting is the behavior his creatures exhibit simply through the summation of the multiple layers' behaviors. Creatures navigate through a busy environment, avoiding objects, changing courses, as if they were consulting a map. In fact, however, creatures do not construct map-like representations of their environments – of landmarks, of hazards, of hallways. There is no need for such a map, Brooks asserts, because his creatures are built to exploit the structure in their environment. A model of the world is not necessary, Brooks explains, because the world is its own best model.

Clearly, Brooks' work embodies Gibsonian principles. Just as Gibson urges the rejection of representations in favor of the detection or "picking up" of information in the environment, so Brooks seeks to build robots that forsake rich representational content in exchange for the detection of informationally loaded variables. Instead of building a representation of an object in order to compute a course around it, a creature might just keep the object to its left as it moves forward. When it no longer senses the object, it can turn left again to continue on its course. In

this way, the creature *uses* the object in "figuring out" how to avoid it. It has no need for a representation of the object.

The third example of symbol antagonism I wish to discuss is Esther Thelen's (1995) study of motor development in infants. Thelen has shown that coordinated kicking or stepping behavior in infants is modeled better in terms of a dynamical system than it is in terms of painstakingly calculated neural signals. The large muscles in the legs act as springs under the control of gravity. Stepping and kicking movements, rather than under the direction of a cognitive plan, are nothing more than the oscillatory movements of a spring. Similarly, the coordination involved in stepping, Thelen argues, does not involve any cognitive control, but results from the coupling between single-leg oscillators. In like manner, two pendula mounted on a single wall will, over time, synchronize their swing. The coordination in both cases is simply a product of unguided physical forces.

As with Gibson and Brooks, Thelen favors explanations that are antagonistic to the postulation of symbols. One can imagine her response to a symbolist who insists that an explanation of motor development must employ symbolic representations of trajectories, masses, and other variables that control motion. "That's all very nice," she might say, "but what's really going on is simply the synchronization of oscillators." To insist on a role for symbols would seem, from her perspective, simply gratuitous – an effort to fit the facts to the theory rather than the other way around.

The conclusion I draw from these brief descriptions of paradigm work is that embodied cognition distinguishes itself from symbolism through its promotion of a theory of mind that seeks to overturn traditional views of cognition as symbol manipulation. More specifically, where embodied cognition seems to depart from symbolism is in its preference for explanations

of cognition that tie symbols more directly to idiosyncratic facts about organisms or interactions between organisms and their environments. This is evident in Barsalou's theory of perceptual symbols, which holds that symbols are copies of perceptual states and undergo processing in the same perceptual systems in which the initial perceptual states were produced, thereby tying the notion of representation to the particularities of organisms' sensory equipment. This is also evident in the work of Gibson, Brooks, and Thelen, all of whom try to exchange representation-heavy accounts of cognition for ones that let interactions between organisms and their environments do the work traditionally assigned to amodal symbolic representations.

5. The Debate in the Context of Philosophy of Science

I have so far defended a negative and a positive thesis. The negative thesis is that symbolists are not guilty of the symbolist's folly. There are theories of meaning, such as the causal theory of meaning, that symbolists might pursue to explain why Chinese symbols have no meaning for the man in the Chinese room although Chinese symbols do have meaning for ordinary Chinese speakers. Accordingly, symbolists have a response to those embodied cognition researchers who attack symbolism for committing the symbolist's folly. The positive thesis is that embodied cognition distinguishes itself from traditional symbolic accounts of the mind through, in the case of researchers like Barsalou, an effort to reconceive the traditional notion of a symbol, or, more radically, in an effort to show that symbols play a much more anemic role in cognition than symbolists have claimed, and that instead cognition emerges from the actions of an organism in an informationally laden environment.

I want now to consider how the symbolist might respond to the more radical views of Gibson, Brooks, and Thelen. Doing so will reveal that the debate between embodied cognition theorists and traditional symbolists is an instance of a broader debate in philosophy of science

concerning realism. As such, future work in embodied cognition is likely to benefit from a philosopher of science's perspective.

