Title

Integrating Spatial Language and Spatial Memory: A Dynamical Systems Approach

Permalink

https://escholarship.org/uc/item/6r24w1p9

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN

1069-7977

Authors

Lipinski, John
Spencer, John P.
Samuelson, Larissa K.

Publication Date

2004

Peer reviewed

Integrating Spatial Language and Spatial Memory:
A Dynamical Systems Approach

John Lipinski (john-lipinski@uiowa.edu)
John P. Spencer (john-spencer@uiowa.edu)
Larissa K. Samuelson (larissa-samuelson@uiowa.edu)
Department of Psychology, University of Iowa
Iowa City, IA 52242 USA

Abstract

The domain of spatial language is an ideal testing ground for proposals addressing the representational gap between perceptual-motor and language systems precisely because it is an unambiguous case of these systems coming together. To date, however, efforts addressing this representational gap within the domain of spatial language have generated conflicting results. Focusing on an “above” ratings task, we provide here a dynamical systems approach to spatial language performance and supporting empirical results that address this impasse. The development of a dynamical systems model linking spatial language and spatial memory is also discussed.

Representation and Spatial Language

A current focus in cognitive science is understanding how the sensory-motor and linguistic systems interact. Because spatial language brings words and physical space together so directly, it is the ideal vehicle for exploring this interaction. To date, two general approaches to representation speak to this issue of interaction in spatial language (Barsalou, 1999): amodal symbolic systems and perceptual symbol systems.

Amodal symbolic systems presume representational independence between symbolic processes like language and sensory-motor systems (Harnad, 1990; Anderson, 2000). The amodal view thus requires a transduction process that permits “communication” between linguistic and non-linguistic systems. This transduction process is best described by Jackendoff’s representational interface (1992; 1996; 2002) in which communication between different types of representations (e.g. auditory and visual) is achieved through a process of schematization– the simplifying and filtering out of information within one representational format for use in another representational system (Talmy, 1983). The representational interface approach ultimately permits abstract conceptual structures that can encode spatial representations but still capture the core characteristics of the symbolic view (e.g. pointers to sensory modalities, type-token distinctions, taxonomies).

There is significant empirical support for this perspective. Talmy (1983), for example, showed that language uses closed-class prepositions (such as “above”, “below”, or “near”) to provide an abstracted, skeletal structure of a scene that narrows the listener’s attention to a particular relationship between two objects by disregarding other available information (Talmy, 1983; Hayward & Tarr, 1995). Thus, in the sentence “The bike stood near the house”, all of the specific information about the bike (e.g. size, shape, orientation) is disregarded and the bike is instead treated as a dimensionless point (Hayward & Tarr, 1995). As a result of this schematization, linguistic representations of relational states can be extended to a variety of visual scenes and objects with little regard to the individual object characteristics.

In contrast to transduction and the amodal approach, Barsalou’s Perceptual Symbol Systems (1999) posits perceptual symbols: “records of neural states that underlie perception” (p.583) that are both inherently grounded in the given sensory modality and capable of replicating the flexible, productive, and hierarchical capacities of amodal symbolic systems. These perceptual symbols are implemented when top-down processes partially reactivate sensory-motor areas and organize the perceptual components around a common frame. Ultimately, perceptual components implement a simulator that captures both perceptual memories and core symbolic behaviors (e.g. type-token distinctions, hierarchies). Because these symbols are grounded in sensory-motor processes, they do not require pointers or transduction to become “meaningful”.

A growing empirical literature supports Barsalou’s (1999) PSS as well. For example, Stanfield and Zwaan (2001) argued that if symbolic, linguistic representations are integrated with perceptual symbol systems, people should be faster to recognize visual objects described in a sentence as the similarity between the perceived object and the description increase. Consistent with this prediction, they found that people were faster to recognize an object (e.g. a vertically oriented pencil) as part of a previous sentence when that sentence matched the orientation (e.g. He placed the pencil in the cup) than when it conflicted (e.g. He placed the pencil in the drawer). Visual information has also been shown to facilitate real-time resolution of temporarily syntactically ambiguous sentences (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), further evidence against a hard separation between linguistic and sensory systems. Finally, recent work by Richardson et al. (2003) shows that verbal stimuli interact with visual discrimination performance, additional evidence that linguistic processing can directly impact the processing of visual space.

