

Understanding is not simulating: A reply to Gibbs and Perlman

Daniel A. Weiskopf

Mailing address: Department of Philosophy, Georgia State University, P.O. Box 4089, Atlanta, GA 30302. Email: dweiskopf@gsu.edu

Abstract: In this response, I do four things. First, I defend the claim that the action compatibility effect does not distinguish between embodied and traditional accounts of language comprehension. Second, I present neuroimaging and neuropsychological results that seem to support the traditional account. Third, I argue that metaphorical language poses no special challenge to the arguments I gave against embodied theories of comprehension. Fourth, I lay out the architecture of language I advocate and suggest the sorts of data that would decide between traditional and embodied accounts.

Gibbs and Perlman (henceforth G&P) raise a number of pointed challenges to the case I laid out against the Embodied Language Comprehension (ELC) view and in favor of the Traditional Language Comprehension (TLC) view. I'm grateful for their thoughtful criticisms and hope we can make some progress towards clarifying the current state of the debate, as well as what future results might help to decide these thorny questions concerning the architecture of the language faculty and the larger cognitive system that it's part of.

1. What does ELC predict?

Whether the evidence supports ELC depends on what predictions it makes relative to its competitors. G&P charge that I have a mistaken view about what sorts of experiential simulations ELC posits, which in turn leads me to misinterpret the evidence in favor of ELC. In the case of the action compatibility effect (ACE), I argued that it was hard to see why one would expect that sentences describing motion towards oneself would prime a motor response towards oneself, since the natural motion one would make in such a transfer situation involves motion away from oneself—an outstretched hand, for example.

G&P reply that “a person simulates the meaning of the sentence... not how that person may possibly respond to the meaning of the sentence once it has been interpreted” (p. 4). Meanings here are just identified with perceptually imagined situations. Glenberg seems to disagree; his Indexical Hypothesis states that simulating afforded actions is part of the comprehension process itself. It temporally follows simulating the perceived situation, but it is still part of what happens in understanding proper. And Zwaan’s Immersed Experiencer Framework claims that both perception and action are simulated as part of understanding. These might seem to be terminological disputes, but they aren’t. The real issue is where to draw the line between processes that are part of understanding a sentence and processes that are subsequent to understanding, and hence not part of the linguistic system itself.

Even if one grants that the simulations that are part of understanding proper are simulations of the *perceived* situation, not how one would *act* in it, the puzzle I raised still applies. The standard explanation of priming effects is that they arise from the re-activation of a representation that was active earlier: either the very same one again, or one associated in memory with it. But here we have a case where a visually imagined path starting with another

person and terminating at me primes a motor command to move my hand in a direction that follows the same path. There are only two ways in which priming could explain this: (1) perception of motion and motor commands at least partially share a representational code, or (2) there is a common code at a level of abstraction above both perception and action.

The ACE evidence doesn't distinguish between (1) and (2). It's compatible with participants having in mind (as part of the experimental setup) that they need to move their hands towards themselves, and then *inferring* from what was said ("Josh gave me the notebook") that the notebook moved from Josh to me. It's the occurrence of the same representation of the motion-path in the latter representation and in one's standing intention to move one's hand that produces facilitation in this case. This account doesn't refer to embodied simulations at all, but would equally predict the ACE results.

This doesn't show that ELC is false, of course, just that the evidence is equivocal. The amendment G&P offer, drawing on cognitive grammar, seems not to help here; indeed, it seems to make things worse. They suggest that what we simulate is influenced by the thematic roles encoded in the sentence; in particular, we simulate the event from the perspective of the AGENT of the sentence. So I simulate being Josh (or just being in Josh's place), extending my arm to give "me" the notebook.¹ But in this case, both the direction of perceived motion and the accompanying motor command encode a path leading *away* from the egocentrically recentered perspective I occupy in the simulation. We would then expect this sentence to prime movement away from my real-world body, but what we find is the opposite, namely that this sentence

¹ A fascinating study by Markman & Brandl (2005) suggests that it is not one's physical location that is important for priming, but where one represents one's *self* as being. They interpret this as showing the need for higher-level symbolic representations in addition to sensorimotor representations in explaining motion-congruency results.

