frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 22 November 2010
doi: 10.3389/fpsyg.2010.00197

A connectionist approach to embodied conceptual metaphor

Stephen J. Flusberg*, Paul H. Thibodeau, Daniel A. Sternberg and Jeremy J. Glick

Department of Psychology, Stanford University, Stanford, CA, USA

Edited by:
Anna M. Borghi, University of Bologna, Italy

Reviewed by:
Elisa Magosso, University of Bologna, Italy
Giovanni Pezzulo, Institute of Cognitive Sciences and Technologies, National Research Council, Italy

***Correspondence:**
Stephen J. Flusberg, Department of Psychology, Stanford University, Jordan Hall, 450 Serra Mall, Building 420, Stanford, CA 94305, USA.
e-mail: sflus@stanford.edu

A growing body of data has been gathered in support of the view that the mind is embodied and that cognition is grounded in sensory-motor processes. Some researchers have gone so far as to claim that this paradigm poses a serious challenge to central tenets of cognitive science, including the widely held view that the mind can be analyzed in terms of abstract computational principles. On the other hand, computational approaches to the study of mind have led to the development of specific models that help researchers understand complex cognitive processes at a level of detail that theories of embodied cognition (EC) have sometimes lacked. Here we make the case that connectionist architectures in particular can illuminate many surprising results from the EC literature. These models can learn the statistical structure in their environments, providing an ideal framework for understanding how simple sensory-motor mechanisms could give rise to higher-level cognitive behavior over the course of learning. Crucially, they form overlapping, distributed representations, which have exactly the properties required by many embodied accounts of cognition. We illustrate this idea by extending an existing connectionist model of semantic cognition in order to simulate findings from the embodied conceptual metaphor literature. Specifically, we explore how the abstract domain of time may be structured by concrete experience with space (including experience with culturally specific spatial and linguistic cues). We suggest that both EC researchers and connectionist modelers can benefit from an integrated approach to understanding these models and the empirical findings they seek to explain.

Keywords: connectionism, models, embodiment, conceptual metaphor, time, space

INTRODUCTION

In recent years, a growing body of data has been gathered in support of the idea that the mind is situated and embodied and that cognition is grounded in sensory-motor interactions with the world (Varela et al., 1991; Clark, 1998; Barsalou, 1999, 2008; Lakoff and Johnson, 1999; Gibbs, 2006; Spivey, 2007; Chemero, 2009). The guiding tenet of the embodied cognition (EC) movement holds that cognitive processes are shaped and structured by the fact that an agent has a particular kind of body and is embedded in a particular kind of environment. Crucially, the effects of embodiment can and should be observed at all levels of cognitive processing, from vision and memory (Glenberg, 1997; Noë, 2004; Proffitt, 2006), to emotion and action perception (Rizzolatti and Craighero, 2004; Niedenthal et al., 2005), to language and abstract thought (Barsalou, 1999; Lakoff and Johnson, 1999; Feldman, 2006; Gibbs, 2006; Barsalou, 2008). It has been argued that this "body-up" approach to cognition poses a serious challenge to more traditional "mind-down" approaches in cognitive science (Lakoff and Johnson, 1999; Spivey, 2007; Barsalou, 2008; Chemero, 2009), which have attempted to define cognition in terms of discrete, amodal, symbolic information-processing mechanisms divorced of any particular physical instantiation (Fodor, 1975; Marr, 1982; Kemp and Tenenbaum, 2008).

This debate has been particularly contentious in discussions of high-level cognition, where the amodal symbolic view has typically dominated. As a result, the embodiment of metaphor and abstract thought has become one of the most hotly researched, discussed, and debated issues within cognitive science (Lakoff and Johnson, 1980, 1999; Gibbs, 1994, 1996, 2006; Murphy, 1996, 1997; Boroditsky, 2000; Boroditsky and Ramscar, 2002; Feldman, 2006; Pinker, 2007). Lakoff and Johnson (1980, 1999) famously observed that natural language is exceedingly figurative. When we talk about complex or abstract topics, we rely heavily on systems of metaphors, borrowing words and phrases from other, more concrete domains. For example, to talk about *theories*, people often rely on *building* metaphors. Indeed, theories must have a *solid foundation* and be *well-supported* by the data or they might *fall apart*, and you can *build them up, tear them down*, or even *explode them* in light of new findings.

While traditional theories of language treat metaphor as mere ornamental flourish (e.g., Grice, 1975; Searle, 1979; Pinker, 2007), Lakoff and Johnson (1980, 1999) argue that metaphor is not simply the way we *talk* about abstract things, but how we *think* about them as well. On this view, we understand and reason about abstract domains like theories, time, and love through our concrete, embodied experiences (e.g., of interacting with physical buildings). Thus, our perceptual and motor experiences actually structure our ability to engage in abstract thinking. Empirical demonstrations of embodied metaphor have taken the form of experiments showing that activating a concrete source domain (e.g., space) influences responses and inferences in the abstract target domain (e.g., time; Boroditsky and Ramscar, 2002; Casasanto and Boroditsky, 2008; Jostmann et al., 2009; Ackerman et al., 2010).

One important challenge facing researchers is to account for this view of metaphorical thought at a more precise, mechanistic level of description (Murphy, 1996, 1997; Barsalou, 2008). This may be particularly problematic because EC is not a singular, unified framework, but rather a collection of heterogeneous viewpoints that may be only loosely related to one another in terms of theoretical commitments and empirical investigation (Wilson, 2002; Ziemke, 2003; Gibbs, 2006; Barsalou, 2008; Chemero, 2009). In addition, because these competing perspectives are commonly described only verbally, it can be difficult to use them to generate the precise predictions that might allow us to directly compare them (but see Lakoff and Johnson, 1999; Feldman, 2006).

Taking a computational modeling approach may provide a potential remedy to these issues. The development of specific, simplified models can help researchers understand complex cognitive processes at a level of detail that theories of EC have sometimes lacked (see, e.g., Broadbent, 1987; Smolensky, 1988; Hintzman, 1991; Seidenberg, 1993; Barsalou, 1999, 2008; Spivey, 2007; McClelland, 2009). The process of constructing a model differs from a verbally described theory in that it forces the researcher to commit at least temporarily to a particular *internally consistent* instantiation of the environment and the agent that acts within it. As a result, computational models can make precise predictions that can be tested empirically. Grounding empirical findings in terms of a model and making principled modifications to that model in order to accommodate these findings can help researchers explore and clarify ideas (McClelland, 2009). In addition, because models can often reveal principles that underlie a given set of phenomena, modeling frameworks can sometimes help unify various areas of empirical inquiry (Estes, 1955; Rescorla, 1988; McClelland et al., 1995; Ramscar et al., 2010). This special topic of Frontiers in Cognition is evidence that more researchers are starting to take computational modeling seriously as a method for exploring the principles and mechanisms that support EC (see Spivey, 2007 for a call to arms on this issue).

At the same time, the findings from EC outlined above provide computational modelers the opportunity to look to for evidence of the ways in which cognition naturally unfolds in a real, embodied agent (for a recent review, see Barsalou, 2008). This will strongly influence not only the details of the model environment, but also the choice of the learning problem to be solved by the model. Modelers focused on understanding learning processes should attend to the fact that the information reaching the cognitive system is always structured by the relationship between the organism and its environment, which may lead to surprising new ways of thinking about everything from visual perception (Noë, 2004) to semantics (Barsalou, 1999).