In summary, the contrasting amodal and modal perspectives both appear to be substantially supported.

iven the clear contrast between the two theories, however, both cannot be correct. Thus, despite a vigorous debate and valuable empirical data on both sides, the fundamental question of how linguistic and non-linguistic systems relate remains unanswered.

Limits of the Current Approaches

Further consideration suggests two critical limits of the proposals and empirical support discussed above. First, they rely on descriptive, conceptual accounts of representational structure. Though critical at initial stages of theory development, the flexibility of conceptual accounts makes them ultimately difficult to critically test and falsify. Consequently, data collected in support of one view can be reinterpreted by the other view. Jackendoff (2002), for example, incorporated the resolution of syntactic ambiguity through visual processing (Tanenhaus, et al., 1995) using characteristics of a syntax-semantics interface.

The second, related limit of the current literature is treatment of representational structure in the abstract. In particular, with the exception of recent tests of the PSS theory (e.g. Richardson et al., 2003), spatial language studies have tended to focus on the nature of representational structure without considering the second-to-second processes that give rise to those structures. This can lead to an impasse because representations are not strongly grounded in task-specific performance. Consideration of an ongoing debate within spatial language illustrates this point. Because this debate is central to our empirical work, it is considered in some detail.

Evidence for Shared Representations

In order to explore the possible correspondence between the linguistic and sensory-motor representations of space, Hayward and Tarr (1995) conducted a series of experiments designed to compare how object relations are linguistically and visually encoded. In the first experiment, participants were presented with a visual scene depicting a referent object and a target object and asked to generate a preposition describing the relationship. Results suggested that the prototypical spatial positions for “above” and “below” lie along a vertical reference axis and prototypical spatial positions for “left” and “right” lie along a horizontal axis. In addition, use of these terms declines as target positions deviate from the horizontal and vertical reference axes.

Next, Hayward and Tarr built on these findings by using a preposition ratings task. In the ratings task, participants were asked to rate on a scale of 1 (least applicable) to 7 (most applicable) the applicability of a given spatial term (e.g. above) to a relationship between two objects. This ratings task is particularly valuable because it permits quantification and metric manipulation of an otherwise gross measure of linguistic output (e.g. above/not above). As such, this task provides a means of empirically bridging the continuity of sensory-motor representations with the discreteness of linguistic representations. Results from this ratings task showed strong metric effects of spatial language use around the vertical and horizontal axes. For instance, “above” ratings were highest along the vertical axis and systematically decreased as the target object’s position deviated from the vertical axis. Hayward and Tarr concluded that this ratings gradient across spatial positions reflected the use of prototypical vertical and horizontal reference axes.

To compare the representational prototypes of spatial language with visual representations of space, Hayward and Tarr compared these findings with performance on location memory and same-different discrimination tasks. Importantly, the areas of highest spatial recall accuracy were vertically aligned with the reference axes used as prototypes in the ratings task. Performance in the same-different location task yielded similar findings, showing that discrimination was best along the vertical and horizontal axes. Collectively, data from these four experiments point to a shared representational spatial structure between linguistic and sensory-motor systems, a result consistent with Barsalou’s PSS approach.

