Studies in History and Philosophy of Science 41 (2010) 294–304

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

Embodied cognition and linguistic comprehension

Daniel A. Weiskopf

Department of Philosophy, Georgia State University, PO Box 4089, Atlanta, GA 30302, USA

ARTICLE INFO

Keywords:
Language
Comprehension
Perception
Action
Embodied cognition
Representation

ABSTRACT

Traditionally, the language faculty was supposed to be a device that maps linguistic inputs to semantic or conceptual representations. These representations themselves were supposed to be distinct from the representations manipulated by the hearer's perceptual and motor systems. Recently this view of language has been challenged by advocates of embodied cognition. Drawing on empirical studies of linguistic comprehension, they have proposed that the language faculty reuses the very representations and processes deployed in perceiving and acting. I review some of the evidence and arguments in favor of the embodied view of language comprehension, and argue that none of it is conclusive. Moreover, the embodied view itself blurs two important distinctions: first, the distinction between linguistic comprehension and its typical consequences; and second, the distinction between representational content and vehicles. Given that these distinctions are well-motivated, we have good reason to reject the embodied view of linguistic understanding.

1. Introduction

While all creatures that can interact with the world have perceptual and motor capacities, humans also have capacities for language and conceptualized thought. A natural question is how these additional capacities are related to the sensorimotor capacities that we share with other organisms. Focusing here on their synchronic (as opposed to evolutionary or developmental) relationship, the question is: do human conceptual and linguistic systems share a representational code and processes with sensorimotor systems, or do they have their own, proprietary coding scheme?

This question has particular force with respect to language. Attempts to model the psychological structures that underlie language date back to the start of the cognitive revolution. One common—but by no means ubiquitous—assumption has been that the language faculty is a separate computational module, possessing its own characteristic representations, database of rules, and processing operations. On this view, there are dedicated rules and representations the operation of which accounts for the normal speaker's ability to produce and comprehend indefinitely many sentences having novel syntactic, semantic, and phonological properties. Moreover, language understood in this way is computationally autonomous from other cognitive systems, including both perceptual and general reasoning capacities.[1]

Advocates of embodied cognition argue that conceptual and linguistic representation is not *sui generis*, but is essentially a matter of reusing sensorimotor representational capacities. This thesis, which has been the focus of attention from psychologists, philosophers, and linguists in recent years, comes in various forms corresponding to different strengths and cognitive domains.[2] For example, consider two claims that are invoked in recent work in language comprehension, the Immersed Experiencer Framework (IEF) and the Indexical Hypothesis (IH):

> (IEF): Language is a set of cues to the comprehender to construct an experiential (perception plus action) simulation of the described situation. In this conceptualization, the comprehender is an immersed experiencer of the described situation, and comprehension is the vicarious experience of the described situation. (Zwaan, 2004, p. 36)

E-mail address: dweiskopf@gsu.edu

[1] For one detailed development of a modular picture of the semantic properties of language, see Borg (2004).

[2] See Anderson (2003), Clark (1997, 2006), Gallagher (2005), Gibbs (2006), Robbins & Aydede (2008), Shapiro (2004), and Wilson (2002) for discussions of embodied cognition more generally, and comprehensive reviews of evidence in its favor.

0039-3681/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsa.2010.07.005

> (IH): The first step [in understanding sentences] is to index words and phrases to objects, analogical representations of the objects such as pictures, or to perceptual symbols. . . . The second step is to derive affordances from the indexed objects. The third step is to mesh those affordances into a coherent (i.e. doable or envisionable) set of actions. (Glenberg & Robertson, 1999, p. 7)

Both of these claims are variations on the fundamental premise of embodied cognition, which is that cognitive capacities are shaped and structured by the bodily capacities of a creature, including the sensorimotor capacities that make possible its basic interactions with the world. I will call the view that linguistic understanding involves embodied representations the *embodied linguistic comprehension* (ELC) thesis to distinguish it from claims about embodied cognition that are more general, or restricted to different domains (e.g. memory).

I will argue that ELC involves an implausible view of what is involved in the process of linguistic comprehension. I focus here on recent studies of language comprehension carried out by Arthur Glenberg, Michael Kaschak, Rolf Zwaan, and their collaborators. These studies are significant for two reasons. First, language is a multimodal device: we can combine and convey information linguistically about things that have been detected in any sense modality. This has suggested to many that language encodes this information amodally. ELC, however, threatens this plausible idea. Second, understanding sentences involves mapping them onto conceptual representations of the thoughts they express. So these studies may also support the claim that our concepts, too, are grounded in our bodily structure. In particular, they may support the view that concepts are recombinations of perceptual representations, as some have recently argued (Barsalou, 1999b; Prinz & Barsalou, 2000).[3]

The outline of the discussion is as follows. First, I lay out the traditional view that the embodied view of language comprehension is opposed to. I next present three lines of evidence that have been taken to support ELC. Each focuses on the existence of an effect of language on subsequent or simultaneous perceptual or motor representation, or on how language processing is influenced by properties that are available in sensorimotor processing. These effects tend to support the claim that comprehension and sensorimotor capacities are linked.

However, this evidence can also be accounted for by advocates of the traditional view. On this picture, central cognitive capacities such as language and conceptualized thought are intimately related to, but not constituted by, the functioning of embodied, sensorimotor capacities. In arguing otherwise, proponents of ELC run together two distinctions that should be kept in place: first, the distinction between constitutive and contingent relations between cognitive systems; and second, the distinction between representational vehicles and content. If we take these distinctions seriously, the case for ELC loses much of its strength. Finally, I argue that the debate between ELC and its opponents stems from disagreement about what the function of the language faculty actually is, and suggest that the traditional view is closer to the truth here than ELC is.

2. The traditional view

To clarify the content of ELC, it will help to contrast it with what I have been calling the *traditional* view of language comprehension (TLC). The view is comprised of two principles that describe the function of language and how that function is implemented in the mind:

> *The Translation Hypothesis*: The language faculty is a device whose main function is to translate between an external representational medium (words and sentences) and an internal representational medium (semantic or conceptual representations).
>
> *The Amodality Hypothesis*: The internal representations that the language faculty translates to and from are amodal, or distinct from those used by the creature's various sensorimotor systems.

As Gibbs (2006) puts it, effectively summarizing this view: 'The traditional belief among many cognitive scientists is that meaning is an abstract entity divorced from bodily experience. Understanding language is assumed to require breaking down the physical information (e.g. speech sounds) into a language-independent medium that constitutes the "language of thought"' (ibid., p. 158). The traditional view has been defended at any number of places across many different fields: in AI (Schank, 1972), in linguistics (Jackendoff, 1983, 1997, 2002), in psychology (Kintsch, 1998; Levelt, 1989; Miller & Johnson-Laird, 1976), and in philosophy (Fodor, 1975; Katz, 1972; Katz & Fodor, 1963).

ELC primarily aims to challenge the notion that the vehicle of semantic or conceptual representation is amodal. As Gibbs puts it: 'We must not assume cognition to be purely internal, symbolic, computational, and disembodied, but seek out the gross and detailed ways that language and thought are inextricably shaped by embodied action' (Gibbs, 2003, p. 2).[4] To say that the semantic code is amodal is merely to make a negative claim: it is not the same as the codes used by any of our sensory or effector systems (nor is it a subset of those codes). Some also claim that language and perception require different *kinds* of code: perception might use a geometric, image-like, or spatial form of representation, while the linguistic system uses a compositional, predicate-argument format. The further claim about different formats isn't necessary; all that's needed is that the codes differ. Even perception could use a predicate-argument style code, as long as it differed from the one in which semantic content is represented.