The present paper has both a narrow and a broad goal. The narrow goal is to capture the effects of embodied conceptual metaphor using a connectionist model. In lieu of instantiating a particular EC theory of metaphor, we repurpose an existing connectionist model of semantic cognition (Rogers and McClelland, 2004) to explore how our experience of space can structure how we think and reason about time. This approach may be especially fruitful because it promises to bring together more established modeling principles with the novel findings from EC.

Our network receives direct experience with both space and time in its simplified environment, including experience that is analogous to the use of linguistic or cultural cues. However, the network's experience in the spatial domain is more richly structured than its experience in the temporal domain, in much the same way that we can freely move around and interact with our spatial, but not temporal, environment. Because the model is sensitive to the ways in which the structure of time is similar to the structure of space, it develops representations of time that are partially constituted by its knowledge of space. Therefore, even in the absence of direct co-occurrence of space and time during learning, the network is able to exploit this structural similarity to draw inferences about temporal events by using what it knows about space. This demonstrates a novel learning mechanism that operates over the course of development and gives rise to deeply metaphorical semantic representations, which may serve as a tractable implementation of existing theories of metaphorically structured thought (e.g., Lakoff and Johnson, 1980; Boroditsky, 2000).

The broad goal of this paper is to serve as an example of how computational models and EC can reciprocally inform one another. In particular, we make the case that connectionist architectures can help explain many surprising results from the EC literature (for related views, see Bechtel, 1997; Clark, 1998; Spivey, 2007). Crucially, our model focuses on the learning process and forms overlapping, distributed representations, which have exactly the properties required by many embodied accounts of cognition. In particular, these representations, together with the learning process, support the integration of experience from multiple modalities, including perceptual-motor, linguistic, and cultural information. At the same time, extending the scope of the model to incorporate insights from EC transforms our interpretation of the modeling approach more generally. This can lead to new ways of thinking about how to set up and investigate particular ideas within this modeling framework. Ultimately, we suggest that this integrative approach can serve as a unifying framework that may help drive future progress within cognitive science.

MATERIALS AND METHODS

GENERAL MODELING FRAMEWORK

The network can be thought of as an agent experiencing its world. Over the course of "training" the agent repeatedly experiences events in the world, predicts their outcomes, and learns something about how the actual events differ from its predictions. The environment and the agent are simplified so as to render the learning process tractable, while still retaining those aspects of environmental structure which are crucial for producing the phenomena the model is supposed to explain, and to make it possible to analyze what the agent has learned (for a discussion of this issue, see McClelland, 2009).

In this model, the environment consists of the various items in the world that the agent experiences in their various relational contexts (collectively forming the input patterns), together with the subsequent states of the world that the network attempts to predict (the target output patterns). The network that comprises the agent is wired up in a strictly feed-forward fashion, as shown in **Figure 1**. While we assume that in reality agents interact with the world in a dynamic fashion, for simplicity we consider only one portion of this dynamic interaction. On each trial, the agent experiences some portion of the world (e.g., that it is standing in a particular section of space and moving in a particular direction), makes a prediction about what it will experience next (e.g., that

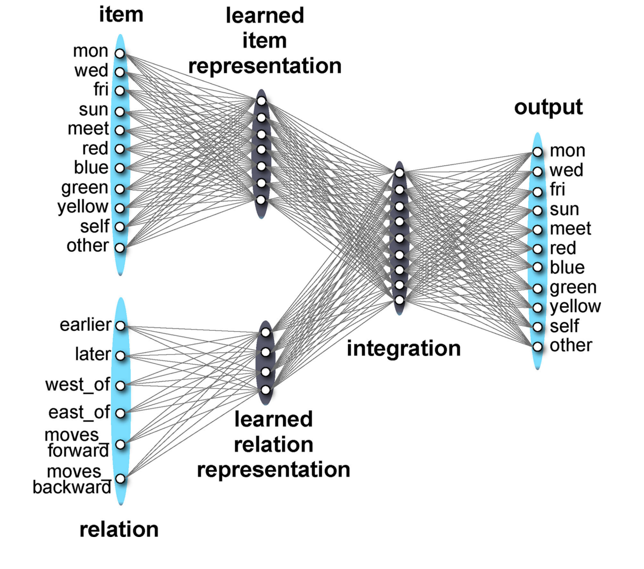

FIGURE 1 | Diagram of network architecture.

it will encounter another particular section of space), and learns about the ways in which it was incorrect, thereby improving future predictions.

The network's knowledge is stored in the weights between the layers. When a pattern of activation occurs across one of the layers, that activation propagates forward through the weights to the next layer. The patterns of activation at the input layers are thought of as multimodal sensory-motor input from the environment. In the *Item* layer, these inputs stand for the experience of physical locations in space and temporally extended events such as the days of the week or a meeting. In the *Relation* layer, the inputs stand for different kinds of relationships that these items can have to each other; for example, we might ask the network what day is *earlier than Wednesday*, or what section of space is *West of* the *blue* section.

While these layers consist of labeled units, they are best thought of as standing for distributed representations that were learned from other, lower-level (possibly modality-specific) patterns of perceptual-motor experience. This simplification does not strongly affect how the model works because the network is forced to create its own distributed representations of these perceptual inputs in the layers that immediately follow (see, e.g., Rogers and McClelland, 2004). In particular, the *Learned Item Representation* is a re-representation of the *Item* inputs, integrating all of the information it has learned across all relations to create a densely overlapping set of patterns that encode the structural regularities that hold between the items. The *Learned Relation Representation* serves the complementary function for the *Relation* inputs. Activation in these layers then propagates forward to the *Integration* layer. Here, information about the two input pathways is combined in a way that we presume is similar to how modality-specific information is integrated at earlier layers. This integrated representation is used to make a prediction about the target pattern, which is represented by activations of the *Output* layer. In the current model, the target pattern consists of another item (or set of items) that bears the appropriate relation to the input.

Initially, the network is instantiated with small random weights connecting each of the layers. As a result, its internal representations of all items and all relations will be similar, and therefore its predictions about the world will be the same for all inputs. Whenever the network's output fails to match the target pattern, however, it receives an error signal in proportion to the squared output error. This error signal informs the network both when it has predicted events that do not occur and when it has failed to predict an event that did occur. In practice, this error signal serves to adjust the weights from the inputs to the outputs in proportion to the error that they caused, using the standard backpropagation learning algorithm.

Since different input patterns predict different events "in the world," the network will gradually learn to differentiate the items from each other, and the relations from each other. This process of differentiation is driven by differences in what the various items predict about what else may happen in the world, not directly by, for example, the degree of overlap in the perceptual inputs (for related views, see Gibson and Gibson, 1955; Rogers and McClelland, 2008; Ramscar et al., 2010). However, wherever there is similarity between different items, these similarities will be encoded in the learned, distributed representations. The "similarity," as we will show, can be similarity either in the explicit overlap between their predictions or in the systematic structural relationships among the various items within a domain. These internal representations therefore capture, in a graded and sub-symbolic fashion, both the similarities and the differences between the items. In the simulations that follow, we examine whether this framework can account for some of the empirical findings from the conceptual metaphor literature. In order to motivate the simulations, we begin with a discussion of a specific example of conceptual metaphor.