primes movement towards myself. My taking Josh's perspective in the simulation won't explain this.²

G&P also charge that I underestimate the *richness* of what is simulated. I argued that (for example) waiting on a corner is perceptually indistinguishable from merely standing on a corner, hence perceptual simulations are sometimes more coarse-grained than truth-conditions. They respond that simulations may include, in addition to perceptual appearances, "a sense of the man's intentions, expectations, and even emotions" (p. 5). If emotional states have a significant bodily phenomenology (Damasio, 1999; Prinz, 2004), that phenomenology might be part of the content of simulations. Still, the difference between standing on the corner and waiting on the corner still does not seem to be something that can be captured in terms of bodily plus sensorimotor phenomenology. The reason is that the two can doubly dissociate: waiting need not give rise to any sort of anxious or impatient bodily sensations, and one can merely stand around feeling just as anxious as if one were waiting for something. Adding bodily sensations to sensorimotor ones doesn't capture the difference in meaning between standing and waiting.

What does capture it? Intentions and expectations do. When one waits, one waits *for* something. So why can't one just simulate *anticipating* the bus, one's girlfriend, etc.? On this model, in understanding "The man waited on the corner for the bus", one straightforwardly simulates *waiting*; that is, one imaginatively places oneself in a state relevantly like that of someone waiting for the bus. Glenberg and collaborators have taken this step and extended the notion of simulation to include emotional states as well as sensorimotor ones (Havas, Glenberg, & Rinck, 2007). They induced participants to either smile or frown while judging the emotional

² I also disagree with the claim that (1) "Josh gave me the notebook" and (2) "I took the notebook from Josh" are two ways of describing the same event. Their truth conditions differ, for one thing. (1) may entail (2), but the converse isn't so. I might take Josh's notebook as one takes candy from a baby, a situation in which the victim most definitely does not also *give* me what I've taken; hence the loud protests that often ensue in asymmetrical taking/giving scenarios.

valence (positive or negative) of sentences. Individuals in the smiling condition were faster to make judgments about positive sentences, while frowners were faster with negative sentences, as would be predicted, since smiling or frowning influences emotional state.

While this might seem to support an emotional simulation account of understanding, it is compatible as well with a TLC-style account on which emotional states tend to activate the relevant emotion concepts, and the activation of these concepts primes sentence-level responding. See the discussion of Wilson & Gibbs (2007) in Section 3 for further details and an example of how this sort of explanation might be applied to action simulation in metaphor comprehension.

2. Further empirical evidence

As G&P note, recent neuroimaging studies provide converging evidence that seems to support ELC. That there is a link between motor activation and action verbs is indisputable. However, as Willems and Hagoort (2007) say in a recent review, it is difficult to tell whether this shows that motor representations are part of semantic representation proper or part of post-comprehension imagery processes.³

For example, in one fMRI study, Boulenger, Hauk, & Pulvermuller (2009) asked participants to read both literal and idiomatic sentences that involved verbs describing movement of the arms or the legs (e.g., “John grasped the idea” vs. “John grasped the object”). The activation produced in these conditions was compared to that performed when viewing a meaningless string of characters, and to the activity produced during motion of the hand and foot. One key result among many is that reading both literal and idiomatic sentences is followed

³ I have argued elsewhere (Weiskopf, 2007) that we cannot simply assume that all regions activated during a conceptual processing task are part of the neural representation of the concept itself. These arguments apply as well to neural studies of lexical semantics.

by activation in the part of motor cortex involved in producing the relevant movements.

However, as the authors also note, this activation occurs relatively late in the process—about 3 second after the sentence ends. This suggests that motor activation is a post-comprehension process (Boulenger et al., p. 1913), which is what the TLC model would predict: simulation is one mechanism used to elaborate on our understanding of what is said, but it doesn't constitute that understanding.