Evidence Against Shared Representations

Follow-up results from Crawford, Regier, & Huttenlocher (2000) present a different picture. To probe both linguistic and visual representations of space, they analyzed “above” ratings as well as spatial memory performance. Although results showed an “above” ratings gradient aligned with the vertical axis similar to that of Hayward and Tarr (1995), Crawford et. al. also found location memory bias away from the vertical axis when participants had to recall the locations of targets to the left and right of this axis. These researchers proposed that the cardinal axes that appear to function as prototypes in the linguistic task instead serve as category boundaries in the spatial memory task. Thus, while both linguistic and sensory-motor spatial representations use the same cardinal axes, these axes serve functionally distinct representational roles in the two tasks. It therefore appears that the linguistic and sensory-motor representations of space differ in critical ways, a conclusion consistent with an amodal representational interface perspective.

Considered together, these results illustrate the limits of dealing with representation in the abstract: both sets of researchers used similar experimental tasks and reported largely similar findings, yet they draw starkly different conclusions, conclusions that depend critically on abstract definitions of representational structure. Because we do not yet know the process that selects, creates, and encodes spatial prototypes nor the process used to create a spatial rating, we cannot go beyond abstract representational descriptions to make predictions about the similarities and differences across tasks. Notably, the failure to resolve this particular debate within spatial language mirrors the larger failure to resolve the modal-amodal conflict. In most general terms, the empirical support offered in both cases fails to delineate between the proposed accounts.

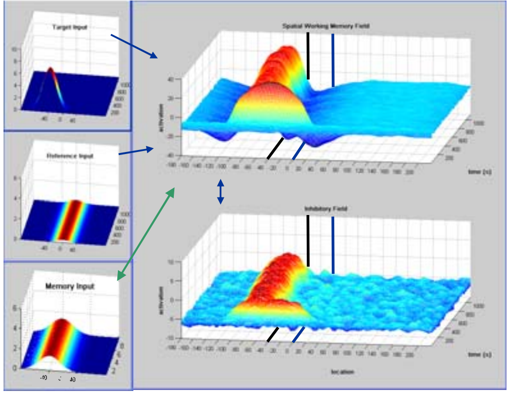
A Process Approach to Spatial Language

The theory and data discussed so far appear to be at an impasse, due in part to an emphasis on descriptive accounts of representational systems and a focus on representation in the abstract. To move beyond these fundamental limitations, the current proposal seeks to establish and test a process model that relates spatial memory and verbal performance. Such a process model can move beyond description and representation in the abstract and provide strong, testable predictions.

To lay the foundation of this proposed model, consider again the results of Crawford et al. (2000). The distinguishing result was the finding of spatial memory biases away from the vertical axis. They interpreted this movement away from midline to be a function of bias towards spatial categories. This interpretation is derived from Huttenlocher et al.'s (1991) Category Adjustment (CA) model. According to the CA model, people encode spatial location at two levels. The first level encodes <u>fine-grained</u> information about target location (e.g. angular deviation), while the second level encodes the region or category of target location. Specifically, the CA model proposes that people represent a central or prototypical value within a category that is most representative of that category. To remember a location, these two levels of detail are then combined to produce a remembered target location. Importantly, fine-grained and categorical information can be weighed differently. If, for example, the fine-grained detail is less certain, the prototype can be given more weight, resulting in a bias away from midline. Moreover, evidence from Huttenlocher et al. (1991) indicates that these spatial prototypes lie along the diagonal axes. According to Crawford et al. (2000), these spatial prototypes along the diagonals are the source of the observed spatial memory biases away from midline. Recall, however, that spatial prepositions maintained their highest applicability ratings along the vertical and horizontal axes. Thus, spatial prepositions appear to maintain prototypes along vertical and horizontal axes while spatial categories appear to maintain prototypes along the diagonal axes.

But must the drift away from midline observed in spatial memory performance result from spatial prototypes along the diagonal? A recent model suggests no. Specifically, the Dynamic Field Theory (DFT) (Spencer & Schöner, 2003; Schutte, Spencer, & Schöner, 2003) provides a formalized process account of spatial memory bias away from reference axes <u>without positing prototypes</u>. This model specifies how location-related activation is maintained in spatial working memory (SWM) during short-term delays and how perception and memory are integrated within this single representational system.