A CASE STUDY OF EMBODIED CONCEPTUAL METAPHOR: TIME AS SPACE

One the best documented cases of how abstract thinking can be metaphorically structured by concrete experience comes from the domain of time (Clark, 1973; Lakoff and Johnson, 1980; McGlone and Harding, 1998; Boroditsky, 2000, 2001; Boroditsky and Ramscar, 2002; Gentner et al., 2002; Evans, 2004; Matlock et al., 2005; Casasanto and Boroditsky, 2008). The language we use to talk about time is heavily infused with the language we use to talk about space, as when we talk about a *long* meeting or two birthdays being *close* together (Clark, 1973; Lakoff and Johnson, 1980). Consistent with the EC perspective, our actual perception of space can influence how we experience and reason about time (Casasanto and Boroditsky, 2008). For example, Casasanto and Boroditsky (2008) found that the length (in spatial extension) of a line on a computer screen affected how long (in temporal duration) it was judged to remain on the screen: the longer the line, the longer the time.

Like many other abstract, complex domains, there is more than one system of metaphor for talking and thinking about time (Clark, 1973; Lakoff and Johnson, 1980; Gentner et al., 2002). For instance, we can imagine ourselves moving forward through time, like when we talk about *coming up on* the holidays (ego-moving perspective), but we can also imagine remaining stationary as time moves toward us, like when we talk about the holidays *fast approaching* (time-moving perspective). Some spatial words that we use to talk about temporal events are ambiguous because they can be interpreted differently depending on which metaphorical perspective is adopted. For example, if you are told that Wednesday's meeting has been moved *forward* 2 days and you had adopted the ego-moving perspective, you would conclude that the meeting is now on Friday. However, if you had adopted the time-moving perspective you would conclude that the meeting is now on Monday (McGlone and Harding, 1998; Boroditsky, 2000; Boroditsky and Ramscar, 2002). Several experiments have demonstrated that the way people are currently thinking about space directly affects which of these perspectives they select and therefore how they reason about time (Boroditsky, 2000; Boroditsky and Ramscar, 2002). For example, people who are asked the Wednesday's meeting question at an airport are more likely to take the ego-moving perspective (i.e., give the Friday response) because they are about to take a flight (i.e., move through space) than when they are waiting to pick someone up (i.e., someone is approaching them in space; Boroditsky and Ramscar, 2002).

These findings suggest that we automatically use our online representations of space to structure our thinking about time. Why might this be the case and what mechanisms support this process? Researchers have highlighted at least two rich sources of information in our experience that could give rise to the metaphorical mapping between time and space. First, time and space co-occur in meaningful ways in our experience moving and acting in our environment (Lakoff and Johnson, 1980; Boroditsky, 2000). For instance, walking a longer distance typically takes a longer amount of time. Second, the structure of our linguistic experience, including the specific spatial metaphors we use as well as features of the language such as writing direction, might also influence how the concept of time is structured in terms of space (Boroditsky, 2000, 2001; Boroditsky and Gaby, 2010; Fuhrman and Boroditsky, 2010). For example, in both explicit event sequencing tasks and implicit temporal judgment tasks people represent time as progressing from the past to the future in a manner consistent with the writing direction of their language (Fuhrman and Boroditsky, 2010). In the following simulations we use the general modeling framework described above to explore how these metaphorical mappings may emerge gradually over the course of learning. Unlike previous proposals about the emergence of metaphor over developmental time, the mappings are not driven primarily by direct co-occurrence, but by the similarities in the structural regularities across domains.

MODEL SIMULATION 1

In the first simulation, the network learns about space and time through experience trying to predict how space and time are structured in the model environment (see **Table 1** for detailed model specifications). In the simplified environment of the model, space is laid out along a single dimension running from *West* to *East* (unlike our own environment, in which space is three-dimensional and also includes north and south, up and down! See **Figure 2**). To make the simulation easier to talk about and understand, space is divided into sections of different colors, going from *red* to *blue* to *green* to *yellow* as you move toward the *East*. Throughout the course of training the network will attempt to learn that two relations – *East of* and *West of* – structure the spatial arrangement of the colored sections in the environment. Training proceeds by asking the model to predict what color section of space it will "see" if it looks toward the *East* or *West* of its current position (and adjusting the weights in proportion to the error of this prediction, as described above). In practice, this works by presenting the network with one or more items along with a relation in the input layers and asking it to generate all appropriate outputs. For instance, if the network were presented with *blue* and *West of* it would have to output *green* and *yellow* (i.e., what the network would see if it looked toward the *East* while standing on the *blue* section: the sections of space that the *blue* section lies to the *West of*).

Time is also laid out in a single dimension from *earlier* to *later* events. Time is divided up into distinct moments, the days of the week, which follow a specific temporal sequence (going from

Table 1 | Detailed simulation parameters.

	Sim 1	Sim 2
LAYERS (# UNITS)		
Item	11	10
Relation	6	6
Learned item representation	7	7
Learned relation representation	4	4
Integration	9	9
Output	11	10
		OTHER PARAMETERS
Initial weight range (–/+)		–0.05/0.05
Activation function		Sigmoid
Error measure		SSE
Learning rate		0.1
Momentum		0

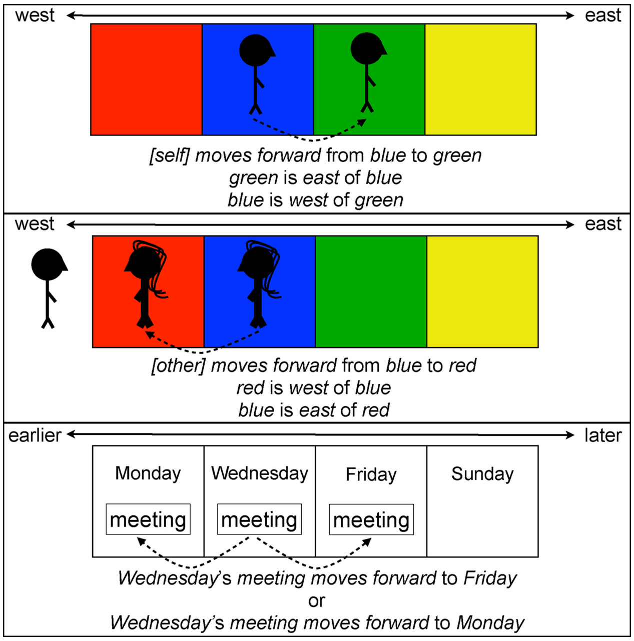
west
east
[self] moves forward from blue to green
green is east of blue
blue is west of green
west
east
[other] moves forward from blue to red
red is west of blue
blue is east of red
earlier
later
Monday
Wednesday
Friday
Sunday
meeting
meeting
meeting
Wednesday's meeting moves forward to Friday
or
Wednesday's meeting moves forward to Monday

FIGURE 2 | A diagram illustrating the structure of spatial and temporal relations in the model environment. The network learns about the consequences of both itself and other agents moving in the environment, though movement in the temporal domain is ambiguous.

Monday to *Wednesday* to *Friday* to *Sunday* as you progress *later* in time). During training the network will attempt to learn that two relations – *earlier than* and *later than* – structure the temporal sequence of the days of the week. The network learns about time in the same way that it learns about space. Thus, if the network were presented with *Monday* and *earlier than* as inputs it would have to generate *Wednesday*, *Friday*, and *Sunday* as outputs (i.e., the days of the week that Monday is *earlier than*).