On the traditional view, linguistic comprehension is a process of mapping seen or heard words onto representations of their meaning. The output of the process is a semantic representation that captures the content of the input sentence. I will say that, on the traditional view, being able to *understand* (or comprehend) a sentence is a matter of being able to produce such a semantic representation. Sentence meaning is at least one thing that can be computed by the linguistic system on encountering a novel sentence. That isn't to say that on every occasion when a sentence is heard, the sentence meaning is computed first. It is even possible that sentence meanings are not routinely computed in on-line comprehension of language. It may be that something like utterance meaning is primary, in the sense of being the kind of content that is routinely computed in the process of comprehending language. All that I require is the claim that having the *capacity* to understand a sentence, for the traditional view, is a matter of being able to compute its semantic representation. As we will see, this slim conception of understanding is enough to pose problems for ELC.

[3] See Weiskopf (2007) for critical discussion of some of the evidence advanced in support of this claim about the perceptual foundations of concepts.

[4] I should note that Gibbs is here running together several notions that can be disentangled. Being internal, symbolic, computational, and disembodied are, *prima facie*, different properties. The heart of the embodied research program is, strictly speaking, the claim that cognition reuses sensorimotor representations and processes. This only addresses the latter claim of the traditional model. See Sect. 7 for some discussion of relations between computational models of cognition and embodied accounts.

3. Evidence for embodied comprehension

Three major effects provide support for ELC: the *appearance*, *affordance*, and *action* compatibility effects (the ACEs). The appearance compatibility effect is evinced in studies of sentence-picture mapping. Zwaan, Stanfield, & Yaxley (2002) had participants read sentences describing an object in situations that result in its having a different physical appearance. For example, an egg can be located in a nest or in a skillet; in a nest an egg is typically whole and ovoid-shaped, while in a skillet it is broken. They then saw a picture of the object and decided whether the pictured object was mentioned in the sentence. Where the picture matched the shape the object would have in the described scenario, people are faster to respond than when the object does not match. They are also faster to name pictured objects if the picture matches the appearance of the object in the scenario.

Similar effects appear in a study by Stanfield & Zwaan (2001), who asked participants to read sentences describing an action that results in an object either having a vertical or horizontal orientation, then judge whether a pictured object was described in the sentence. Response times were faster when the pictured object matched the orientation that the object would be in, given the scenario; if a nail is described as being hammered into a wall, people respond faster to a picture of a horizontal nail than to a vertical one. These studies suggest that people tend to generate perceptual images of described scenarios in sentence comprehension. This result is predicted by ELC, which claims that comprehension just *is* creating such embodied scenarios.

Appearance effects also appear with dynamic stimuli. Zwaan et al. (2004) found that if participants were presented with sentences describing an object's motion towards or away from them (e.g. 'You throw the ball to John'/'John throws the ball to you'), then shown two slides of an object in a sequence that makes it appear as if the object is coming closer or moving farther away, they were faster to judge that the same object is pictured in each slide when the direction of motion depicted agrees with the direction that the object would be moving in the scenario described by the sentence. This suggests that in hearing the sentences, the participants are simulating the appearance of the moving object as it occupies a larger or smaller portion of the visual field over time.

Even comparatively abstract moving stimuli can produce these effects. Asking participants to listen to sentences that imply motion while viewing visual displays that instantiate the same direction of motion *slows* their judgments that the sentences are sensible compared to displays that instantiate different directions of motion (Kaschak et al., 2005). (Speeded responses to mismatching stimuli are explained by resource competition in matching cases.) The stimuli that imply motion need not even be visual. Heard sentences describing objects moving up/down or towards/away from the hearer are judged to be sensible more rapidly when they are accompanied by artificial auditory stimuli that 'sound' as if they are moving in a different direction from the one described (Kaschak et al., 2006). Here, again, the slower replies to *matching* stimuli allegedly result from resource competition: constructing a simulation of movement is harder when a competing representation is trying to use the same resources at the same time.

Other evidence for ELC comes from Glenberg & Robertson (2000), who asked participants to read descriptions of situations in which objects could be used in a way that they naturally afford versus ways that they don't readily afford. Affordances, in Gibson's (1979) terms, are properties of objects that make them available for use in certain kinds of immediate action. What an object affords depends not just on the object, but on the perceiver and how its physical construction enables it to interact with the object. One paired scenario involved a person using a newspaper to cover his face from the wind (afforded action) versus using a matchbook (non-afforded). Sentences involving objects whose affordances fit the use to which they were being put were rated as being both more sensible than those describing non-afforded uses, and also as being easier to envision. The fact that sentences describing affordance-compatible uses of objects are rated as more sensible suggests that comprehending those sentences involves representing the situation described in terms of the affordances of participating objects.

Affordance compatibility effects were also found by Kaschak & Glenberg (2000). In one study, participants read about transfer of an item from one person to another using an object that either affords such a use (e.g. a wheeled chair to help move a heavy box) or doesn't (a chair with no wheels). Participants were faster to comprehend the sentence describing the transfer itself when an object affording that action was used. In another study, participants again read passages about objects having situationally relevant but atypical affordances (e.g. using a crutch to pass someone an apple, rather than support one's weight), then read probe sentences describing the object as having a salient affordance (for the crutch, being long), a non-salient affordance (being sturdy), or a merely stereotypical property (being helpful with injuries). Sentences describing salient affordances of objects were verified faster than non-salient affordance or associated but non-affordance related properties. These results would be predicted if, as ELC claims, linguistic understanding involves constructing sensorimotor scenarios in which the affordances of objects are readily available.

Finally, the action compatibility effect arises when sentences describing the movement or transfer of an object align with the motor response a person is required to make. Glenberg & Kaschak (2002) showed that when participants judged the sensibility of sentences describing the transfer of an object, their response time to press a button is shorter when the direction they need to move in order to press the button—either towards their body or away from their body—aligns with the direction of the described transfer (from them to someone else, or vice versa). For instance, when reading the sentence 'Andy delivered the pizza to you', participants were faster to move towards themselves—in the direction of the pizza's described motion—to press a response key than away.

Interestingly, this effect arises not just for transfer of physical objects like pizzas and notebooks, but also for abstract entities like stories and messages. So judging that 'You radioed the message to the policeman' is sensible facilitates movement away from one's body—again, in the direction of the transfer—versus movement towards one's body. Glenberg & Kaschak (2004) suggest that this shows that 'the meaning of transfer, even the transfer of abstract entities such as information, seems to be embodied as action' (ibid., p. 24). Understanding a sentence involves creating an action-oriented motor representation, and when this aligns with a pre-existing motor representation, it facilitates the execution of the corresponding bodily movements. When it conflicts with a prior motor plan, it interferes with that plan's execution. So language functions not just to activate perceptual representations, but also motor representations. Putting these pieces together seems to give rise to a fully embodied picture of linguistic comprehension.

The three ACEs, then, give ELC a strong hand. They suggest that the process of understanding sentences can affect visual and motor processing, and can be affected by how plausibly what they describe can be enacted by creatures with bodies like ours. This puts pressure on TLC's amodality hypothesis. However, as we will see, all three effects can also be explained from within the traditional view of linguistic understanding.

4. Amodal actions

Take the action compatibility effect first. On ELC, linguistic understanding is a simulation process in the sense of Goldman (2006): it involves re-use of the same processes that are involved in seeing and acting. Just as mental imagery involves re-use of perceptual systems and representations (Kosslyn, 1994), and hence involves simulation of perception and action, so too does linguistic understanding involve simulation of perception and action in the described scenarios. I will call the process of generating a sensorimotor representation of a linguistically described situation *enactive simulation*. Recall that the Indexical Hypothesis claims that the final stage of understanding a sentence is to derive action plans from a perceptual simulation of the affordances of objects described by the sentence. Sentences, Glenberg claims, *imply* actions in virtue of making perceptual scenarios available (Glenberg & Kaschak, 2002, p. 559).