The traditional symbolist is likely to look on the positive thesis of embodied cognition as a simple confusion. Perhaps, the symbolist might concede, there is more information in the environment than has been typically appreciated. Gibson's uncovering of invariances in the ambient optic array, Brooks' success building robots that spend less time planning and representing than traditional robots, Thelen's discovery that motor development does not require sophisticated neural representations, and so on, are all very interesting, the symbolist might agree. Nevertheless, the symbolist might insist, none of these discoveries impugns the view that cognition is symbolic.

The symbolist's apology would go like this. Whether there is little information in the environment or lots, the basic fact remains that information cannot move from the environment into the organism without an intervening representational step. It is all well and good to notice that the environment is chock-full of information, but this information can be of service to an organism only if the organism is able to respond to it; and responding to information is possible only by way of representing that information. What's out there is of no use if it remains out there; but in order to bring it in here – into the mind – it is necessary to represent it.

The symbolist is free as well to be suspicious of claims like Gibson's that organisms simply "pickup" information, or, as he will sometimes say, "resonate" to information. These are just metaphors, the symbolist might charge, that when cashed out suggest a commitment to representation. One can grant Gibson the idea of invariants in the ambient optic array, but there remains the question how organisms come to use these invariants. The invariants are of no value unless an organism has a means by which to detect them, but what is detection if not

representation? Likewise for Brooks' creatures. Perhaps they don't construct a map of their world, but they do represent objects in front of them, objects to the left or right, the command to go straight ahead, and other such simple things. Furthermore, perhaps rich motor plans are not necessary for motor development, as Thelen claims. Still, an infant's legs can move in synchrony only if able to respond to each other, and they can respond to each other only if there is some sort of representation of each leg's activity.

This symbolist response to the positive idea in embodied cognition raises many issues, and I cannot hope to resolve any of them in the present context. However, let me try to sort out some of them. First, the symbolist's claim that representation is always necessary to get information from "out there" to a place where an organism can use it is, in some cases, simply false; but in other cases it misses the more important lesson that embodied cognition teaches.

The claim is false when the explanandum is the coordination of infant stepping behavior. Assuming that Thelen's model of stepping behavior is correct, there is no need to attribute representations for the purpose of understanding how the infant's legs synchronize. The principles responsible for the legs' synchrony are the same as those that explain the synchrony of two pendula on a wall. The pendula clearly affect each other, but this is a case of response without representation. It's closer in spirit to the effect a strong wind has on a palm tree than it is to a computation. I think the symbolist should be prepared to admit that some behaviors that might once have seemed to fall within their province are in fact behaviors that are better explained non-symbolically.[2]

---

[2] I have significant reservations about whether Thelen's work should be relevant in this dispute. Perhaps we should not consider the kinds of behavior Thelen examines to be cognitive at all. What exactly is definitive of cognitive behavior?

The sense in which the symbolist response misses the point of embodied cognition becomes apparent when considering Gibson and Brooks' research. One might believe that Gibson and Brooks overstate their cases against representation yet still believe that their work marks a significant departure from symbolic explanations of cognition. On a charitable reading of their work, they are less concerned with whether some aspects of cognition are representational than they are with investigating the extent to which the active exploration of an environment can play the role that has traditionally been assigned to symbolic computation. Much of Gibson's work is concerned with discovering what invariants are present in light such that an organism's computational burdens might be reduced. Similarly, Brooks' creatures are attempts to maximize behavioral versatility while minimizing computational load.