Crucially, the network enjoys a richer, more structured set of experiences in the spatial domain because it can observe the consequences of its own movements in space (as well as the consequences of the movements of other agents in the environment). We can imagine that, like most mobile organisms, the network has both a front and back and can move forward and backward in space. To keep things simple, let us imagine that the network is standing on the *blue* section of space facing toward the *East*. If the network *moves forward*, it will move toward the *East* end up on the *green* section of space, while if it *moves backward* it will move toward the *West* and end up on the *red* section of space. However, *forward* and *backward* movements in space are not simply the same as moving toward the *East* or *West*. Now imagine that the network is observing another agent in the environment that is standing on the *blue* section of space and facing *West*. If this other agent *moves forward*, it will move toward the *West* end up on the *red* section of space, while if it *moves backward* it will move toward the *East* and end up on the *green* section of space. Thus the network has to learn that the consequences of moving *backward* and *forward* in space depends on whether it is attending to its own movements or to the movements of another agent. In practice, this works by including *self* and *other* items in the input layer to let the network know whose movements it is observing (see **Figure 2**). To keep things simple, we assume that the model is always facing toward the *East* and the *other* agent in the environment is always facing toward the *West*.

While the effects of movement in the spatial environment are unambiguous in the presence of either the *self* or *other* context, the model's experience of "movement" in the temporal domain is ambiguous in that there is no consistent mapping between *forward/backward* and *earlier/later*. The model learns that when a *Wednesday* meeting *moves forward*, it sometimes is moved to *Monday*, and other times it is moved to *Friday*. The same can occur when a meeting *moves backward*. Structuring the temporal domain in this way allows us to study to the ambiguity explored in Boroditsky (2000) and Boroditsky and Ramscar (2002). In particular, while the network has no experience with the *self/other* distinction in the temporal domain, we can examine whether it can use its experience with the effects of these contexts in the spatial domain to resolve the ambiguity of "movement" in the temporal domain. That is, we can test whether activating a particular spatial frame of reference (i.e., the *self* or *other* perspective) in the context of reasoning about a temporal event (i.e., moving the *Wednesday* meeting *forward*) will influence the network's expectations about the effects of "movement" in the temporal domain.

MODEL SIMULATION 2

The first simulation investigated whether the network would learn to metaphorically map its relatively rich experience with space onto the parallel but experientially impoverished domain of time in order to resolve an ambiguous temporal reasoning task. In Simulation 2, we explore whether the network can learn to map the directionality of time (from *earlier* to *later*) onto other spatial cues in the environment (e.g., the directionality of space, from *West* to *East*). Several studies have demonstrated that culturally specific spatial cues, such as writing direction (Fuhrman and Boroditsky, 2010) and absolute spatial coordinate systems (Boroditsky and Gaby, 2010), can influence and structure how people think about the directional "flow" of time.

The model was set up in a very similar manner as in Simulation 1. However, where Simulation 1 included an ambiguity in the temporal domain, Simulation 2 removes that ambiguity in order to closely align the meanings of the temporal and spatial relations. In particular, the *moves forward* relation was made unambiguous in both the spatial and temporal domains, by removing the *other* item. In the temporal domain, *moves forward* always predicted that the *Wednesday meeting* should occur on *Friday*, never *Monday,* and *moves backward* always predicted that *Wednesday* meetings should occur on *Monday*, never *Friday*. This might be interpreted as a culturally specific bias, analogous to the experience of reading temporally sequenced material like calendars and comics from left to right (or even writing direction itself, see Fuhrman and Boroditsky, 2010). In the spatial domain, we removed the patterns in which the *other* agent *moves forward* from *blue* to *red* and *moves backward* from *blue* to *green*, again rendering the situation unambiguous. Removing these four patterns, two from the temporal domain and two from the spatial domain, leaves the *moves forward* relation consistent with the *earlier than* relation in the temporal domain and with the *West of* relation in the spatial domain. This can be seen in the predicted outcomes of the events: *moves forward* from *Wednesday* predicts *Friday*, and *Wednesday* is *earlier than Friday*, and so on for the spatial domain.

If the model is sensitive to the structural similarities present in this environment, it should learn that the *West of* and *earlier than* relations make similar predictions in their respective domains, as do the *East of* and *later than* relations. As a result, the learned distributed representations for these pairs of relations should become similar as a function of experience – allowing, for instance, spatial words like *East of* and *West of* to be sensibly interpreted in the temporal domain (e.g., *Wednesday* is *East of Monday* or *Wednesday* is *West of Friday*). The model only ever observes *West of* and *East of* in the spatial domain, and *earlier than* and *later than* in the temporal domain, so an account based on direct co-occurrence would not generate the same prediction. This would provide a demonstration of how culturally specific features of the environment such as writing direction or dominant spatial coordinate systems could come to organize our representations of abstract domains such as time.

RESULTS

SIMULATION 1

In the first simulation we explored whether an ambiguity in the temporal domain (i.e., that a *Wednesday* meeting sometimes *moves forward* to *Monday* and sometimes *moves forward* to *Friday*) can be resolved by activating a particular spatial frame of reference. That is, even though the model has no experience with the *self/other* distinction in the temporal domain, we can nevertheless activate one of these spatial frames of reference in the temporal domain when asking the model whether it thinks *Wednesday* meetings *move forward* to *Monday* or *Friday*. If these reference frames influence the model's interpretation of *moves forward* in a way that is consistent with empirical results (Boroditsky, 2000; Boroditsky and Ramscar, 2002), we would expect that including *self* as an input (along with *Wednesday*, *meeting*, and *moves forward*) would yield relatively more activation in the *Friday* output unit than the *Monday* output unit. Alternatively, we would expect that including *other* as an input instead would result in relatively more activation in the *Monday* output unit than the *Friday* output unit.

To investigate this, we exposed the network to 10,000 epochs of training in the simplified environment, at which point we froze the weights to prevent further learning and began the testing phase. The statistics reported for both simulations include activation values that have been averaged across 10 instances of the model to ensure that any effects are not the result of a random bias in a particular instance. First, we tested whether the network had learned the unambiguous spatial and temporal structure of its environment by presenting it with the same input–output pairings that it was trained on. Indeed, the network performed quite well on this test (mean tss = 2.31, SD = 0.32)[1], demonstrating that it had correctly learned the features of its environment that it had been directly exposed to during training. Next we tested whether including a spatial frame of reference (i.e., *self* vs. *other*) influenced the network's predictions for the effect of *moves forward* in the temporal domain. We measured this effect by comparing three test patterns: (1) the ambiguous pattern that the network was trained on in which the *Wednesday* and *meeting* items were paired with the *moves forward* relation, (2) this same pattern with the *self* item included as an input, and (3) this same pattern with the *other* item included as an input (instead of *self*).

The model learned that *moves forward* was ambiguous in the temporal domain when the *self* and *other* items were not included as inputs. Specifically, when tested on the ambiguous pattern, the model fully predicted *meeting* on the output (mean = 0.981, SD = 0.0043) and partially predicted both *Monday* (mean = 0.500, SD = 0.0075) and *Friday* (mean = 0.503, SD = 0.0069). No other units had average activations greater than 0.02. A regression model that predicted output activation of the two target units (*Monday* and *Friday*) with contrast-coded predictors for Day (*Monday*, *Friday*) and Perspective (*self*, *other*) as well as a Day × Perspective interaction term, was fit to the two test patterns. Both main effects were significant (Day: $\beta = 0.062$, $p < 0.05$; Perspective: $\beta = -0.100$, $p < 0.01$) as was the interaction term ($\beta = 0.081$, $p < 0.01$), indicating that including the perspective units shifted the degree to which the model predicted that *Wednesday's meeting* would *move forward* to *Monday* or *Friday* (see **Figure 3**).