Moreover, it is sometimes unclear what should count as 'motor activation' in these studies. Postle, McMahon, Ashton, Meredith, & de Zubicaray (2008) ran a complex series of tasks intended to localize these regions more precisely. Participants underwent fMRI scans as they first read a series of words (action verbs involving the hand, leg, and mouth; unrelated nouns; nonwords; and hashes), viewed short videos of actions involving the hand, leg, and mouth, and executed brief actions involving those body parts (forming a fist, wiggling one's tongue, etc.). The regions of interest (ROIs) activated by each task were then projected onto cytoarchitectural maps for canonical motor areas BA4 (primary motor cortex) and BA6 (premotor cortex).

Relative to other studies, the results were surprising. Action execution and observation, as predicted, produced somatotopically organized activity in BA4 and BA6. Action words, however, failed reliably to show activation in those regions. This is strikingly not what ELC would predict. In response, the authors determined the peak ROI common to the execution and observation tasks and assessed its selectivity to action words. This region fell within the pre-supplementary motor area (pre-SMA). The pre-SMA is often taken to have a hybrid cognitive and motor function, and in particular to be connected with modality-independent and effector-independent maintenance of task-relevant information (Pickard & Strick, 2001). In other words,

the main association of action verbs in this study was with an area that seems to be involved in retrieving and maintaining information needed for behavioral planning, as well as associative learning. This is closer to the classic role of amodal representations, namely the orchestration of how various low-level processes are to be deployed. These results are generally inconsistent with the emphasis by ELC on the re-use of specific low-level sensory and motor systems.

Chronometric studies and more fine-grained imaging results are important for testing the predictions of TLC and ELC, but lesion studies (and studies of selective inactivation via, e.g., repeated transcranial magnetic stimulation) provide one of the best possible sources of evidence, since ELC predicts that deficits in sensorimotor capacity should entail deficits in linguistic comprehension, while TLC predicts that there should be no such deficits (or that they should be much less profound). A direct test of this hypothesis was performed by Negri, Rumiati, Zadini, Ukmar, Mahon, & Caramazza (2007). They studied patients with ideational apraxia, a deficit in using objects not based in impairments to basic sensory and motor functions, by giving them a sequence of tasks intended to tap linguistic and motor performance: object recognition (naming), recognition of pantomimed actions performed without the use of objects (e.g., hammering motions without the use of a hammer), imitation of pantomimes without objects, appropriate object use, and imitation of intransitive actions. Among their findings were double dissociations between object use and object naming, and between object use and pantomimed action recognition. So some patients can successfully name objects but not use them, and others can use them but not describe the behaviors characteristic of their use, as well as the reverse, in each case. This, again, would not be predicted if successfully naming an object always required retrieving motor representations corresponding to its proper use, as ELC entails.

I don't suggest that these studies provide conclusive evidence against ELC, although the double dissociations in particular are a striking anomaly that need to be explained. They are, however, consistent with the TLC perspective on the relationship between comprehension and simulation. I will say more about this relationship in Section 4.

3. Metaphor and nonliteral meaning

I will only make a few brief comments on metaphor and nonliteral meaning. Let me state that I'm not at all committed to the claim that on-line sentence comprehension first involves computing a "context-free, semantic, literal representation" (p. 10). On-line comprehension is always aimed at understanding in a particular context of use, and this purpose is plausibly best served by computing what is said by a speaker making an utterance, rather than the meaning of the sentence (Recanati, 2004). So intrusions of pragmatic or background knowledge in processing, even at early stages, is compatible with my view.

I do think, however, that we have the capacity to consider the meanings of sentences taken apart from their possible contexts of use, and that this capacity consists at least in part of our having a system that can compute their semantic representations. If we didn't have such an ability, it's hard to see how the enterprise of semantic theorizing could get off the ground in the first place. Semantics takes as its data the fact that sentences have certain logical properties and relations such as entailment, compatibility, contradiction, thematic relations, synonymy, ambiguity, and so on (Chierchia & McConnell-Ginet, 2000; Dowty, Wall, & Peters, 1980; Heim & Kratzer, 1998; Larson & Siegel, 1995), and semantic theories aim to characterize the information that this system encodes. But the system need not be modular (in Fodor's sense), and the process of comprehension need not be serial.