We can now see the sense in which the symbolist's response is off target. Researchers who pursue embodied cognition are calling into question the need for traditional symbolic representation in cognitive explanation. Seen this way, embodied cognition is just the latest chapter in a debate that is very familiar in the broader context of philosophy of science. The debate concerns the justification of explanatory posits. Symbols, like atoms, forces, and other unobservables, will earn their keep insofar as they do the kind of work that these other unobservables have done. If the attribution of symbols is necessary for modeling cognitive processes, for predicting cognitive behavior, for building cognitive systems, and so on, then the symbolists will have done as much as any scientist can do to establish the existence of their theoretical entities. On the other hand, if embodied cognition researchers are able to explain cognitive behavior without assuming the existence of amodal, computational symbols, and if traditional symbolic explanations prove themselves to be as ill-founded as, e.g., explanations of

life that assume the presence of an élan vital, or of heat transfer that assume the existence of phlogiston, then so much the worse for symbolism.

Interestingly, Barsalou (1999) picks up on some of these broader issues. In commentary on Barsalou's theory of perceptual symbols, Murat Aydede (1999) raises an objection similar to one I raised earlier. Amodal symbols of the kind that the person in the Chinese room is manipulating *can* be grounded, Aydede assumes, given appropriate causal connections to the world. Thus, Aydede would presumably agree with me that Baralou and others attack a straw man when arguing that amodal symbols face a *special* kind of grounding problem. In response to this criticism, Barsalou seems to soften his stance. He claims that "Aydede notes that perceptual representations could indeed be a part of this [causal] sequence, yet he fails to consider the implication that potentially follows" (1999: 638). What follows, Barsalou argues, is that perceptual symbols might by themselves suffice for cognition – there is no need to go the extra step in positing the creation of amodal symbols from perceptual symbols. As Barsalou says, "why include an additional layer of amodal symbols in the causal sequence, especially given all the problems they face? Nothing in Aydede's commentary makes a case for their existence or necessity" (1999: 638).

Barsalou's response to Aydede makes clear that the success of embodied cognition will depend in part on whether embodied theories of symbols and symbolic processing can offer better and simpler accounts of cognition than those currently available. It is in this context that the conceptual tools of the philosopher of science grow in importance. What counts as better? What counts as simpler? Answers to these questions requires a critical study of the research that participants in this conference, as well as many others, eagerly pursue.

References

Aydede, M. (1999). "What Makes Perceptual Symbols Perceptual?," *Behavioral and Brain Sciences* 22: 610-611.

Barsalou, L. (1999). "Perceptual Symbol Systems," *Behavioral and Brain Sciences* 22: 577-609.

Brooks, R. (1991). "Intelligence Without Representation," *Artificial Intelligence* 47: 139-159.

Dretske, F. (1981). *Knowledge and the Flow of Information* (Cambridge: MIT Press).

Fodor, J. (1975). *The Language of Thought* (Cambridge: Harvard University Press).

Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge: MIT Press).

Gibson, J. (1979). *The Ecological Approach to Visual Perception* (New York: Houghton Mifflin).

Glenberg, A., De Vega, M., and Graesser, A. (XXXX). "Framing the Debate," in A. Glenberg and M. De Vega (eds). *Symbols, Embodiment, and Meaning* (Oxford: Oxford University Press: xx-xx).

Glenberg, A. M. & Kaschak, M. P. (2002). "Grounding language in action," *Psychonomic Bulletin & Review* 9: 558-565.

Marr, D. (1984). *Vision: A Computational Investigation in the Human Representation and Processing of Visual Information* (San Francisco: W.H. Freeman).

Searle, J. (1980). "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3: 417-424.

Stampe, D. (1977). "Towards a Causal Theory of Linguistic Representation," in P. French, T. Uehling, and H. Wettstein (eds). *Midwest Studies in Philosophy 2: Contemporary Perspectives in the Philosophy of Language* (Minneapolis: University Press: 42-63).

Thelen, E. (1995). "Motor development: A New Synthesis," *American Psychologist* 50: 79-95.

Transcript

Graesser: In the first draft of the framing chapter, I don't know whether it was you Manolo or Art who wanted to keep representation out of there, that it was bringing up a bunch of problems. And I wanted it in there. But here's the question. Is it really true that the embodied position is denying representation? It was my understanding that it allowed representation but it was modality specific and all this sort of stuff, but it was representational.