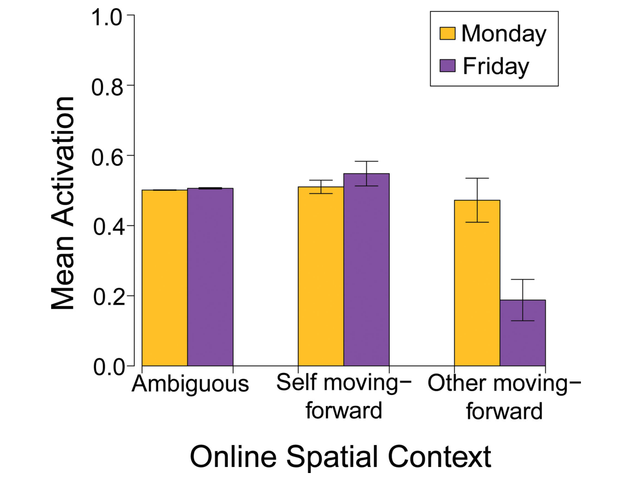

FIGURE 3 | The results of Simulation 1 showing that activating a particular spatial frame of reference biases the network's predictions about movement in the temporal domain.

SIMULATION 2

In the second simulation, we explored whether the network could in principle learn to map the directionality of time (from *earlier* to *later*) onto the directionality of space (from *West* to *East*). In order to clearly explore this possibility, we modified the model's environment slightly from that of Simulation 1 so that the *moves forward* relation was consistent with *later than* in the temporal domain and *East of* in the spatial domain (and so that *moves backward* was consistent with *earlier than* and *West of*). If the network is able to take advantage of this similarity in a way that is consistent with empirical findings (e.g., Fuhrman and Boroditsky, 2010; Boroditsky and Gaby,2010), then it should be able to interpret, for example, the relations *East of* and *West of* in the domain of time (i.e., *Wednesday* is *East of Monday* and *Wednesday* is *West of Friday*) even though these relations were never explicitly paired with temporal inputs or outputs in training.

We investigated this issue by exposing the network to 10,000 epochs of training and then freezing the weights. At this point the network had learned to make correct predictions for each of the training patterns (mean tss = 0.0898, SD = 0.0375). We then presented the network with two novel test patterns: one pairing *Wednesday* with *East of* on the input, the other pairing *Wednesday* with *West of* on the input.

When *Wednesday* was paired with the *East of* relation, the model predicted *Monday* (mean = 0.525, SD = 0.421) more than *Friday* (mean = 0.022, SD = 0.046), whereas when *Wednesday* was paired with the *West of* relation, the model predicted *Friday* (mean = 0.324, SD = 0.387) more than *Monday* (mean = 0.098, SD = 0.306). A regression model that predicted output activation with contrast-coded predictors for Day (*Monday*, *Friday*) and Relation (*East of*, *West of*) as well as an interaction term was fit to the two test patterns. Neither main effect was significant (Day: $\beta = 0.032$, $p = 0.54$; Relation: $\beta = -0.069$, $p = 0.19$); however, the Day × Perspective interaction was significant ($\beta = -0.182$, $p < 0.01$), indicating that these relation units held meaning in the domain of time even though this was the first time the model had encountered them in the temporal context (**Figure 4**).

In order to determine why this effect occurred, we submitted the representations for each of the relations of interest (i.e., *West of*, *East of*, *earlier*, *later*, *moves forward*, and *moves backward*) in the *Learned Representation Layer* to a hierarchical clustering analysis (shown in **Figure 5**). This analysis shows that the representation of *West of* is very similar to the representation of *earlier*, and the representation of *East of* is similar to the representation of *later*. It also shows that *moves forward* is more similar to *West of* and *earlier*, and *moves backward* is more similar to *East of* and *later*.

[1] The imposed ambiguity in the temporal domain (i.e., that *Wednesday's* meeting can move both *forward* and *backward* to *Monday* or *Friday*) made it impossible for the network's tss to improve beyond 2.

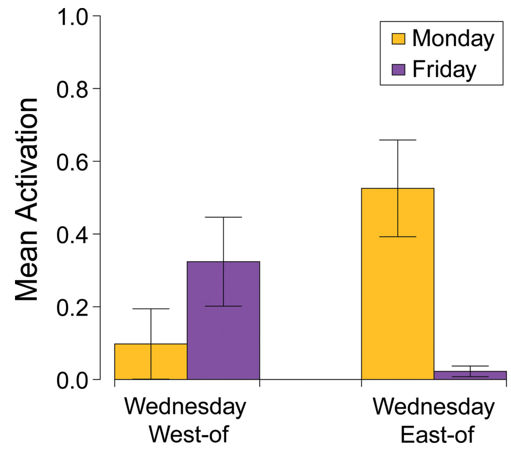

FIGURE 4 | The results of Simulation 2 showing that spatial relations that were never directly experienced in the temporal domain can still structure temporal reasoning.

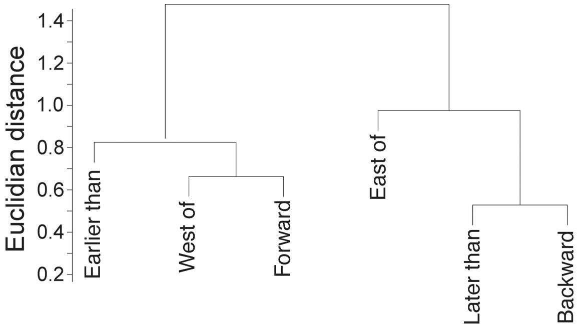

FIGURE 5 | The structure of the *learned relation representation* layer in Simulation 2.

SUMMARY OF FINDINGS

Both model simulations successfully learned a representation of the temporal structure of the world based partially on their experience with space. In Simulation 1, this allowed the network to resolve an ambiguity in the temporal domain by relying on additional structure only present in the spatial domain: a true application of conceptual metaphor to aid cognition. In Simulation 2, the network's representations of the spatial and temporal domains were shaped by a structural homology between the domains: in this case, a "culturally driven" bias to scan from *West* to *East* through time. In both cases, the network's metaphoric concepts were not driven by direct co-occurrence between concepts within the domains (e.g., distance with duration). Rather, the available information for learning the metaphor was the second-order relations between items within each domain (e.g., things move around in space in a similar way to how events can be sequenced in time).

DISCUSSION

WHY THE MODEL WORKS

Several theorists have proposed that the grounding of abstract thought in concrete knowledge may be due in part to the direct co-occurrence of certain domains in experience, for example, of time with space or of love with physical warmth (Lakoff and Johnson, 1999; see also the afterward to the 2003 edition of Lakoff and Johnson, 1980). The model presented here demonstrates another, more general, yet equally grounded pathway to metaphor: structural similarity between the target and the source domain (see also Boroditsky, 2000). However, this use of structural similarity is not a distinct online, rule-based algorithm operating over symbolic representations, as in other theories of structural alignment in metaphor comprehension and analogical reasoning (e.g., Falkenhainer et al., 1989); rather, it is a result of the gradual process of differentiation that takes place over the entire course of learning. It is fair to say that the network's knowledge of time is partially constituted by the learned structural relations in the spatial domain. This is demonstrated by the metaphoric remapping between time and space, which, in this model, share almost no input (only the *moves forward* and *moves backward* units) and no output units at all. Merely having distributed representations, as most connectionist models do, is not sufficient for this kind of behavior to emerge; the process by which those representations were acquired through experience with the environment is also critical.