However, there are two possible ways to understand this implication. First, a sentence might make me call to mind plans of action for how *I* might respond to finding myself in the described situation. For instance, hearing 'The dog is growling' might lead me to imagine seeing a growling dog, which would lead me to begin planning my escape route. The motor representations, then, correspond to my own plausible movements. When these match the movements required by the experimental response, they facilitate execution of the response.

It should be clear that this cannot be the explanation for the effect, though. Consider the away/towards pair: (A) 'I give Josh the notebook', (T) 'Josh gives me the notebook'. In (A), movement away from myself is facilitated. This is consistent with imagining myself making the transfer motion. But this doesn't explain why movement *towards* myself is facilitated by (T). Why isn't extending one's hand just as sensible a response, since that would be the motor pattern required to take the notebook from Josh? The direction of the transfer here seems opposed to the direction of the required bodily movement. This suggests that one's own movements in response to the perceived situation cannot be part of what is simulated—or if they are, they do not explain the compatibility effect.

Moreover, this should be clear from the abstract transfer results, as Glenberg & Kaschak (2004) point out. They note (ibid., p. 24) that the motor pattern activated in transferring information involves movement of the lips, tongue, and so on. But it's far from clear how this could interfere with arm movements, since different motor representations are needed in each case. Again, enactive simulation of my own actions seems irrelevant to explaining the effect.

A second explanation is offered by Borreggine & Kaschak (2006). They propose that perceiving and acting have a common representational code (Prinz, 1990). Hence, perceiving a movement activates the same representations that are used in planning movements. On this view, reading a sentence causes construction of an enactive simulation of perceiving a situation. This simulation includes representations of how entities (concrete or abstract) are transferred from place to place; since these motion representations are shared with the action planning system, there is less cost associated with performing acts that require movement in the same direction as the simulated entity.

This common coding approach, however, is difficult to distinguish from a disembodied, amodal account of the same phenomenon, such as the following:[5]

(1) Hearing a sentence results in construction of an amodal, propositional representation of its truth conditions;
(2) Planning an action involves forming an intention to act in a certain way;
(3) The amodal code used to represent sentence meaning is the same as the code used to form intentions and other propositional attitudes.

Since shared representational codes can explain facilitation/inhibition effects, the amodal account here has the exact same structure as the embodied account does. In other words, the fact that a representational code is commonly deployed in two tasks doesn't by itself decide the issue of whether it is sensorimotor or amodal.

To make this a little more concrete, consider an example. I hear 'Josh gives you the notebook', and interpret it as meaning: GIVES(JOSH, NOTEBOOK, ME). From this I can immediately infer: MOVES(NOTEBOOK, TO ME). Given the setup of the experiment, I have a standing intention to move my hand in a certain direction if the sentence I hear makes sense. Since it does, I presently intend: MOVE(HAND, TO ME). This standing intention shares a directional representation with my earlier belief about the notebook's trajectory, which facilitates execution of the intended behavior. Both an embodied account and an amodal account can make an appeal to shared representational codes in explaining the action compatibility effect; the available evidence does not distinguish them.

In itself, this conclusion is not surprising. Neither the embodied view nor the traditional, amodal view of language comprehension are completely well specified models; rather, they are loosely organized families of models that share commitments to certain core representational principles. But in itself the available evidence is not sufficiently fine-grained to support one or the other set of principles. It is only at the more detailed level of particular models that the evidence can weigh in most heavily.

However, the conclusion I aim to reach here is more than a scotch verdict. As I will argue in the next two sections, there are considerations that favor the traditional view over ELC. These arguments turn on what facts a theory of linguistic understanding should be required to explain. I now turn to the first of these arguments.

5. Understanding without appearances

Like action compatibility effects, appearance compatibility effects, too, are explained by enactive simulation according to ELC. It is plausible that understanding a sentence, particularly a sentence that involves an observable physical event, may often cause us to imagine the event being described. Part of what makes fiction enjoyable is that it causes us to have such experiences, or puts us in the state referred to by novelist John Gardner as a 'vivid, continuous dream' (Gardner, 1985). The appearance compatibility effect suggests that in many cases this is what people do. When we read about an eagle in a tree we visualize it looking one way, but when we read about an eagle in flight we visualize it looking quite different.

There are three possible positions—strong, medium, and weak—that one could take on the relationship between enactive simulation and linguistic understanding. They are:

(ELC_s): Linguistic understanding just is an enactive simulation process;

[5] Prinz & Barsalou (2000), p. 63, define a disembodied representation as 'one harbored by a system that lacks a sensory-motor system [SMS] capable of interacting with the environment'. This definition is too strong, though. In my terms, disembodied representations are only those that are *type-distinct* from those used by SMS. This doesn't imply either that they are (or must be) had by creatures without SMS, or even that they *can* be had by creatures without SMS. It might be that SMS is necessary for disembodied representations to exist. Still, if there are any disembodied representations, their dependence on sensorimotor representations wouldn't entail that they *just are* sensorimotor representations.

(ELC_m): Linguistic understanding requires, but is not identified with, enactive simulation;
(ELC_w): Linguistic understanding can use, but does not require, enactive simulation.

Defenders of ELC often do not distinguish these three possibilities, but they frequently seem to have ELC_s in mind. Zwaan (2004), for instance, says that language 'is a set of cues' (ibid., p. 36) to construct an enactive simulation, but then in the next sentence says that comprehension '*is* the vicarious experience of the described situation' (my emphasis). This seems to move without a pause from something like ELC_w to ELC_s. I suspect that it is the latter, more radical view that he ultimately intends. Glenberg & Robertson (1999) are more explicit in their statement of IH. They straightforwardly identify the output of comprehension with an action plan based on the affordances that the objects described make available. This seems like advocacy of the strong version of ELC.

The major division among these views comes between ELC_s and ELC_m on the one hand and ELC_w on the other. The former two take enactive simulation to be *constitutive* of linguistic understanding, while the latter takes it to be *contingently* involved in it. The difference here is significant. If a capacity C_1 is (partially or wholly) constituted by another, C_2, then it should be impossible to find exercises of C_1 that are not also exercises of C_2. However, if C_2 is merely a contingent adjunct of C_1, it is possible for them to come apart, or to dissociate in some circumstances. This dissociation might be a possibility even if C_1 typically or often calls on C_2.

I suggest that the contingent view, ELC_w, is the most plausible of these possibilities. The argument for this claim turns on what one takes to be necessary to ascribe understanding of a sentence to a hearer. Here we can return to the traditional view for a moment. On TLC, to understand a sentence—to grasp its meaning—one needs to have knowledge of what its words mean and what it means to combine them in the way that the sentence does. This knowledge is essentially knowledge of the compositional semantic structure of the sentence itself. On a broadly Davidsonian perspective, knowledge of meaning is fundamentally knowledge of truth conditions (Higginbotham, 1989; Larson & Segal, 1995). Within this perspective, understanding a (non-indexical and non-context sensitive) sentence requires only knowing its truth conditions as expressed, for example, by the derivation of the relevant T-sentence. I will call any view on which knowledge of a sentence's truth conditions is sufficient for understanding the sentence a *minimal* view of linguistic understanding.

Having this minimal understanding involves at least being able to draw and recognize certain inferences that follow from the sentence. For instance, understanding 'John ate the bagel with gusto at noon' enables me to infer 'Someone ate the bagel with gusto at noon' and 'John ate the bagel'. These sentences just follow given the logical structure of the first sentence, and my drawing them is licensed by my knowledge of that logical (truth-conditional) structure. The issue, then, is whether these inferences necessarily include inferences concerning the appearances and affordances of the objects described in the sentence. Note that on ELC, since the output of comprehension is an embodied representation, language comprehension entails representing these properties as part of grasping what a sentence means.

