
It's in the Eye of the Beholder: Spatial Language and Spatial Memory Use the Same Perceptual Reference Frames

JOHN LIPINSKI, JOHN P. SPENCER,
AND LARISSA K. SAMUELSON

Representations of words are often viewed as discrete and static, while those of sensorimotor systems are seen as continuous and dynamic, a distinction mirroring the larger contrast between amodal and perceptual symbol systems. Spatial language provides an effective domain in which to examine the connection between non-linguistic and linguistic systems because it is an unambiguous case of linguistic and sensorimotor systems coming together. To this end, we reconsider foundational work in spatial language by Hayward & Tarr (1995) and Crawford and colleagues (2000) which emphasizes representation in the abstract. In particular, we use a process-based theory of spatial working memory—the Dynamic Field Theory—to generate and test novel predictions regarding the time-dependent link between spatial memory and spatial language. Our analysis and empirical findings suggest that focusing on the processes underlying spatial language, rather than representations *per se*, can produce more constrained theories of the connection between sensorimotor and linguistic systems.

6.1 Introduction

A fundamental issue in the study of language is the relationship between the representations of words and sensorimotor systems that necessarily operate in the real world in real time (Barsalou 1999; Harnad 1990). Representations of words are typically viewed as discrete, arbitrary, and static, while sensorimotor

systems typically trade in continuous, non-arbitrary, and dynamic representations. From a theoretical standpoint, the challenge is to understand how two such seemingly different representational formats communicate with each other (Bridgeman, Gemmer, Forsman, & Huemer 2000; Bridgeman, Peery, & Anand 1997; Jackendoff 1996). The domain of spatial language is an ideal testing ground for proposals addressing this representational gap precisely because it is an unambiguous case of linguistic and sensorimotor systems coming together.

Within the field of spatial language, the issue of representational formats has a long, rich history, from the extensive linguistic analysis by Talmy (1983), who argued that schematic representations underlie spatial term use, to more recent efforts that have examined the real-time activation of linguistic representations by sensory inputs (Spivey-Knowlton, Tanenhaus, Eberhard, & Sedivy 1998). This diversity of approaches has led to a diversity of perspectives regarding the nature of the relationship between spatial language on the one hand and spatial perception, spatial memory, and spatial action on the other hand. Some researchers contend that linguistic and non-linguistic representations overlap in fundamental ways (Avraamides 2003; Hayward & Tarr 1995; Loomis, Lippa, Klatzky, & Golledge 2002), while other researchers contend that these are distinctly different classes of representation (Crawford, Regier, & Huttenlocher 2000; Jackendoff 1996).

Although the rich literature on spatial representations has led to important insights about the nature of linguistic and non-linguistic spatial systems, the central thesis of the present chapter is that this work suffers from a heavy emphasis on static representations. This, combined with the often conceptual nature of the theories proposed in the spatial language domain, leads to theories that are under-constrained and empirical findings that can be interpreted in multiple ways. We contend that the current state of affairs warrants a new approach that emphasizes the *processes that give rise to representational states*, that is, the second-to-second processes that connect the sensorimotor to the cognitive—both linguistic and non-linguistic—in the context of a specific task. We use the term ‘representational state’ to contrast our emphasis on process with previous work that has emphasized static representations. A representational state by our view is a time-dependent state in which a particular pattern of neural activation that reflects, for instance, some event in the world is re-presented to the nervous system in the absence of the input that specified that event. Note that this view of re-presentation is related to recent ideas that the brain runs ‘simulations’ of past events during many cognitive tasks (see e.g. Damasio & Damasio 1994; for further discussion see Johnson, Spencer, & Schöner, in press; Spencer & Schöner 2003).

There are three key advantages to emphasizing the processes that give rise to representational states. First, process models are more constrained than models that focus primarily on static representations because they must specify two things: the processes that give rise to representational states as well as the nature of the representational states themselves. In our experience, handling the first issue provides strong constraints on possible answers to the second issue (Spencer & Schöner 2003). Second, theories that focus too narrowly on static representations tend to sidestep the central issue we began with: how to connect the dynamic world of the sensorimotor to the seemingly discrete world of the linguistic. By contrast, process-based theories provide useful grounding, forcing researchers to take the real-time details of the task and context seriously. Third, we contend that an emphasis on process can lead to new empirical questions and new methods to answer them. We illustrate this with a novel set of findings that probe the link between spatial language and spatial memory. These empirical efforts build upon other recent insights gained from thinking of language and cognition as ‘embodied’, that is, intricately connected with the sensorimotor world (see Barsalou 1999; Spivey-Knowlton et al. 1998; Stanfield & Zwaan 2001; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy 1995; Zwaan, Madden, Yaxley, & Aveyard 2004; Zwaan, Stanfield, & Yaxley 2002).