To understand why the remapping occurs, recall that the model initially treats all items and relations as equivalent (due to its small and random initial weights) and only discriminates objects as it is forced to do so by the flow of information from the world. Over the course of this differentiation process, the model constructs several high-dimensional and highly overlapping representations for the items, the relations, and the item–relation conjunctions, all passing through the same sets of weights and patterns of activation over the same sets of units. If the network can reuse certain dimensions of its representations because of similarity in the structural relationships between and among items and contexts, it will tend to do so. Since the spatial and temporal domains share most, though not all, of their respective relational structure, the network learns a partially unified representation for the two domains. These overlapping representations, which are a direct result of the differentiation process, give rise to the influence of the concrete spatial experience on the abstract temporal reasoning task.

This model demonstrates that the structural homology between domains of experience is one aspect of the environment that can drive generalization (or, more properly, partial lack of differentiation) for metaphorical inference. But it is not the only way that metaphors can be learned. As mentioned above, co-occurrence of more abstract with more concrete domains of experience may also cause the learner to build metaphorical semantic representations. This is because the experience with the abstract domain will often predict properties that are also predicted during experience with the concrete domain, which may drive the representations to become more similar than they would otherwise be. In Simulation 2, it is indeed co-occurrence that drives learning, but it is indirect, not direct, co-occurrence. The *moves forward* unit in this simulation is unambiguously similar to *West of* when it occurs in the spatial domain, and to *earlier than* when it occurs in the temporal domain. Notice that *West of* and *earlier than* never predict similar outputs in a way that would cause them to become similar, so this is not a matter of raw co-occurrence. Still, the model is encouraged to draw its representations of *moves forward*, *West of*, and *earlier than* into a similar semantic structure because these relations must be re-represented in the *Learned Representation Layer*. In a parallel fashion, *moves backward* draws together the cross-domain relations *East of* and *later than*. The bridging between structures can occur because of similarity in the structural relationships among the items within each domain, or because of some direct or indirect co-occurrence (or co-prediction) in the environment, or (as we believe is probably the case in most natural settings) both.

Another mechanism that drives metaphorical structuring is language use. There are at least two routes through which language can bring about metaphoric alignment, one slow and one fast. In the current framework, linguistic experience is considered to be another aspect of the environment (this point is discussed in more detail below). The influence of language here would be across the (slow) course of development, serving as additional scaffolding for similar high-order structures (as in Simulation 1) or as an indirect co-occurrence cue or outcome (as in Simulation 2). On the other hand, language might be used online to point out a novel metaphorical structural mapping, such as "an atom is like the solar system." The agent's task is then to take two existing semantic structures and figure out what aspects of those structures the conversational partner intends to highlight. Our model deals with a very slow process of learning and differentiation, but does not have a way of rapidly integrating new information, so this kind of novel metaphorical language is a problem in this model. However, we are not claiming that our model describes the whole story, and any model of learning like ours will eventually need to take into account fast-learning processes as well (McClelland et al., 1995). Our model is nevertheless a novel contribution to the literature, as existing models of metaphor (and analogy) that do deal with online structural alignment (e.g., Falkenhainer et al., 1989) do not attempt to slowly integrate structural information over the course of development.

The possibility that speakers of different languages might categorize or even perceive the world in different ways has been a focus of scrutiny in recent work (e.g., Boroditsky et al., 2003; Majid et al., 2004; Winawer et al., 2007). One might expect that if embodiment holds, then the environment itself would fully determine the semantic representations possessed by the agents within that environment (and therefore language use would not really have any effect on conceptual representation; for a related position, see Gleitman and Papafragou, 2005). While this viewpoint recognizes the importance of the statistics of the environment in semantic learning, it fails to appreciate that linguistic information is itself another rich source of environmental statistics. The modeling approach described here provides a principled way of integrating the effects of language on cognition with EC (see also Dilkina et al., 2007; Boroditsky and Prinz, 2008; Andrews et al., 2009).

In our modeling framework, the key to this integration is to allow the network to experience a linguistic relational context alongside contexts conveying other kinds of perceptual and motor information. The network integrates information for each item across many different relational contexts, though these are limited to a fairly small set of physical and temporal relations. However, in other related models the contexts can take on a much broader meaning (e.g., Rogers and McClelland, 2004; Thibodeau et al., 2009). For example, in the semantic cognition model of Rogers and McClelland (2004, 2008), contexts include the Quillian-like *is-a*, *is, can*, and *has* relations. These relations may be thought of as different kinds of world contexts within which the items might be encountered. For example, when first observing a robin, one might notice that it is red; when attempting to catch a robin, one might observe that it can fly away; and when discussing robins with other people, one might be informed that a robin is a bird and an animal. If language is used to describe things in the world, we would expect that the relational context within which linguistic information is acquired should bear some structural resemblance to relational contexts grounded in perceptual-motor experience. In this case, as described above, the structural information may shape the representations even across different relations, leading to just the effects of language on thought that have been shown in experimental work.

IMPLICATIONS OF THE MODEL FOR EMBODIMENT

We have presented this model as an exploration of effects in the embodied conceptual metaphor literature, and as having implications for theories of EC as a whole. Thus it is important to address the possible criticism that this model neither is embodied nor speaks to issues in EC.

For one, because the inputs and outputs of this network are labeled units, of which only a few are active at a given time, the model may appear to support a more classically symbolic approach to cognition than EC would endorse. Indeed, these representations are highly simplified and abstracted from realistic sensorimotor information. We make the claim that this is an acceptable simplification because falling back to a relatively more localist representation does not fundamentally change the way the model works. Simulations by Rogers and McClelland (2004) using a similar model demonstrated that replacing the localist input units with a distributed input representation (e.g., of the visual features of animals, rather than their names) did not affect the model's performance in any significant way. There is reason to expect this result, since the model is not allowed to manipulate these localist inputs in any direct fashion; rather, as we noted earlier, it is forced to create its own distributed representations of these inputs in the layers that immediately follow. Our inputs may look like "linguistic" rather than "sensorimotor" representations precisely because they are localist, and many in the field think of linguistic units as localist symbols. While we do not exclude linguistic information as part of the experience relevant to the time/space effects (and neither do most researchers in the field), we do intend our model to stand for the entire space of experiences available to the agent.

Of course, this localist/distributed argument is somewhat distinct from the question of whether our training patterns accurately reflect the sensorimotor inputs to an agent, which in this case they do not and cannot. Even distributed representations would have to be greatly simplified and abstracted relative to the enormous flow of information that continuously impinges on the sensory receptors of any biological agent. Better input and output representations are surely possible. One promising approach would be to use unsupervised learning mechanisms such as the deep belief networks of Hinton and Salakhutdinov (2006) in order to extract distributed representations from more ecologically valid datasets. As this is not the focus of the current research, we used minimally distributed item representations, which nevertheless allowed us to capture the learning processes of interest.