There is reason to doubt whether comprehension requires representing these properties, however. Take a non-afforded sentence from Glenberg & Robertson (2000) such as:

(1) Adam pulled out of his golf bag a ham sandwich and used that to chisel an inch of ice off his windshield.

Plainly this is no way to clean one's windshield. Participants assigned the sentence the lowest rating on sensibleness and ease of envisioning. Yet the sentence can perfectly well be *understood*. For instance, on reading the sentence one can infer:

(2) Adam used something to remove the ice from his windshield.

(2) follows from (1) because the construction 'used x to y' conveys that x is an instrument for y-ing. Assuming the sentence is true, one can also infer:

(3) The thing that removed the ice was stored in the golf bag.

This fact can be derived from (1) by interpreting the pronoun 'that' as anaphoric on 'a ham sandwich'. Finally, one can infer:

(4) The sandwich was the kind of thing that could chisel ice.

This follows from (1), since if x is used to y, then x is capable of doing y. Given the non-semantic background knowledge that things that chisel ice have to be pretty hard, we might go on to infer that the sandwich was pretty hard. The fact that it's not *easy* to imagine a sandwich having these characteristics doesn't seem relevant to whether one grasps these basic facts about the sentence's truth conditions. These facts seem available to anyone with a grasp of the sentence's syntactic and semantic features, independent of whether they can enactively simulate the scenario it describes.

Even sentences that are wildly bizarre can support such inferences: if colorless green ideas slept furiously, then they slept, and they were both colorless and green. Visualizing what sort of scenario is being described by Chomsky's famous sentence is impossible. But some elementary facts about its truth conditions remain accessible nevertheless. This suggests that linguistic understanding in a minimal sense is independent of the ability to visualize what is described, even if in many cases we do make use of such visual images.

This dovetails with another point, namely that simulated perceptual experiences do not seem to be sufficiently fine-grained to distinguish between truth conditionally distinct sentences. First, consider the pair of sentences:

(5a) The man stood on the corner.
(5b) The man waited on the corner.

In what way does the enactive simulation prompted by (5a) differ from the one prompted by (5b)? It is hard to see any principled grounds for forming different images here, since the two situations just involve the placement of an adult male on a corner. Waiting and standing don't perceptually differ from each other.[6] But the sentences do differ in truth conditions. For example, waiting is an inherently goal-directed activity. There is always something one is waiting *for* (even if, like Godot, it never arrives). Not so for standing. Hence the difference between:

(6a) *The man stood on the corner for his girlfriend to arrive.
(6b) The man waited on the corner for his girlfriend to arrive.
(7a) *The man stood for the bus to arrive.
(7b) The man waited for the bus to arrive.

These pairs suggest that waiting and standing differ in whether there is an implied goal. This is a semantic difference—a difference in truth conditions—between 'stand' and 'wait'. Moreover, it is a difference that is easily recognized by those that understand (5)–(7). But it is hard to see how to capture this difference in terms of sensorimotor simulation.

[6] Assuming, as I do here, that the default way in which one waits is by standing. This is not part of the meaning of 'to wait', of course—one can sit down to wait, or lie down, and so on.

Note that this sort of case can be multiplied. For another example (due to Landau & Jackendoff, 1993), consider the difference between:

(8a) *Bill threw the ball at Harry without meaning to hit him.
(8b) Bill threw the ball toward Harry without meaning to hit him.

Sentence (8a) is obviously strange, since 'threw at', as opposed to 'threw toward', connotes intention on the part of the thrower. But it is hard to see how one would imagine the situations described (throwing toward vs. throwing at) differently, at least if one was restricted to a sensorimotor vocabulary.

Finally, the fact that we can often make inferences about truth conditions even when dealing with novel lexical items suggests that visualization is not essential to understanding. On hearing 'The dax was sleeping on the bed', you can infer that a dax is a living creature, since only living things can sleep. That minimal inference is licensed by the lexical entry for 'sleep', which requires an animate subject. But no information is available about what a dax looks like, or how it sleeps (does it curl up, or stretch out and sprawl over the edge of the bed?). Nor does such information seem necessary to making inferences that are licensed by lexical entries.

What these three cases suggest is that sentence understanding can happen without enactive simulation. Simulation, then, is at most a contingent accompaniment of such understanding. This minimal conception of linguistic understanding is consistent with ELC_w, but not with either of the stronger conceptions.[7] And since ELC_w is consistent with TLC, this is no objection to the traditional view.

6. Vehicle and content in grasping affordances

I turn now to the second major argument against ELC, which focuses on the affordance compatibility effects. According to Shapiro, these effects show that 'an ability to understand sentences seems, at least in many cases, to incorporate an organism's knowledge about how its body might interact with objects in its environment' (Shapiro, 2004, p. 212). I have already argued that understanding *per se* should be separated from capacities to imagine described situations. However, this leaves it open that judging whether these situations are sensible or coherent (as opposed to merely understanding the sentences that describe them) might depend on embodied knowledge. In assessing this claim, though, we need to recall the crucial distinction between psychological *content* and representational *vehicle*.

Contents are what is represented (individuals, properties, states of affairs), while vehicles are the actual structures—the mental particulars—that do the representing, or that carry the content. The same content can potentially be encoded in many different vehicles. This point is easily seen by thinking of content in terms of information (Dretske, 1981). I may convey to you the information that the cat is sleeping by drawing you a picture, showing you a photograph, pantomiming, or simply telling you. In particular, perceptual content—the detailed information represented by a creature's sensory states—can be expressed in many different formats as well. This fact is what makes it possible for us to truly describe what we see. Accurate perception of the dozing cat gives rise to a rich perceptual representation of her; while I may lose some of this rich, perspectival content in telling you what I see, there is nevertheless a common informational core to what is conveyed to you by my sentence and what is conveyed to me by my perception. This informational core is just the information—the proposition—that the cat is asleep.

Affordances are (possibly transient) properties of objects that make them suitable for certain kinds of actions. In Gibson's (1979) theory, we directly perceive these properties. They are themselves part of the content of perception. In seeing a chair, I directly perceive its property of being sit-on-able. We need not adopt so controversial a view, though. Information about affordances might be spontaneously extracted by a complex inferential process that is fed directly by our perceptual systems. When it comes to everyday objects, thinking of them might immediately bring to mind information about what sorts of acts they typically afford. The thought of a mug brings to mind the fact that it can be grasped in such and such a way and used for drinking hot liquids safely. Information about affordances, like information about the other properties objects have, is part of the rich body of background knowledge that we have about our world.

Insofar as they are mentally represented, then, affordances can be understood in two ways: (1) as being part of the content of *perceptual* states; or (2) as part of the content of informational states that are immediately derived from perception—for example, beliefs or other *cognitive* states. What affordance compatibility effects show is that we draw on states that carry information about affordances in comprehending scenarios. These are part of our background knowledge of objects and their possible behaviors. This doesn't, however, show that this background knowledge is carried by the same vehicles that are used in sensorimotor interactions with those objects. From evidence concerning content, we cannot immediately infer anything about the *vehicles* of that content. So we can't conclude that sentence comprehension re-uses the same representational vehicles that are deployed in sensorimotor interactions. And it is the thesis about vehicles that ELC needs to establish.

Moreover, the effect doesn't even demonstrate that the content that includes information about affordances is perceptual content rather than conceptualized content. In many of the scenarios, it would be possible to figure out that an object would or wouldn't be useful in a situation just by thinking it through—that is, by reasoning on the basis of what one believes about objects of that sort and what can be done with them. To borrow another of Glenberg and Robertson's examples, it's clear that you can't fill a sweater with water and successfully use it as a pillow, whereas you can fill it with leaves for that purpose. I might arrive at that belief just by noticing that sweaters can't be sealed shut and aren't made of material that will keep water on the inside. Coming to judge that these scenarios are or aren't sensible might rely on general abductive competence—the ability to make inferences about real-world situations drawing on relevant knowledge—rather than any specifically sensorimotor capacities.