The DFT can be best understood within the context of a location memory task used to test predictions of the model. In this task, participants are seated at a large empty table and a spaceship-shaped object is flashed for 2 seconds on the table. After a variable delay, participants are asked to indicate the location of the ship using a computer mouse. Participants in this task show the same biases away from midline reported by Huttenlocher and colleagues. The focus of the DFT is to explain this performance through activation in the SWM field.



<u>Figure 1</u> Dynamic Field Theory of spatial working memory

Figure 1 shows the structure of the DFT model. The large box shows the excitatory and inhibitory layers of neurons that together form the SWM field. Each layer has a collection of spatially tuned neurons that respond selectively to a specific location. Spatial location is indicated by position along the x-axis, where 0˚ is the center of the space; positive locations are rightward, and negative locations are leftward. The y-axis represents time which is moving away as a particular experimental trial proceeds from start to finish. The z-axis captures the activation of each neuron in the field.

In addition to the excitatory and inhibitory layers of the SWM field, there are input fields: target input, reference input, and memory input. The upper left portion of Figure 1 shows the target input which feeds into the excitatory layer of the SWM field. This target input turns on when the target is visible and turns off when the target is hidden. Figure 1 also shows the reference input. This reference input captures perception of the midline or vertical symmetry axis, the same axis central to the linguistic and non-linguistic representations of space discussed above. The third input is the long-term memory field which reflects the activation history of the SWM. This field also reciprocally feeds into the SWM field to impact real-time spatial memory processes.

The integration of these inputs in working memory is governed by an interaction function that determines how activation at one site in the SWM field influences activation at other sites. The DFT uses a local excitation and lateral inhibition function. Thus, activation at one site increases the activation of its neighbors and decreases the activation of sites further away. There are two main consequences of the interaction function. First, strong target input can lead to a self-sustaining peak of activation. These self-sustaining peaks of activation maintain themselves even after the target input is removed. In this way the field can maintain a memory of the target location during short-term delays.

The second consequence of the interaction function is that self-sustaining peaks can drift away from reference axes such as midline during memory delays. The process that gives rise to such delay-dependent spatial drift is illustrated in Figure 2. The short activation profile in this figure was generated by running a simulation of the model shown in

Figure 1 with only a single input—the reference input—and taking a time slice through the excitatory layer (top layer in the large panel of Figure 1) of the resultant SWM field (at time 8.00 s). Thus, this short activation profile reflects the influence the reference input has on each neuron in SWM at a particular moment in time (note that, because the reference input in Figure 1 is constant throughout the memory delay, the short activation profile actually captures the resultant influence of the reference input throughout the trial).

As can be seen in Figure 2, the resultant reference profile has stronger activation around midline; however, there are also two troughs in activation to the left and right of midline. These troughs cause systematic delay-dependent drift away from midline when targets are positioned to the left and right of this axis. This is captured schematically by the tall activation profile in Figure 2. As can be seen in Figure 2, this tall activation profile receives slightly more reference-related input on its left side than its right side. As a consequence, neurons on the left side of the activation peak will tend to join into the locally-excitatory interaction, while neurons on the right side of the peak will tend to be laterally inhibited.

The excitatory (top) layer of the SWM field in Figure 1 shows that as this interactive process propagates through time, activation peaks can spatially drift. In particular, Figure 1 shows a simulation of the model during a single trial to a -40º location. At the start of the trial, activation in the excitatory layer of SWM is relatively uniform because no strong inputs are present. At 2 s, the target appears at -40º and the strong target input associated with this event builds a peak of activation in SWM. Importantly, this activation peak sustains itself even after the target disappears at 4 s. And, during the ensuing memory delay, the peak drifts systematically away from midline (i.e., away from 0º). Note that this effect is partially counteracted by the long-term memory input at -40º.