For all that we believe that our simplifications are both justified and necessary, it might still be argued that the cumulative effect is to render the model "disembodied." However, to the extent that this model has consequences for EC, we would characterize it as a model of embodiment. We take our model as a kind of metaphor (as all models ultimately are), which points to a certain kind of statistical learning process that could help explain many of the results in the EC literature. We show that the statistical support available for learning environmental regularities is much stronger than the raw co-occurrence-based mechanisms previously proposed. Importantly, this helps explain how simple learning mechanisms (of the sort that may be plausibly instantiated by perceptual and motor brain regions) can give rise to "higher-level" cognitive processes such as conceptual metaphor. This is an example of the sort of back-and-forth engagement between connectionist and embodied approaches that we hope to foster in this paper. By situating our model in the EC perspective, we provide stronger support for the validity of the EC approach in general, and in particular, for the generality of the learning mechanisms that underlie many embodied theories.

FUTURE DIRECTIONS

The current model could be improved upon by including a more ecologically valid environment structure and set of training data. At present, we have made several simplifying assumptions that do not realistically map onto the ways in which humans experience space and time. For example, the network receives the same amount of experience moving forward and backward in space, only ever faces in one direction, and does not actually experience moving into every location in the surrounding space. While these simplifications allowed us to more easily explore the mechanisms underlying a small number of relevant findings, future versions of the model could incorporate a more realistic environment structure based on empirical findings in order to generate more precise and accurate predictions and explanations.

In the process of further developing this model, it will become increasingly important to explore the relationship between the model and the way in which conceptual metaphor is realized in the brain (see also Lakoff and Johnson, 1999; Feldman, 2006). One way to approach this question would be to use the model to make predictions about how neurological damage should affect metaphorical knowledge or the ability to reason metaphorically. We believe this approach could be fruitful given that Rogers and McClelland (2004) used a variant of the model that we adapted for our simulations to understand the degradation of semantic knowledge in patients with particular patterns of neurological damage. This work has led to more biologically plausible models of semantic dementia that highlight the specific role multimodal integration layers in the anterior temporal lobes play in semantic representation (Rogers et al., 2004).

Research on conceptual metaphor suggests that the effects of damage to particular brain regions will depend on the metaphorical domain in question. As described above, the reason that the conceptual metaphor approach fits naturally within the wider scope of EC theories is that abstract knowledge is thought to be grounded in lower-level sensory and motor mechanisms. This view suggests, for example, that to understand metaphors that rely on mappings to the motor domain (e.g., "The reader *grasped* the ideas in the paper"), we would draw on neural mechanisms that support actual motor planning or execution. Recently, neuroimaging evidence has been gathered in support of this claim (Boulenger et al., 2009). Other work has shown that processing metaphorical language about movement in space (e.g., "Stock prices *soared*") is sufficient to adapt direction-selective perceptual circuits and lead to a visual motion after-effect (Toskos Dils and Boroditsky, 2010).

In light of these findings, we would predict that brain areas responsible for representing spatial experience would also be important for certain aspects of temporal reasoning. In fact, recent research has implicated parietal cortex in representing space, time, number, and other domains that involve magnitudes (for a review, see Hubbard et al., 2005; Bueti and Walsh, 2009). Other researchers have suggested that the cerebellum, which is important for coordinating fine motor movements and balance in space, might also play a role in representing the temporal aspects of linguistic processing (Oliveri et al., 2009). The results of our current simulations suggest that any brain networks that represent the structure of space or time in experience might play a role in these metaphorical processes. Future research will explore the relationship between the model and the brain more directly.

CONCLUSION

In the introduction we outlined both a narrow and a broad goal for the modeling approach described in this paper. The narrow goal, capturing embodied effects in conceptual metaphor using a connectionist model, has been described in some detail above. We would now like to return to the broader goal of showing how connectionist models and theories of EC can mutually inform one another, and how marrying these approaches can benefit cognitive science as a whole.

For one, we have demonstrated that it is both possible and useful for proponents of EC to engage with the rich literature on domain-general learning mechanisms for insight into how to construct models of their findings and generate testable, mechanistic theories. This approach promises to provide an implementation of many ideas that EC theorists have proposed. Lakoff and Johnson (1980, 1999) saw that the conceptual system is deeply metaphorical, and we can now understand why this might be the case for a particular kind of learner embedded in a particular kind of environment. Our model provides an illustration of how conceptual metaphor naturally emerges within a system that learns the statistical structure of the environment through progressive differentiation and stores its representations as distributed and overlapping patterns of activation.

Connectionists, in turn, can gain a new understanding of their own models by examining the empirical findings from EC. The model of semantic cognition we extended here was originally tested on a simple Quillian semantic hierarchy (Collins and Quillian, 1969; Rogers and McClelland, 2004), and showed the right patterns of learning to account for traditional ideas of conceptual development. However, as the EC critique has ably pointed out, the physical abilities and limitations of the agent provide an extremely powerful source of statistics that pervades the agent's interactions with all features of its environment (Noë, 2004; Gibbs, 2006). This observation transforms the implications of the semantic cognition model, allowing us to think about it not as an observer gradually learning to construct a mirror image of the world inside its head, but as an active and embodied agent learning to predict the world around it. Similarly, embodiment effects, together with metaphorical overlap between learned contexts, provide new ways of thinking about traditional controversies. For example, language learning might be thought of as a task that occurs not in symbolic isolation, but within the broader context of learning to discriminate sounds in general, to produce sounds with the mouth in general, to predict the behavior of other agents in general, and so on. Thus the EC way of thinking seems to fit naturally into the domain-general approach to cognition that has been championed by connectionist researchers for decades.

Finally, we would like to suggest that this integrative approach to thinking about both EC and computational models might itself serve as a model for future research within cognitive science. We believe that we are now approaching a point in the evolution of cognitive science where various and diverse theoretical and experimental approaches can be usefully synthesized into a greater whole. Movement in this direction is ongoing (see, for example, the recent volume edited by Spencer et al., 2009) and we hope that this paper may serve as an additional nudge to the field.

ACKNOWLEDGMENTS

The authors would like to thank Lera Boroditsky and Jay McClelland for inspiration and helpful, critical discussion of the content of this paper and the issues it addresses. We would also like to thank Alexia Toskos Dils, Teenie Matlock, Dharshan Kumaran and Adriana Weisleder for additional help in thinking through the logic of our simulations and for comments on an earlier draft. This material is based on work supported under a National Science Foundation Graduate Research Fellowship and a Stanford Graduate Fellowship.

REFERENCES

Ackerman, J., Nocera, C., and Bargh, J. (2010). Incidental haptic sensations influence social judgments and decisions. *Science* 328, 1712.

Andrews, M., Vigliocco, G., and Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychol. Rev.* 116, 463–498.

Barsalou, L. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Bechtel, W. (1997). "Embodied connectionism," in *The Future of the Cognitive Revolution*, eds D. M. Johnson and C. E. Erneling (Oxford, UK: Oxford University Press), 187–208.

Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition* 75, 1–28.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cogn. Psychol.* 43, 1–22.

Boroditsky, L., and Gaby, A. (2010). Absolute spatial representations of time in an aboriginal Australian community. *Psychol. Sci.* doi:10.1177/0956797610386621

Boroditsky, L., and Prinz, J. (2008). "What thoughts are made of," in *Embodied Grounding: Social, Cognitive, Affective, and Neuroscientific Approaches*, eds G. Semin and E. Smith (New York, NY: Cambridge University Press), 98.

Boroditsky, L., and Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychol. Sci.* 13, 185–189.

Boroditsky, L., Schmidt, L. A., and Phillips, W. (2003). "Sex, syntax, and semantics," in *Language in Mind: Advances in the Study of Language and Thought*, eds D. Gentner and S. Goldin-Meadow (Cambridge, MA: MIT Press), 61–79.