The power of abductive reasoning is that it can potentially operate on any sort of information, including perceptual information concerning affordances. Presumably we have a lot of such information, including information about how our own bodies are put together, how the general physical laws of the world operate, how things look and behave, and so on. Figuring out how to make programmed computer systems draw on this sort of information in comprehending natural language has been a long-running project (Schank & Abelson, 1977). Since we have considerable reason to think that discourse understanding (as well as much of everyday life) involves abductive reasoning that taps this information

[7] Something like this idea is expressed by Marconi (1997), p. 106, although he doesn't ultimately endorse it: 'Intuitively, to understand a sentence is to know under what conditions it would be *true*; it is *not* to know how one would go about verifying the sentence or to be able to construct a mental model corresponding to the sentence or anything else of a cognitive nature'. I would amend this slightly: understanding might in fact be a matter of having the right cognitive state, namely a state in which one represents a sentence's truth conditions.

(Kintsch, 1998), the embodied cognition claim seems superfluous in explaining how these inferences are made. We reason about affordances in deciding whether scenarios make sense. But this modest claim about the content of our reasoning doesn't show that we are reusing sensorimotor representations in so doing.

Similar content-vehicle mistakes appear elsewhere in the literature on embodied cognition. One sometimes hears, for example, that our concept CHAIR could be grasped only by creatures that have our kind of sensory organs, bodies, and motor systems.[8] This is because part of the concept of a chair is that it is for sitting on, and understanding what is meant by 'sitting' here requires knowing that humans sit on their posteriors, that their legs bend at the knee, that they require back support to sit comfortably, and so on. All of these things go into understanding what it is for a human to sit in a chair. (A cat sits on a chair in a much different way than a human does.) And not knowing these things impairs one's ability to grasp the concept CHAIR, at least in the way that humans do. Understanding the word 'chair', then, requires the right sort of embodied encounters with chairs themselves.

But this line of argument conflates *having* information about humans' physical structure and *representing* that information in a particular way (e.g. a way that only a creature with humanlike sensorimotor capacities could enjoy). Presumably, any sufficiently sophisticated informationally sensitive creature can represent the information that goes into the concept CHAIR as described above. Perhaps even a creature with limited mobility but sophisticated perceptual and reasoning capacities could acquire such a concept by observation alone. Indeed, it has been common practice for decades to train neural networks to acquire such concepts (Rumelhart et al., 1986). Starting with a set of features that reflect physical, functional, and social properties an object might have, such networks develop sensitivity to chairs as such by learning associations among these features. The point here is not that these are realistic models of human concepts in general—or even of humans' CHAIR concept—but rather that disembodied computational systems can readily learn the content of this concept, insofar as it is constituted by properties and their interrelations. The same content can be realized in a wide range of representational structures. There is no reason yet to think that one needs to have the right sort of body in order to grasp this content.

A final content-vehicle conflation is exemplified in Gibbs et al.'s (1994) study of how the verb 'stand' can be interpreted in various literal and non-literal contexts. They asked participants to decide how central various 'image schemas' (picked out by descriptors like VERTICALITY, BALANCE, RESISTANCE, etc.) were to their experience of standing. A second group of participants judged how similar thirty-five different senses of 'stand' are in meaning. Finally, a last group of participants were given thirty-two different senses of 'stand' and asked to rate how related the most important image schemas were to each one of them. The major results were that participants classified literal and non-literal uses of 'stand' together, and that these groupings could be predicted from the associations of those uses with different images schemas. So if two distinct uses were associated with the same pattern of image schemas, they were likely to be judged to be close in meaning. Gibbs et al. conclude from this that image schemas derived from perception are central to explaining the distribution of senses of 'stand'.

The difficulty here, however, is that while participants were asked to stand and reflect on various features of their experience prior to carrying out several of these studies, the image schemas themselves were conveyed by linguistic descriptions; for example, 'Balance refers to your sense of symmetry or stability relative to some point in your body' (Gibbs et al., 1994, p. 236). Participants may find it easy to classify their experience in terms of the properties picked out by those descriptions, and these assignments of properties may predict semantic relatedness judgments. But none of this shows that in tokening the meaning of 'stand' in each one of these different contexts people are re-activating the very proprioceptive or somatosensory representations that they experience when standing. At most it shows that properties which they use to classify their experience can also be associated with different senses of 'stand'. In other words, the same content that is involved in judgments concerning one's experience can also be used in semantic judgments. Despite the fact that Gibbs et al. refer to 'image schemas', there is no evidence that this content is in fact *encoded* in an imagistic format, or in a format appropriate to enactive simulation.[9] Since they are silent on the question of what vehicles carry the content of 'stand', nothing in these studies is incompatible with TLC.

7. A priori arguments for embodied comprehension

In addition to the empirical arguments surveyed so far, advocates of ELC have also offered a priori arguments against the traditional view. One such argument owes to Stevan Harnad (1990), who claims that purely symbolic models of mental phenomena such as language comprehension all face the 'symbol grounding' problem. The problem is essentially this: given (1) that human cognition involves thoughts with semantic content or meaning, (2) that computation is just a matter of manipulating arbitrary symbols according to formally specified rules, and (3) that no purely formal relations among symbols are sufficient for those symbols to mean anything, how can cognition be computational? A system that simply manipulates meaningless symbols cannot be a model of meaningful cognition.

Embodied cognition theorists take this problem to be a serious challenge for computational theories of comprehension and other cognitive processes (see, for example, Barsalou, 1999b). It's worth noting, however, that Harnad himself thinks the problem is solvable if arbitrary symbols are linked with devices that connect them to perceptual systems in the right sort of way; for example, a CHAIR symbol might be connected to a perceptual template matching device that activates when a stereotypical chair is seen. Symbols get grounded, for him, by having such relations to perceptual categorizers. In criticizing connectionist models that employ only devices for categorizing sensory input, Harnad (1993) says that symbols are '[u]nnecessary for sensory spatial analysis, perhaps, but cognitive modeling also needs to scale up to the rest of our cognitive competence—including *concepts*, which have systematic language-of-thought properties'. If embodied cognition theorists adopt Harnad's understanding of the problem, then, it would seem

[8] Shapiro (2004), p. 212, makes a very close variant of this argument. He does not go so far as to say that *only* something with our body and sensory capacities could grasp this concept, but he does say that comprehending sentences involving 'chair' depends on these bodily facts about us. There is a weak interpretation of dependence that leaves open the possibility that something physically different from us *could* nevertheless grasp the same CHAIR concept.

[9] This point applies more generally to much work in embodied cognition that focuses on the centrality of the body in various metaphorical domains. Gibbs (2006), Ch. 4, cites research in this tradition in arguing that human concepts are influenced by representations of bodily properties. In itself, though, this fact is hardly surprising. Our bodies are surely among the most salient objects in our lives—if indeed they are correctly viewed as 'objects' at all. It is not surprising that we spend a large amount of time thinking about them, and that projecting their structure and features onto other parts of the world is a virtually irresistible cognitive temptation. The claim that cognition is 'embodied' in this sense, though, is trivial. It is a consequence of the importance of our own bodies to us, combined with the fact that people tend to think and talk in terms of what is closest and most familiar when trying to understand the distant and unfamiliar.

that they need to give some account of why his own solution to it would be unacceptable.[10]

A distinct a priori argument against symbolic models of language is given by Glenberg & Robertson (2000). Rather than appealing to the ungroundedness of symbols, they suggest that no purely symbolic system can account for the affordance compatibility effects surveyed above. Specifically, they claim that no symbolic process can account for the observed differences in judgments of sensibility between afforded and non-afforded sentences. They pose their challenge as follows: 'consider asking a symbol system if the following novel combination is sensible: Can symbol 10011001 be put in relation 11110001 with symbol 10011010?' (ibid., p. 397). This way of putting the point muddies things a little, since it makes it seem as if the system is evaluating facts about symbols rather than facts about what the symbols stand for. Reformulating slightly, the question is how a system that manipulates arbitrary symbols can appropriately compute relations among objects that bear complex perceptual and functional relations to one another. The issue here centers on *how* a symbol system would have to compute facts about novel relations among objects and events—that is, whether it is possible to arrive at appropriately graded judgments of sensibility given only the resources of an abstract computational system.