In summary, the DFT provides a process-based alternative to Huttenlocher et al.'s (1991) category adjustment model. Critically, this model links spatial memory biases to a process that integrates remembered information in working memory with perceived reference frames, the same reference frames implicated in research by Hayward and Tarr (1995). As a result, the central argument against Hayward and Tarr's claim of shared structure between linguistic and non-linguistic representations of space—that memory is biased away from a category boundary—no longer follows obligatorily from the data. This provides the impetus to once again consider the potentially rich and direct connections between spatial memory and spatial language.

Connecting the DFT with Spatial Language

Inspired by our model of spatial working memory, we recently conducted a set of experiments designed to investigate Hayward and Tarr's (1995) claim that linguistic and non-linguistic representations overlap, not by examining representational structures in the abstract, but by considering the specific representational structures that emerge in our formalized process model. In particular, we asked whether the processes that create delay-dependent spatial drift in spatial working memory might also leave some empirical signature in a spatial language task. Toward this end, we used the ratings task from Hayward and Tarr given its capacity to reveal quantifiable metric effects and its centrality in the spatial language literature (e.g., Hayward & Tarr, 1995; Crawford et al., 2000; Logan & Sadler, 1996; Regier & Carlson, 2001).

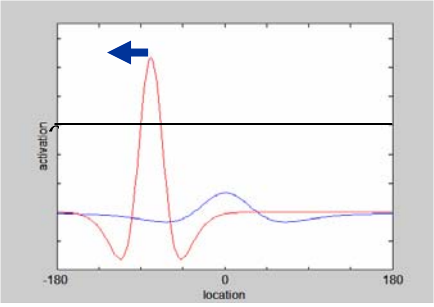


Figure 2 Local excitation/ lateral inhibition

To relate the DFT to performance in the ratings task, we borrowed an idea from Regier and Carlson's (2001) Attentional Vector Sum model and scaled verbal ratings for "above" by the angle between the representation of the target location and the representation of the reference axis, that is, by the spatial distance between the center of the activation peak in SWM and the midline axis (0º). Ratings should be highest when activation is centered at 0º and should fall off systematically as the activation peak is shifted to the left or right. Based on this proposal and the dynamic properties of the DFT, we predicted that if spatial language and spatial memory use the same representational system—spatial working memory—then ratings performance should show delay-dependent "drift", giving systematically lower "above" ratings as memory delays increase (i.e. as the distance between the activation peak in SWM and the midline axis increases).

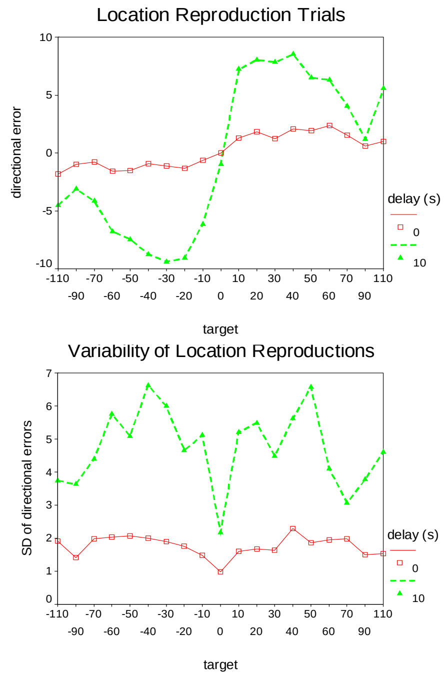


Figure 3. Spatial memory performance

Experimental Support

Subjects. 15 University of Iowa undergraduates participated in this study in exchange for class credit or payment.

Method. Experimental sessions were conducted in dim lighting in a room with black curtains covering all external landmarks. A curved border occluded the corners of the table (and therefore the diagonal symmetry axes).

A single referent disc appeared at midline 30cm in front of the participant and remained visible throughout each presentation trial. At the start of each trial the participant moved a computer mouse to this disk. A number (100-500) then appeared and participants begin counting backwards by 1s aloud until they made a response. This counting task prevented the verbal encoding and maintenance of the spaceship position or rating. A small, spaceship-shaped target then appeared on the screen for two seconds.