With our broad issues now framed, here are the details of how we will proceed. First, we give a brief overview of how the link between linguistic and non-linguistic representations has been conceptualized within the domain of spatial language (section 6.2). Although these approaches are rich conceptually, they have not provided a theoretical framework constrained enough to produce critical empirical tests (section 6.3). Next, we discuss an ongoing debate about spatial preposition use that has attempted to simplify the problem of connecting sensorimotor and linguistic systems by focusing on the representations underlying spatial language (section 6.4). Although data generated in the context of this debate are compelling, the accounts that have been proposed are under-constrained. We claim that thinking about process can shed new light on such debates. Thus, in section 6.5, we apply a new theory of spatial working memory—the Dynamic Field Theory [DFT] (Spencer & Schöner 2003; Spencer, Simmering, Schutte, & Schöner 2007)—to the issue of how people activate and use spatial information in linguistic and non-linguistic tasks. We then test some novel predictions inspired by our model (section 6.6). Finally, we return to the wider literature and highlight some implications of our process-based approach as well as some of the future challenges for our viewpoint (section 6.7).

6.2 Two approaches to the linguistic/non-linguistic connection

A fundamental strength of language is its ability to connect abstract symbols that refer to objects in the real world to the dynamic sensorimotor systems that perceive and interact with these objects. 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 speak to this issue of the linguistic/non-linguistic connection in spatial language: amodal symbol systems and perceptual symbol systems.

6.2.1 Amodal symbol systems

Amodal symbol systems presume representational independence between symbolic processes like language and sensorimotor systems (Anderson 2000; Harnad 1990). 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 (Jackendoff 1992; 1996; 2002). Representational interfaces account for communication between different types of representation (e.g. verbal and visual) by proposing 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 to encode spatial representations while still capturing the core characteristics of the symbolic view (e.g. pointers to sensory modalities, type-token distinctions, taxonomies).

There is significant empirical support for this view. Consistent with Jackendoff’s representational interface, for example, Talmy (1983) 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. In the sentence ‘The bike stood near the house’, for example, Talmy shows that 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, such a linguistic representation of a relational state can be extended to a variety of visual scenes and objects without much regard to the individual object characteristics (Landau & Jackendoff 1993).

6.2.2 Perceptual symbol systems

In contrast to the transduction view of the amodal approach, Barsalou’s Perceptual Symbol Systems [PSS] (1999) posits a more intricate connection

between the linguistic and non-linguistic. By this view, transduction is not needed because symbols—perceptual symbols—arise from the same neural states that underlie perception. In particular, perceptual symbols arise when top-down processes partially reactivate sensorimotor areas and, over time, organize perceptual memories around a common frame. Once such a frame is established, the perceptual components of the frame can be reactivated, forming a ‘simulator’ that captures key elements of past experiences as well as core symbolic aspects of behavior such as productivity, type-token distinctions, and hierarchical relations. In this way, perceptual symbols are both inherently grounded in the cortical activations produced by a given sensory modality and capable of replicating the flexible, productive, and hierarchical capacities of amodal symbolic systems. Moreover, because these symbols are grounded in sensorimotor processes, they do not require pointers or transduction to become ‘meaningful’.

A growing empirical literature supports Barsalou’s (1999) PSS. For example, Stanfield and Zwaan (2001) argued that if symbolic, linguistic representations are integrated with perceptual 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’). Additional evidence for the tight integration of visual and linguistic representations comes from head-mounted eye-tracking data acquired during linguistic processing tasks. Such data show that eye movements used to scan a visual scene are time-locked to verbal instructions to pick up items within that scene (Spivey-Knowlton et al. 1998). Visual information has also been shown to facilitate real-time resolution of temporarily syntactically ambiguous sentences (Tanenhaus et al. 1995)—further evidence against a hard separation between linguistic and sensory systems. Finally, work by Richardson, Spivey, Barsalou, & McRae (2003) shows that spatially grounded verbal stimuli interact with visual discrimination performance, providing additional evidence that linguistic processing can directly impact the processing of visual space.