First, Glenberg and Robertson say, '[b]ecause the symbols are arbitrarily related to the objects, the symbols cannot literally be juxtaposed (e.g. '1111000110011010') to produce anything sensible. This is true even if the symbols included a very fine-grained coding of perceptual features because by definition those features are arbitrarily related to real shape' (ibid., p. 397). The point, then, is that a visual image of a matchbook adjacent to one's face in some sense makes it clear that it can't be used to shield oneself from the wind, a symbolic representation like ADJACENT_TO[MATCHBOOK, FACE] does not. Because properties of the objects like their relative size are not encoded as part of the representation of the objects themselves, the comprehension system cannot derive facts about whether this juxtaposition is sensible just by comparing the symbols that comprise the propositional representation itself. Images and other sensory representations, however, do encode such information. Given the right sorts of processes, then, it should be possible to recover facts about such affordances from embodied representations directly.

But, one might reply, surely one can derive the fact that a matchbook isn't useful for covering one's face by making abductive inferences on the basis of one's background knowledge (of e.g. the physics of wind)? To this Glenberg and Robertson argue, in effect, that symbolic systems will invariably fall prey to the frame problem if they attempt to derive sensibility judgments abductively:

> Attempting to find a logical relation between the two original symbols results in a combinatoric explosion that would soon overwhelm even the most powerful computers. More devastatingly, given enough time and enough facts, a logical path can be found to connect almost any two objects. For example, a matchbook is made of cardboard. Cardboard can be recycled. Recycled cardboard can be turned into a large sheet of newsprint ... and so a matchbook can be used to cover the face. (Ibid., p. 398)

The central claim here is that symbolic computational systems cannot draw only the appropriate, situationally relevant inferences to derive these judgments. So, since the information needed to make sensibility judgments is not contained in the propositional representations themselves, and cannot be derived by any tractable, constrained sort of inferential process, symbolic models of comprehension must be untenable (*contra* the view sketched in Sect. 6).

Suppose for the sake of argument that we grant that symbolic systems are subject to the dilemma posed here. In particular, suppose that we grant the assumption that these systems face something that might appropriately be called the frame problem (or the relevance problem) when it comes to explaining how appropriate information is retrieved from memory and appropriate inferences are drawn in the process of comprehension. The question then is, since we obviously do retrieve such appropriate information in comprehension, reasoning, planning, and so on, what sorts of mechanisms allow us to do so?

Following Barsalou (1999b), I will call a mental mechanism that is capable of producing an enactive simulation a *simulator*. Now consider the kinds of information and processes that a simulator would have to contain in order to solve the sorts of complex problems that embodied cognition theorists assign to them. A simulator that can decide whether an object affords a certain sort of action would have to contain information concerning the object's physical and functional properties, the properties of the cognizer's body, innumerable pieces of specific and general information about the physical world, about other living things and their behavior, and so on. If this information is not somehow in place within the system, it would be impossible to explain how it can generate *accurate* simulations of the results of taking some action, for example, trying to block the wind with a matchbook. The simulation depends on our having a more or less correct grasp of how the physics of solid objects of a certain size works. Moreover, if the simulator is not itself to fall prey to the frame problem, it must retrieve and use only the appropriate, situationally relevant information. The inferences it draws using this information must similarly be the relevant ones given the task at hand. And it is just as easy to simulate irrelevant details of a scenario as it is to simulate relevant ones. Some mechanism must be present that activates and runs only the right inferences. That is: simulators themselves must embody some way of solving the frame problem if they are to do the work needed in embodied accounts of comprehension.

But the problem now is that we simply have no account of what sort of mechanism might actually effect this solution. While it is certainly true that overcoming the problems of relevance and (potential) combinatorial explosion is part of our cognitive profile, there are essentially no computational systems that do this for a realistic range of cases. So while it might be true that classical computational devices face this problem, and perhaps cannot overcome it, it is similarly true that we don't know what other sort of system might do it. Embodied cognition theorists have, in effect, just stipulated that simulators are capable of doing so.

Since *something* must be capable of doing so, this might seem not to be a suspect move—but in the context of a debate over the appropriate representational format for comprehension and cognition more generally, it actually is. The reason is because, as indicated above, frame problems arise more or less *independently* of what sort of representational vehicles one chooses to build a system. This is because the frame problem is a problem about computational processes, inferential rules, and mechanisms, not a problem about format. Consider again the example of the matchbook sentence. Forming a perceptual image of the scene in which I try to use the matchbook to cover my face is one thing; using it to infer whether this action is sensible is another. The representation contains information about the relative sizes of my face and the matchbook, and much else besides. But to reach the *right* conclusion—to answer the *relevant* question—I need to extract and use

[10] I don't mean to suggest here that I endorse Harnad's solution, only to challenge the use of it against amodal theorists without taking into account the fact that he thinks it can be overcome within a kind of hybrid cognitive architecture.

only some of this information, and do the right thing with it. But given a rich enough repertoire of inferential processes, the option exists, just as in the case involving propositional representations, of drawing innumerable irrelevant or pointless inferences as well. Nothing about the format itself can foreclose this possibility.

One might argue that the frame problem, as such, is only well defined for classical computational systems, and hence that embodied systems such as simulators don't need to provide an answer to it. Presumably the same thing would be said about other non-classical systems such as connectionist systems and cognitive dynamical systems. But this doesn't seem right. If the problem is how a system can rapidly retrieve relevant information to solving a particular problem, draw relevant inferences, and update its stored database of information in a computationally tractable way, this seems to arise independently of the kind of representations the system uses. And choice of representation alone won't solve the problem—processing assumptions play a crucial role here. It is worth noting in this context that some connectionists appear to think that something that is essentially the classical frame problem arises for their style of cognitive modeling as well (see especially Horgan, 1997; Horgan & Tienson, 1996). Some ways of stating the 'frame problem' do restrict it to classical computational systems. But given that generalizations of the problem arise for other architectures, it seems likely that they arise for embodied simulators just as much as for connectionist nets, dynamical systems, and so on.

The dialectical situation, then, is as follows. It's true that perceptual vehicles make immediately available information for solving certain sorts of problems, and abstract propositional vehicles don't. It's also true that any adequate mechanistic account of the mind—whether computational or any other variety—needs to solve the frame or relevance problem, since *we* solve it (or seem to in many circumstances, at least). But the problem is not solved simply by choosing one representational format or another. So until embodied cognition theorists can show how a simulator might actually work such that it solves the problems of information retrieval and inference, they are not entitled to use considerations of representational format to support their position.