Trial Types For spatial memory trials, participants were instructed to move the mouse to the location corresponding the ship's location when the computer says "Ready-Set-Go". For spatial language rating trials, on the other hand, participants are instructed to rate on a scale of 1 ("definitely not above") to 9 ("definitely above") the extent to which the word "above" describes the spaceship's location relative to the reference disk and say their rating when the computer says "Please give your 'Above' rating." The spoken stimuli that indicated which response to provide were each 1500ms in duration. In No Delay conditions, completion of the spoken stimulus was timed to coincide with the offset of the spaceship target. In the 10s Delay conditions, completion of the spoken stimulus occurred exactly 10 seconds after the disappearance of the spaceship target. Spaceship targets appeared at a constant radius of 15cm at 19 different locations relative to the midline axis (0º): every 10º from -70º to +70º as well as ±90º and ±110º to map onto previous research.

Results

Participants in our modified spaceship task either gave a spatial memory response or a verbal ratings response (1 = "definitely not above", 9 = "definitely above") after a 0 s or 10 s delay. The top portion of Figure 3 shows directional errors on the memory trials across target locations and delays. Positive errors were clockwise, while negative errors were counterclockwise. Consistent with previous work (Spencer & Hund, 2002), directional error was larger in the 10 s delay condition and responses were systematically biased away from midline (responses to negative or leftward targets showed counterclockwise bias; responses to positive or rightward targets showed clockwise bias). Similar effects were found for variable error (see lower portion of Figure 3). Specifically, variability was higher in the 10s delay condition, and responses to targets to the left and right of midline were more variable than responses to the target aligned with 0º.

Critically, we also found the predicted delay-dependent drift effect in participants' ratings performance. The top portion of Figure 4 shows that "above" ratings in the spaceship task followed a gradient similar to that obtained by Hayward and Tarr (1995) and Crawford et al. (2000). However, there was a systematic and significant decrease in ratings in the 10 s delay condition. Examination of ratings variability revealed effects of delay comparable to those found on the spatial memory trials (see Figure 4). Specifically, ratings variability was higher at the long delay and lower for targets near the midline axis.

In a final analysis, we compared spatial memory and ratings responses directly by converting the ratings "drift" apparent in Figure 4 into a spatial deviation measure (e.g., deviation at 10º target = (change in 10 s delay rating between 10º and 20º) * 10º / (change in 0 s delay rating between 10º and 20º)). This analysis revealed a high degree of overlap in delay-dependent spatial drift across the two tasks (see Figure 5). These results support the prediction we generated from the DFT and suggest that a shared working memory representation underlies performance in both tasks.