SHAPIRO: Well, this is the question that I had for you when you were talking about embodied representations. If you have an embodied representation, what does that mean? Does that mean it's a non-symbolic representation? See, I thought that the work of people like Gibson, and Brooks, and Thelen was intended to be antagonistic toward symbols. Look, you say you need a cognitive map, Brooks might say, but here, my creatures get around the environment - no cognitive map

Graesser: I see them as not symbol antagonists but as representational antagonisst, is the way I view them. But I'm just… Well, I don't know, maybe Manolo can speak to this.

De Vega: I was surprised by your identification of symbols with embodied theories and non-representational theories. Because I think it is quite natural to consider embodied representations. What does it mean? Well, first of all you can establish a neurological description in which you find out some ways of overlapping between embodied meaning and perception and action. This means, okay this is an off-line embodied representation. Secondly, you can have some sort of behavioral study in which you demonstrate, for instance, some sort of interference or facilitation between perceptual and motor events and sentence meaning. In all these cases you can say, okay, this means that you are starting with a symbolic code which is language, that something occurs in your brain that is some sort of representational activity, and this sort of activity is important because it interferes or overlaps with motor-perceptual activity.

SHAPIRO: Well, now I'm getting confused what the contrast is supposed to be between embodied approaches to cognition and symbolic approaches. If embodied approaches are consistent with symbolic approaches, then what's the debate about? I'm really confused.

Gomila: May I continue? A historical point, which makes me feel old; you're younger. In the late 70's, early 80's the discussion took the form of the imagist debate, whether there are images in the mind and the discussion was images against propositions.

SHAPIRO: Yes, between Pylyshyn and Kosslyn.

Gomila: Well, it still goes on but I would say that the result of that debate was that we have to acknowledge the existence of non-propositional forms of representation. But not like pictures in the head that have to be seen. But as a way to understand this imagistic representation we have this perceptual systems theory that tries to account in terms of sensory or motor activity. The difficult issue here or the terminological issue is whether for this approach, embodied meaning is just grounded symbols. Because if it's the same, so if embodied meaning in the way that it was

discussed this morning, is just grounded symbols then it's a symbolist position, but the theory of the grounding of the symbols is very different, it cannot be a causal theory.

SHAPIRO: Well, again I want to resist this idea that to be a symbolist means you have to ground the symbols in a way that Searle's Chinese room suggests they have to be grounded. Now, I'm perfectly happy to accept, speaking as a symbolist – and you know, I don't know what I am – but if I were a symbolist, I'd be happy to accept this idea that some of my thoughts have meaning in virtue of my bodily interactions with the environment. But people in embodied cognition have to say more than that to distinguish themselves from traditional conceptions of cognition. I mean no one doubts that, or no one should doubt, that my perceptual states acquire meaning because there's a world outside me and I have to open my eyes and interact with the world in order to have perceptions with content. But embodied cognition people have to say something more distinctive than that. They have to say something more distinctive than our perceptions, our thoughts, get meaning by interaction with the environment. This is something a symbolist would easily accept. They'd say well, of course if your brain were in a black box, then you wouldn't have any thoughts. So what more does embodied cognition say? How do you distinguish or make novel what embodied cognition is about? And my efforts to answer that question led me to this view that, well, given that everyone agrees that we have to interact with the environment, and it's our body that's going to allow this interaction, the real difference has to be in the role that embodied people think symbols play in cognition. And what they want to try to do is, as I think Thelen, Gibson and Brooks do try to do, they've tried to develop accounts of cognition that minimize the role of representation.

Sanford: Yes, I was interested in your ideas about phlogiston and élan vital. I mean, do you see the problem as being that we just don't know enough about how the brain actually does things, that is, do you think the distinction between symbol activity and embodied activity would disappear under one umbrella if one could get to the appropriate level of neuroscience? Is that what you mean or do you mean something different, or something completely unknown?