8. The function of linguistic comprehension

Let me now take a step back and put together the three TLC-based explanations laid out so far. The core of linguistic comprehension involves grasping the minimal truth conditions of a sentence. This process, for non-indexical and non-context sensitive sentences at least, is straightforwardly compositional: it relies just on the meanings of the words of the sentence and their syntactic and semantic mode of combination. Grasping a sentence's truth conditions gives us a propositional characterization of a scenario. We can then decide whether this scenario strikes us as a reasonable one or not, a process that may involve wide-ranging abductive inferences drawing on an array of relevant bodies of knowledge. Among this knowledge might be facts about our own bodily form and capabilities, as well as the ways in which we might interact with the objects in the scenario. The content we make reference to in coming to these judgments has to do with bodily/perceptual properties, affordances, and so on; but making these inferences does not involve redeploying our sensorimotor capacities in any essential way (although it is always open that they play a significant heuristic role). Finally, we can also use knowledge of what sentences mean to interact with our action plans. And, like abductively driven theoretical reasoning, this practical reasoning involves using a central code that is independent of the particular way that actions and perceptions themselves are represented.

The disagreement between ELC and TLC ultimately turns, I think, on differing conceptions of what we should require of a theory of linguistic understanding. The approach defended here supposes that this understanding is relatively thin: it includes not much more than the information needed to compute the truth conditions of context-free sentences. But Glenberg & Robertson (1999) seem to have a richer notion in mind. In an early test of their Indexical Hypothesis, they taught participants to use a compass to navigate and locate landmarks on maps. The training was either listening to the instructions, listening plus reading along, listening plus seeing pictures of the compass and map, or listening while seeing pictures and having salient parts indexed (pointed out). They were then given a verbal post-test of comprehension and a test of their ability to use the compass to find new landmarks.

Just listening to the instructions produces the worst performance on both the post-test and the navigation task. But listening while *reading* produces post-test performance comparable to listening and indexing.[11] The listen and read group was slower at navigation, but they made no more errors than the listen and index group, and they had never seen the equipment before the task itself. If we take comprehension to be measured by the ability to answer questions after encountering a body of text, both groups are equivalent. If we take it to be measured by how well one can interact with and use the described objects in a complex task, the listen and index group has better comprehension. This is clearly a *rich* notion of what is required for linguistic comprehension. It is unsurprising that rich comprehension is difficult to acquire from text alone—although given the equivalent error rates for the reading and indexing groups, it isn't as hard as it might seem.

Glenberg & Robertson (1999) take their results to support IH (and thus, ELC). The debate over embodied linguistic understanding is, then, fundamentally a debate over how rich a notion of linguistic competence we ought to adopt. The minimal notion of competence separates the ability to grasp the truth-conditional content of a sentence from the ability to judge how sensible, plausible, or easily visualized the scenario it describes is. Advocates of the minimal conception of understanding can readily accept that enactive simulation processes can (perhaps spontaneously) be recruited as consequences of this understanding. But that doesn't entail that they are part of the process of coming to understand a sentence itself. Given that separating these abilities is theoretically useful, the minimal notion has the advantage over the richer notion required by ELC.

However, the richer conception of linguistic understanding is sometimes motivated by functional considerations. These take the form of hypotheses about the purpose or function of linguistic understanding, and the role of the language faculty more generally in human cognition. For example, Glenberg (1997) proposes that 'the goal of language comprehension is the creation of a conceptualization of meshed patterns of action' (ibid., p. 13). Barsalou (1999a) also gives an evolutionary argument in favor of the functional claim that 'language comprehension prepares agents for situated action' (ibid., p. 62). Language, he claims, evolved for the purpose of facilitating social coordination. On this view, language aids social coordination insofar as it is useful 'to establish shared beliefs about the environment', 'for describing actions to perform on the environment', and 'to specify roles in groups, thereby implementing a division of labor' (ibid., p. 65). Given this evolutionary

[11] Interestingly, listening while seeing pictures produces post-test performance that is only slightly better than the listen-only group's. It's unclear why this is so, if seeing these objects provides rich visual information that enriches understanding. Moreover, the listen and picture group was as slow on the navigation task and made as many compass errors as did the listen-only group. These results are hard to reconcile with the general thrust of ELC, which claims that sensorimotor representations should, all things considered, facilitate performance on these tasks.

function, we should expect that the outputs of the comprehension process proper 'include actions on objects, interactions with other agents, information seeking, and bodily reenactments, such as manual and facial gestures' (ibid., p. 62).

Suppose, then, that this claim is true:

(F) The function of language is to create enactive simulations.

Claim (F), in turn, would support either ELC_s or ELC_m. Recall that the distinction between the strong and medium views is whether linguistic understanding is *entirely* a matter of enactive simulation, or whether understanding merely *requires* it. These correspond to two different ways of carrying out the function in question. Either language itself is wholly a device for producing enactive simulations, or it is at a minimum such a device.

I take it that the previous arguments against ELC_s and ELC_m also cast doubt on claim (F) itself. However, there are also independent reasons to think that, at the very least, it lacks supporting evidence. Such functional claims can be read in two ways: either they describe one function among many that language might have, or they describe a function that is uniquely performed by language (or for which language is uniquely well suited). Read in the former sense, these functional claims are innocuous enough. It's indubitable that in a lot of circumstances being able to understand what a sentence means can induce a person to imagine what it would be like to be present from a certain perspective in the circumstances described. Perhaps one could also make some predictions about how it might be best to act in those circumstances. If the question is whether language at least some of the time can function in these ways, the answer is yes. But this falls far short of arguing for a version of claim (F) that is strong enough to support either ELC_s or ELC_m, since it does not show any constitutive link between enaction and linguistic comprehension. In general, claims that language can be used to carry out a certain task/function are rather mild, since language is like many other cognitive systems in having a large range of possible uses to which it can be put. Once you have it, it comes in handy in any number of ways.

What about the claim that language is uniquely well suited to producing enactive simulations? Again, this seems difficult to sustain. The primary claim of ELC is that the vehicles of linguistic meaning are sensorimotor representations that are constructed under the guidance of syntactically structured linguistic input. But the process of constructing these simulated experiences or patterns of actions does not rely on any specifically linguistic skill, beyond the ability to generate a syntactic parsing of the input (an ability that ELC does not attempt to account for). Rather, it relies on otherwise well attested capacities for creating mental images, including not just visual and auditory images, but also images of bodily location, orientation, and movements. These capacities are substantially language-independent, and have wide application in non-linguistic domains. Rather than language filling a unique role in producing these simulations, language on the ELC account seems to piggyback on these pre-existing capacities. So the claim that language is a system that functions uniquely to create off-line enactive simulations seems misplaced as well.

Finally, Barsalou's evolutionary considerations in support of claim (F) are far from decisive. Of the three ways of achieving social coordination that he lists, only one—describing actions—is directly tied to sensorimotor representation. Coordinating beliefs about the environment is, again, something that is neutral concerning the format in which those beliefs are encoded. Many beliefs about the environment are not about perceivable or afforded properties of surrounding entities. Religious beliefs provide one such example. Interestingly, although Barsalou proposes it as a separate function of language, specifying roles in groups is another. Social relations (e.g. dominance and familial relations) are not directly perceivable, and they are marked in a wide variety of ways in different groups. Discovering how they are indicated in one's culture requires inference from behavioral signs and culturally determined markers. While these three properties may well be part of the evolutionary function of language, they do not obviously require the redeployment of sensorimotor representations, and so don't provide support for the claim that the output of comprehension is enactive simulation. I conclude that language does not play a distinctive cognitive role that would link it with sensorimotor capacities in the way required by ELC.

9. Conclusion

The traditional view of language has the resources to answer the criticisms leveled by proponents of embodied cognition. I should also stress, though, that adopting TLC does not mean taking the view that understanding words and sentences does not involve perception or action at all. To the contrary, information received through linguistic channels might be broadcast promiscuously to other cognitive systems capable of making use of it. Our capacity to create enactive simulations, then, is probably engaged spontaneously by our simultaneous processing of language. The view I have defended here combines some insights from the weakest form of embodied cognition (ELC_w), which is perfectly compatible with the basic claim of TLC, namely that the representational resources of the language faculty are amodal. This could be so even if certain tasks involving linguistic interpretation spin off subsidiary processes that tap sensorimotor faculties.