Boulenger, V., Hauk, O., and Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cereb. Cortex* 19, 1905–1914.

Broadbent, D. (1987). "Simple models for experimentable situations," in *Modelling Cognition*, ed. P. Morris (Chichester: John Wiley & Sons), 169–185.

Bueti, D., and Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 364, 1831–1840.

Casasanto, D., and Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition* 106, 579–593.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Clark, A. (1998). *Being There: Putting Brain, Body, and World Together Again.* Cambridge, MA: MIT Press.

Clark, H. (1973). "Space, time, semantics, and the child," in *Cognitive Development and the Acquisition of Language*, ed. T. E. Moore (New York, NY: Academic Press), 27–63.

Collins, A. M., and Quillian, M. R. (1969). Retrieval time from semantic memory. *J. Verbal Learn. Verbal Behav.* 8, 240–247.

Dilkina, K., McClelland, J. L., and Boroditsky, L. (2007). "How language affects thought in a connectionist model," in *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, eds D. S. McNamara and J. G. Trafton (Austin, TX: Cognitive Science Society), 215–220.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychol. Rev.* 62, 145–154.

Evans, V. (2004). *The Structure of Time: Language, Meaning and Temporal Cognition*. Amsterdam: John Benjamins.

Falkenhainer, B., Forbus, K., and Gentner, D. (1989). The structure-mapping engine: algorithm and examples. *Artif. Intell.* 41, 1–63.

Feldman, J. (2006). *From Molecule to Metaphor: A Neural Theory of Language*. Cambridge, MA: MIT Press.

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

Fuhrman, O., and Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cogn. Sci.* 34, 1430–1451.

Gentner, D., Imai, M., and Boroditsky, L. (2002). As time goes by: evidence for two systems in processing space → time metaphors. *Lang. Cogn. Process.* 17, 537–565.

Gibbs, R. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge, UK: Cambridge University Press.

Gibbs, R. (1996). Why many concepts are metaphorical. *Cognition* 61, 309–319.

Gibbs, R. (2006). *Embodiment and Cognitive Science.* Cambridge, UK: Cambridge University Press.

Gibson, J., and Gibson, E. (1955). Perceptual learning: differentiation or enrichment? *Psychol. Rev.* 62, 32–41.

Gleitman, L., and Papafragou, A. (2005). "Language and thought," in *Cambridge Handbook of Thinking and Reasoning*, eds K. Holyoak and B. Morrison (Cambridge, UK: Cambridge University Press), 633–661.

Glenberg, A. (1997). What memory is for: creating meaning in the service of action. *Behav. Brain Sci.* 20, 41–50.

Grice, H. (1975). "Logic and conversation," in *Readings in Language and Mind*, eds H. Geirsson and M. Losonsky (Cambridge, MA: Blackwell Publishers), 41–58.

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507.

Hintzman, D. (1991). "Why are formal models useful in psychology," in *Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock*, eds W. E. Hockley and S. Lewandowsky (Hillsdale, NJ: Lawrence Erlbaum), 39–56.

Hubbard, E. M., Piazza, M., Pinel, P., and Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nat. Rev. Neurosci.* 6, 435–448.

Jostmann, N. B., Lakens, D., and Schubert, T. W. (2009). Weight as an embodiment of importance. *Psychol. Sci.* 20, 1169–1174.

Kemp, C., and Tenenbaum, J. B. (2008). The discovery of structural form. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10687–10692.

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By.* Chicago, IL: University of Chicago Press.

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought.* New York, NY: Basic Books.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., and Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends Cogn. Sci.* 8, 108–114.

Marr, D. (1982). *Vision.* New York, NY: W.H. Freeman.

Matlock, T., Ramscar, M., and Boroditsky, L. (2005). On the experiential link between spatial and temporal language. *Cogn. Sci.* 29, 655–664.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Top. Cogn. Sci.* 1, 11–38.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–437.

McGlone, M., and Harding, J. (1998). Back (or forward?) to the future: the role of perspective in temporal language comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 1211–1223.

Murphy, G. (1997). Reasons to doubt the present evidence for metaphoric representation. *Cognition* 62, 99–108.

Murphy, G. L. (1996). On metaphoric representation. *Cognition* 60, 173–204.

Niedenthal, P., Barsalou, L., Winkielman, P., Krauth-Gruber, S., and Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Pers. Soc. Psychol. Rev.* 9, 184.

Noë, A. (2004). *Action in Perception.* Cambridge, MA: MIT Press.

Oliveri, M., Bonnì, S., Turriziani, P., Koch, G., Lo Gerfo, E., Torriero, S., Vicario, C. M., Petrosini, L., and Caltagirone, C. (2009). Motor and linguistic linking of space and time in the cerebellum. *PLoS ONE* 4, e7933. doi: 10.1371/journal.pone.0007933.

Pinker, S. (2007). *The Stuff of Thought: Language as a Window into Human Nature.* New York, NY: Viking Press.

Proffitt, D. (2006). Embodied perception and the economy of action. *Perspect. Psychol. Sci.* 1, 110.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., and Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cogn. Sci.* 34, 909–957.

Rescorla, R. A. (1988). Pavlovian conditioning: it's not what you think it is. *Am. Psychol.* 43, 151–160.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.

Rogers, T., and McClelland, J. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach.* Cambridge, MA: MIT Press.

Rogers, T. T., Lambon Ralph, M. A., Garrad, P., Bozeat, S., McClelland, J. L., Hodges, J. R., and Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychol. Rev.* 111, 205–235.

Rogers, T. T., and McClelland, J. L. (2008). Précis of *Semantic Cognition: A Parallel Distributed Processing Approach. Behav. Brain Sci.* 31, 689–749.

Searle, J. (1979). "Metaphor," in *Metaphor and Thought*, ed. A. Ortony (Cambridge, England: Cambridge University Press), 83–111.

Seidenberg, M. (1993). Connectionist models and cognitive theory. *Psychol. Sci.* 4, 228–235.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behav. Brain Sci.* 11, 1–74.

Spencer, J., Thomas, M., and McClelland, J. (2009). *Toward a Unified Theory of Development.* Oxford, UK: Oxford University Press.

Spivey, J. (2007). *The Continuity of Mind.* Oxford, UK: Oxford University Press.

Thibodeau, P. H., McClelland, J. L., and Boroditsky, L. (2009). "When a bad metaphor may not be a victimless crime: the role of metaphor in social policy," *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, eds N. Taatgen and H. van Rijn (Amsterdam: Cognitive Science Society), 809–814.

Toskos Dils, A., and Boroditsky, L. (2010). Visual motion aftereffect from understanding motion language. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16396–16400.

Varela, F., Thompson, E., and Rosch, E. (1991). *The Embodied Mind.* Cambridge, MA: MIT Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7780–7785.

Ziemke, T. (2003). "What's that thing called embodiment," in *Proceedings of the Annual Conference of the Cognitive Science Society*, eds R. Alterman and D. Kirsch (Mahwah, NJ: Cognitive Science Society), 1305–1310.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 July 2010; accepted: 22 October 2010; published online: 22 November 2010.

Citation: Flusberg SJ, Thibodeau PH, Sternberg DA and Glick JJ (2010) A connectionist approach to embodied conceptual metaphor. Front. Psychology ***1****:197. doi: 10.3389/fpsyg.2010.00197*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.