Metaphorical language and thought are two of the most important and difficult aspects of cognition to understand, and it's not at all clear that there is just one thing going on that explains and unifies all of the phenomena in these domains (Camp, 2006). But suppose that metaphorical language is like literal language in the following respect: comprehending it also triggers, and is influenced by, ongoing sensorimotor activity. In this case, as far as I can see, the arguments in favor of TLC that I've given go over in the same way. For example, Wilson & Gibbs (2007) showed that imagining or carrying out an action can speed comprehension of metaphorical sentences that include a verb denoting the action. One possibility is that motor activity that is characteristically associated with a certain concept—CHEW or KICK, say—primes that concept, and the priming of the concept facilitates comprehension of sentences containing the key verb. One might object that participants were had difficulty producing the precise verb included in these sentences when asked to describe the actions separately, which Wilson & Gibbs take as evidence that the effect is not lexically mediated. But, first, mappings from semantic/conceptual representations onto lexical items are not generally unique, particularly for complex actions as opposed to objects; and, second, the effect here is not lexical but conceptual, so failures of lexical retrieval are not dispositive. This explanation is consistent with both the data and the TLC account, and it would also apply to the emotional priming observed by Havas, Glenberg, & Rinck, (2007).

Obviously this is only a sketch, in need of greater fleshing out. The main point remains, however: if metaphorical language and literal language are processed alike, then there is no need for a separate argument against ELC in the domain of metaphor.

4. The structure of the linguistic system

I'll now briefly state what I take the relationship to be between the linguistic system(s) and the systems responsible for our abilities to produce simulations of various kinds. The TLC architecture is laid out in Figure 1.

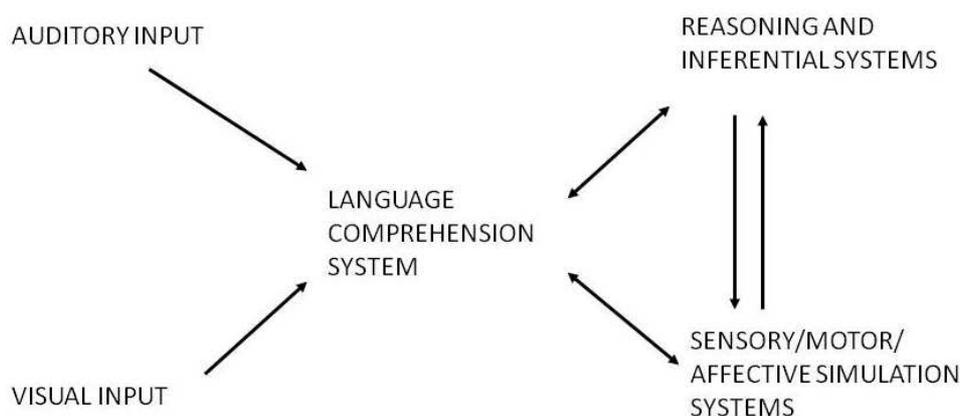


Figure 1: The place of the language comprehension system vis-à-vis other cognitive systems

Obviously, much detail is being suppressed in this analysis. Language can be either heard or seen, but this figure omits how we determine that auditory or visual input should be interpreted by the linguistic system. Similarly, the internal structure of the language comprehension system is entirely left out, but it will need to be extremely complex, including at least mechanisms of phonological interpretation, graphemic interpretation, syntactic analysis, morphological analysis, and semantic analysis. The precise decomposition of this system into interacting subcomponents is not the present concern; what matters is the output the system produces, namely a semantically interpreted representation of the input string's meaning (if it is a meaningful sentence of the language) or an error signal (if it is not). That is what the system itself, taken in isolation, is capable of producing.

However, in interaction with other systems, it is capable of doing more. As I've noted, by interacting with general world knowledge, including knowledge of the current conversational

context and the speaker's intentions, it is capable of producing a contextually appropriate representation of what is being said by the utterance. That is why, in on-line processing, one typically generates utterance meanings, as G&P point out. Moreover, the output of the language system is passed immediately to other cognitive systems for further elaboration and processing. It is sent to general purpose reasoning systems in order to generate useful inferences to guide further conversation and decision-making. It is also sent to various simulation subsystems that produce sensory, motor, and affective (as well as possible higher cognitive) simulations that bear on the proposition that was expressed by the input sentence. These simulations, as ELC advocates have quite correctly noted, play an important role both in guiding future behavior and in helping to produce new inferences that cannot easily be generated by amodal reasoning mechanisms. And information produced by these subsidiary systems may in turn feed back into the linguistic system to guide ongoing interpretation processes.