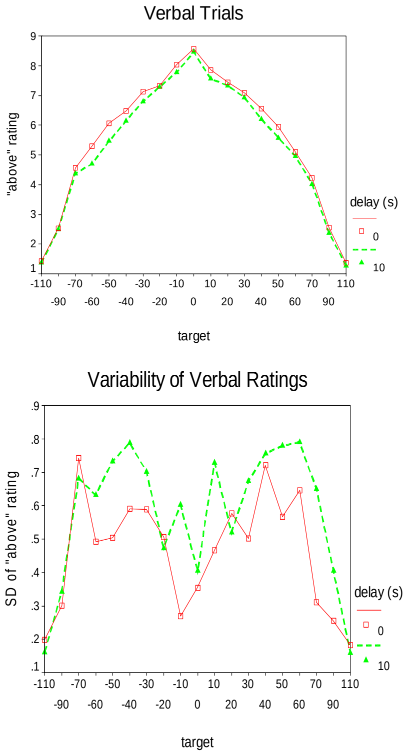


Figure 4 Ratings performance

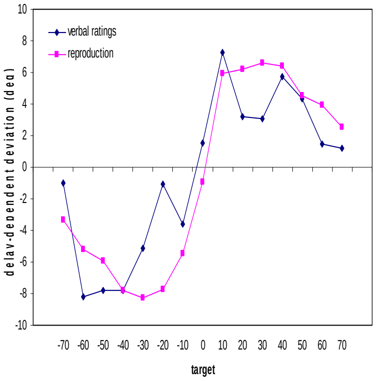


Figure 5 Ratings and drift

Towards a Dynamic Field Model of Spatial Language Performance

Given that we have a formal theory of SWM and encouraging preliminary data, we are in a unique position to develop a process model of both spatial memory and verbal behavior in spatial tasks. The starting point of such a model will be a modified dynamic field model that links two dynamic fields—the SWM field discussed previously and a new spatial prepositions field. Although this new spatial prepositions field has yet to be formalized, the current data suggest two important features. First, this field must be alignable with particular locations in SWM, in particular with perceived reference frames. We are currently developing a process within the field theory that handles the alignment of multiple fields, including the anchoring and scaling necessary in such situations. Critically, these processes must be developed in a way that allows for generalizability to spatial prepositions beyond "above" such as "left", "right", and "below".

Second, consistent with the dynamic nature of the tasks employed here, and cognition more generally, the two fields should be dynamically coupled. This means that activation in SWM can serve as input to the spatial preposition field and vice versa. This dynamic coupling is critical given the presented evidence that verbal ratings reflect the same dynamic processes underlying spatial working memory performance. If these layers are indeed dynamically coupled as we suggest, then establishment of stable activation peaks within one layer should give rise to stable peaks in the other. Similarly, instability and drift within one layer should give rise to a instability and drift within the other layer. Experiments are currently underway to test these specific predictions.

Although this provides only a limited window onto the dynamic processes that underlie a very flexible spatial cognitive system, we contend that this is an appropriate starting point given the novelty of our general theoretical approach. Indeed, the results of our current experiments will provide the empirical foundation for a more extensive formal model that links spatial working memory with spatial language processes.

References

Anderson, J.R. (2000). *Cognitive psychology and it implications*. New York: Worth.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*,577-660.

Crawford, L.E., Regier, T., & Huttenlocher, J. (2000). Linguistic and non-linguistic spatial categorization.*Cognition, 75*, 209-235.

Harnad, S. (1990). The symbol grounding problem.*Physica D, 42*, 335-346.

Hayward, W.G., & Tarr, M.J. (1995). Spatial language and spatial representation. *Cognition, 55*, 39-84.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98*, 352-376.

Jackendoff, R. (1992). *Languages of the mind*. Cambridge, MA: MIT Press.

Jackendoff, R. (1996).The architecture of the linguistic-spatial interface. In P. Bloom, M.A. Peterson, L. Nadel, & M.F. Garret (Eds) *Language and Space*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). Foundations of language. New York: Oxford University Press.

Logan, G.D.,& Sadler, D.D. (1996). The architecture of the linguistic-spatial interface. In P. Bloom, M.A. Peterson, L. Nadel, & M.F. Garret (Eds) Language and Space. Cambridge, MA: MIT Press.

Regier, T., & Carlson, L.A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General, 130*,2373-298

Richardson, D.C., Spivey, M.J., Barsalou, L.W., & McRae, K. (in press). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science.*

Schutte, A. R., Spencer, J. P., & Schöner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. *Child Development, 74*, 1393-1417.

Spencer, J.P. & Hund, A.M. (2002). Prototypes and particulars: Geometric and experience-dependent spatial categories. *Journal of Experimental Psychology: General, 131,* 16-37.

Spencer, J.P. & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science, 6,* 392-412.

Stanfield, R.A., & Zwaan, R.A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*, 153-156.

Talmy, L. (1983). How language structures space. In H. Pick& L. Acredelo (Eds): Spatial orientation: Theory, research, and application. Plenum. New York.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. Science, 268, 1632-1634.