SHAPIRO: Here's what I mean…if we could watch peoples' brains as they behave we wouldn't see any symbols in their brains, okay. Symbols…when we attribute symbols to a cognitive agent or to the process of cognition – these are theoretical posits. We're attributing symbols. This is a way of describing what's going on and it's in virtue of this description we're able to predict and make explanations. And in this regard I think cognitive science is like any other science. No one sees forces. No one sees gravity. These are posits that allow us to predict and model things. Symbols are like those. The question, as I see it, is, are these posits really valuable? Do we need these things? Is there a better way of explaining things? So, we no longer need to talk about the gods' anger to explain why lightning sometimes hits the earth. That's turned out to be a very unuseful posit. And I think it's still an open question whether symbols are a valuable posit. And the kind of ecumenical conclusion I like is that, well, it sure seems that they're necessary for some cognitive processes like language use and perhaps dead reckoning but perhaps they're not….perhaps we can look at Brooks' work to see that, no, if we thought we needed a symbolic representation in the environment to navigate through it, we were wrong about that, so that's one case where they're more like élan vital or phlogiston. That's what I was trying to develop there.

Steels: The game is to grab the microphone here. So, I just wanted to clarify without any debate but particularly this position of Brooks and others in behavior-based robotics. I mean it was really as you said. In fact, you take a behavior like following a wall or avoiding an obstacle….and so the point of this is actually very simple, which is you don't need to categorize this something as an obstacle or something as a wall because you build a direct link between let's say an infrared sensor which is reflecting from the wall. So there is no categorization. I think that's the better way to think about this thing – or the same thing with the wall. You have an attraction to the wall or a repulsion from the wall, as an obstacle. And then if you put two things together and let them go, then you see emergent wall-following without having the category 'wall', so rather than talking about symbols or no symbols… well, representations or no representations, I think the easiest way to understand is how conceptual, you know, these categories are. And I think today…I mean all of this happened ten years ago. Or fifteen years ago. So, today in robotics at least, today, I think you realize you can do many things this way. But as soon as you, let's say, do navigation and you want to, you know, remember what was behind you, you need some sort of representation of that. So I think it's generally recognized that for sensory motor behavior these things would be fine – I mean not having these categorizations, but as soon as you go to something more interesting then you need categorization, which is in fact what you said.

SHAPIRO: Yeah, and that line – at what point do you need them, and at what point you don't…I think the line used to be, you know, oh, you need them for everything. And now, I don't know whether the line is moving up or down, but it's changing position, and this is why I was talking to Deb last night about Mitchel Resnik's work. You know he's got this neat book *Turtles, Termites, and Traffic Jams*. It's full of these neat examples of behavior that looked like, gee, you needed a representation to explain it and it turns out you don't.

Pulvermüller: I have to disagree with one thing you said. You said that if we looked at the brain we wouldn't see any symbols. I think we have…When looking at brain activation when people process symbols - words and other symbols, we see very clear brain correlates of different types of symbols and therefore, in the relevant sense we see brain correlates of symbols.

SHAPIRO: Okay, that's not something I want to take issue with. I mean if you think of a word as a symbol and suppose corresponding to every distinct word there is some brain state, I'd say yes, these are correlates of symbols. My objection was something like this: suppose that your best description of how bees find the source of nectar involves a Randy Gallistel type explanation, that attributes to bees all these sorts of vectors and positions of the sun and things like that. Well, this is a level of description, right? If we look in the brain we're not going to see vectors. And what we do is we impose this symbolic model onto this hardware. But are there really symbols in the brain? No.

Pulvermüller : I think we are very clear that such a symbol exists or has a clear framework or correlation. The question is how to describe it, whether you want to use vector description of the symbol or whether you want to use an embodied description, characterization. I think that's the question as I understand it.

SHAPIRO: I think we're agreeing more than you think we are. Let's talk about this more later.