So what one sees is a dynamic cycle of activity from the linguistic system to other cognitive systems and back again. But it is important to keep in mind that these are *distinct* systems. Mahon & Caramazza (2008) have laid out in greater detail the case for this cognitive division of labor, as well as presenting arguments similar to the ones I give against ELC.⁴ On this sort of architecture, one would *expect* to see sensorimotor activity produced as a result of, and recurrently influencing, comprehension processes. But one would generally predict this activity to succeed comprehension proper, and to be causally independent of it: that is, comprehension can take place without simulation, and vice-versa. And as the studies surveyed in Section 2 suggest, we do observe this pattern.

⁴ Their discussion appeared in print after my own had been drafted and submitted, but I am gratified to see the convergence in our interpretations of the data, as well as in the architecture that we endorse. I refer readers looking for further arguments in favor of TLC to their paper.

The moral of this discussion is that further research will need to focus on how to distinguish the predictions of ELC from those of TLC. Behavioral, imaging, and lesion data are all relevant here, with the key issues being the time course of comprehension and simulation processes, the precise neural regions activated during each process, and their potential dissociability. In addition, a theory of linguistic competence needs not only to mesh with a full account of the mechanisms of on-line processing, but also to explain our ability to draw the sorts of inferences that semanticists have concerned themselves with. The resources of TLC in this respect are considerable, and the case for ELC remains far less certain than its proponents often claim.

References

- Boulenger, V., Hauk, O., & Pulvermuller, F. (2009). Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex*, *19*, 1905-1914.
- Camp, E. (2006). Metaphor in the mind: The cognition of metaphor. *Philosophy Compass*, *1*, 154-170.
- Chierchia, G., & McConnell-Ginet, S. (2000). *Meaning and Grammar: An Introduction to Semantics* (2nd ed.). Cambridge: MIT Press.
- Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Basic Books.
- Dowty, D. R., Wall, R. E., & Peters, S. (1980). *Introduction to Montague Semantics*. Dordrecht: D. Reidel.

- Havas, D. A., Glenberg, A. M., & Rinck, M. (2007). Emotion simulation during language comprehension. *Psychonomic Bulletin & Review*, *14*, 436-441.
- Heim, I., & Kratzer, A. (1998). *Semantics in Generative Grammar*. Malden, MA: Blackwell.
- Larson, R., & Segal, G. (1995). *Knowledge of Meaning*. Cambridge: MIT Press.
- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology - Paris*, *102*, 59-70.
- Markman, A. B., & Brandl, C. M. (2005). Constraining theories of embodied cognition. *Psychological Science*, *16*, 6-10.
- Negri, G. A. L., Rumiati, R. I., Zadini, A., Ukmar, M., Mahon, B. Z., & Caramazza, A. (2007). What is the role of motor simulation in action and object recognition? Evidence from apraxia. *Cognitive Neuropsychology*, *24*, 795-816.
- Pickard, N., & Strick, P. L. (2001). Imaging the premotor areas. *Current Opinion in Neurobiology*, *11*, 663-672.
- Postle, N., McMahan, K. L., Ashton, R., Meredith, M., & de Zubicaray, G. I. (2008). Action word meaning representations in cytoarchitecturally defined primary and premotor cortices. *NeuroImage*, *43*, 634-644.
- Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.
- Recanati, F. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.
- Weiskopf, D. A. (2007). Concept empiricism and the vehicles of thought. *Journal of Consciousness Studies*, *14*, 156-183.

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language, 101*, 278-289.

Wilson, N. L., & Gibbs, R. W., Jr. (2007). Real and imagined body movement primes metaphor comprehension. *Cognitive Science, 31*, 721-731.