On this view, linguistic tasks involve many separate cognitive systems, just as embodied cognitive scientists have emphasized. But this processing fundamentally relies on prior processing by the dedicated language system. Messages are decoded into a separate propositional format, then passed on to other systems for separate elaboration. This joint processing of information in distinct sensory and semantic systems might have a number of advantages. It might, for instance, enable us to both reflect on the content of the message being transmitted to us and rapidly entertain possible ways we might respond to it. However, we shouldn't confuse the enactive simulation of these scenarios and our responses to them with actually grasping the information conveyed by the sentence itself. Grasping that information relies on a different set of inferential skills than does entertaining a perceptual simulation. In short, TLC is right about the structure of the language faculty proper. But ELC_w is right about the place of the language faculty in the larger cognitive system. The two are quite compatible—indeed, complimentary. While I doubt whether any advocates of TLC have denied that language is tightly integrated with other cognitive faculties, they have not often spelled out these connections in much detail. It is a virtue of the embodied research program that it can help to fill in these gaps.

Generally speaking, the claim that the semantic or conceptual system uses an amodal representational code is perfectly consistent with the claim that the semantic properties of words are *influenced* by input from sensorimotor systems. This influence might take the form of a translation or mapping from one domain of representations to another. Landau & Jackendoff (1993), for instance, develop an extensive analysis of the ways in which the spatial representation system interacts with language, particularly the closed-class system of prepositions. They suggest that otherwise puzzling properties of spatial prepositions (e.g. figure-ground asymmetries) can be explained in terms of an underlying geometric system for describing objects, regions, shapes, and paths. But these geometric representations themselves are not part of the semantic system; rather, they are translated into a different

representational format for use in language and non-spatial cognition more generally.

One motivation for embodied approaches to cognition is that in explaining behavior we need not posit any more representational capacities than are already present in our relatively complex sensory systems. This is a kind of parsimony argument. The traditionalist rejoinder to such claims is that we ought to expect that a creature's representational capacities will be shaped by the kinds of tasks that they have to undertake. Representations are like tools. Where they are aimed at doing a different sort of job, they will be shaped differently. Given that the job of representing communicated information poses different demands than the jobs of presenting information about the surrounding world or moving one's muscles in a certain way, one would expect language (and the conceptual system more broadly) to employ a distinct encoding system. So on the one hand we have an argument from parsimony, on the other hand an argument from task-adaptiveness.

These considerations arise at a fairly abstract level. Neither one is decisive in itself. Only careful attention to a range of empirical details and broader theoretical concerns can settle the question. I've argued that the evidence that seems to support the strongly embodied view does so only at the price of ignoring two crucial distinctions: first, between linguistic competence and the cognitive abilities that are contingently linked with it; and second, between content and the vehicles that carry it. Attending to these distinctions shows that, while language use in practical contexts may tap sensorimotor capacities, linguistic understanding as such does not. Linguistic and perceptual representation might be closely linked, but not so intimately intertwined that they should be identified. At the core, the language faculty is not embodied.

Acknowledgements

Thanks to audience members at Cognition: Embodied, Embedded, Enactive, Extended (Orlando, FL, October 2007), especially Fred Adams and Andy Clark, for their comments and discussion on an earlier version of this paper. Thanks also to all of the participants at Computation and Cognitive Science (King's College, Cambridge, July 2008) for much useful feedback as well. I am especially grateful to Mark Sprevak for his editorial advice, and for his efforts in putting together both the conference and this special issue.

References

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence, 149*, 91–130.
Barsalou, L. W. (1999a). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes, 28*, 61–80.
Barsalou, L. W. (1999b). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609.
Borg, E. (2004). *Minimal semantics*. Cambridge: Cambridge University Press.
Borreggine, K. L., & Kaschak, M. P. (2006). The action–sentence compatibility effect: It's all in the timing. *Cognitive Science, 30*, 1097–1112.
Clark, A. (1997). *Being there*. Cambridge, MA: MIT Press.
Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Science, 10*, 370–374.
Dretske, F. I. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
Gallagher, S. (2005). *How the body shapes the mind*. Oxford: Oxford University Press.
Gardner, J. (1985). *The art of fiction*. New York: Vintage.
Gibbs, R. W., Jr., (2003). Embodied experience and linguistic meaning. *Brain and Language, 84*, 1–15.
Gibbs, R. W., Jr., (2006). *Embodiment and cognitive science*. Cambridge: Cambridge University Press.
Gibbs, R. W., Jr., Beitel, D., Harrington, M., & Sanders, P. (1994). Taking a stand on the meanings of 'stand': Bodily experience as motivation for polysemy. *Journal of Semantics, 11*, 231–251.
Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.
Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*, 1–55.
Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558–565.
Glenberg, A. M., & Kaschak, M. P. (2004). Language is grounded in action. In L. Carlson, & E. van der Zee (Eds.), *Functional features in language and space: Insights from perception, categorization, and development* (pp. 11–24). Oxford: Oxford University Press.
Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes, 28*, 1–26.
Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43*, 379–401.
Goldman, A. (2006). *Simulating minds*. Oxford: Oxford University Press.
Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.
Harnad, S. (1993). Symbol grounding is an empirical problem: Neural nets are just a candidate component. *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society* (pp. 169–174). Hillsdale, NJ: Erlbaum.
Higginbotham, J. (1989). Elucidations of meaning. *Linguistics and Philosophy, 12*, 465–518.
Horgan, T. (1997). Connectionism and the philosophical foundations of cognitive science. *Metaphilosophy, 28*, 1–30.
Horgan, T., & Tienson, J. (1996). *Connectionism and the philosophy of psychology*. Cambridge, MA: MIT Press.
Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.
Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language, 43*, 508–529.
Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M. E., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition, 94*, B79–B89.
Kaschak, M. P., Zwaan, R. A., Aveyard, M. E., & Yaxley, R. H. (2006). Perception of auditory motion affects language processing. *Cognitive Science, 30*, 1–12.
Katz, J. J. (1972). *Semantic theory*. New York: Harper and Row.
Katz, J. J., & Fodor, J. (1963). The structure of a semantic theory. *Language, 39*, 170–210.
Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
Landau, B., & Jackendoff, R. (1993). 'What' and 'where' in spatial language and spatial cognition. *Behavioral and Brain Sciences, 16*, 217–265.
Larson, R., & Segal, G. (1995). *Knowledge of meaning: An introduction to semantic theory*. Cambridge, MA: MIT Press.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Marconi, D. (1997). *Lexical competence*. Cambridge, MA: MIT Press.
Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
Prinz, J., & Barsalou, L. W. (2000). Steering a course for embodied representation. In E. Dietrich, & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 51–77). Mahwah, NJ: Erlbaum.
Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann, & W. Prinz (Eds.), *Relationships between perception and action* (pp. 167-201). Berlin: Springer.
Robbins, P., & Aydede, M. (Eds.). (2008). *Cambridge handbook of situated cognition*. Cambridge: Cambridge University Press.
Rumelhart, D., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland, & D. Rumelhart (Eds.), *Parallel distributed processing* (2 vols.) (Vol. 2, pp. 7–57). Cambridge, MA: MIT Press.
Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology, 3*, 552–631.
Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.
Shapiro, L. A. (2004). *The mind incarnate*. Cambridge, MA: MIT Press.
Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 13*, 153–156.
Weiskopf, D. A. (2007). Concept empiricism and the vehicles of thought. *Journal of Consciousness Studies, 14*, 156–183.
Wilson, M. A. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625–636.
Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. *The Psychology of Learning and Motivation, 44*, 35–62.
Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science, 28*, 611–619.
Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*, 168–171. IP29 5QE.