6.3 Limits of the amodal and perceptual symbols system approaches

The amodal and PSS views are opposites conceptually; however, both perspectives appear to be substantially supported within the spatial language

domain. This is not an acceptable state of affairs, because two opposing perspectives proposed to account for the same phenomena cannot both be correct. For instance, if the PSS view were correct, amodal symbols would be superfluous because symbolic processes would fall out of the organization of dynamic, schematic records of neural activation that arise during perception (Barsalou 1999). Thus, despite a vigorous debate and valuable empirical data on both sides, the fundamental question of how spatial linguistic and non-linguistic systems are connected remains unanswered. Further consideration suggests a critical limitation of these proposals: the amodal and PSS views rely on descriptive, conceptual accounts of the linguistic/non-linguistic connection. Though often useful 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, in some cases, be reinterpreted by the other view. Jackendoff (2002), for example, has explained the real-time resolution of syntactic ambiguity through visual processing (Tanenhaus et al. 1995) using characteristics of a syntax-semantics interface.

Conceptual theories are particularly problematic in the context of the linguistic/non-linguistic connection because of the complexity of the theoretical terrain: these theories must explain the process that unites spatial terms with spatial perception, memory, and action. More concretely, such theories have to specify how people perceive a scene, how they identify key spatial relations such as the relation between a target object and a reference object, how such spatial relations are remembered in the context of real-time action of both the observer and the environment, and how these relations are used in discourse to produce a verbal description sufficiently detailed to allow another person to act on that information. The conceptual theories discussed above make reference to processes involved in such situations—transduction processes on one hand, simulation processes on the other—but the formal details of these processes are lacking. Given the complexity of what these theories have to accomplish, this is not surprising.

Although a formal theory seems relatively distant at present, we can ask a simpler question: what might a formal theory of such processes look like? Barsalou's (1999) move to embrace neural reality seems particularly appealing in that it highlights possible connections among conceptual theory (e.g. the PSS view), neurally inspired formal theories (e.g. neural network approaches), and data (e.g. fMRI or single-unit recordings). Indeed, there are several neurally plausible theories of key elements of the linguistic/non-linguistic connection (e.g. Cohen, Braver, & O'Reilly 1996; Gruber & Goschke 2004; Gupta & MacWhinney 1997; Gupta, MacWhinney,

Feldman, & Sacco 2003; McClelland, McNaughton, & O'Reilly 1995; O'Reilly & Munakata 2000). Although these potential links are exciting, they are also daunting given the added complexities of dealing with a densely interconnected and highly non-linear nervous system (e.g. Freeman 2000). For instance, how might a population of neurons that encodes a particular spatial relation link up with other populations that deal with lexical and semantic information? And how might these different populations allow their patterns of activation to mingle and integrate, while at the same time stably maintaining their own unique content in the face of neural noise and changing environments (Johnson et al. in press; Spencer & Schöner 2003; Spencer et al. 2007)?

Perhaps on account of this daunting picture, many researchers have split the linguistic/non-linguistic connection problem up into two parts: (1) what is the nature of the representations used by linguistic and non-linguistic systems, and (2) how are they connected? Within this framework, the vast majority of research has focused on the first question: the representational format used by spatial perception, action, and memory on one hand and spatial language on the other. Although, as before, this view has generated many insightful empirical findings (some of which we describe below), it has led to under-constrained theories of the representations that support performance. We contend that this is a natural by-product of emphasizing representations in the abstract, rather than the processes that give rise to representational states. Moreover, we claim that the latter approach ultimately leads to more constrained theories and, perhaps, a richer view of how the sensorimotor and the linguistic connect.

To illustrate both the limitations of the 'abstract representation' view and the potential of a more process-based approach, we turn to an ongoing debate on spatial prepositions. Within this domain, one group of researchers has claimed that people use overlapping representations in linguistic and non-linguistic tasks, while a second group has claimed that different representations are used in these two types of task. Importantly, both sets of claims focus on representations in the abstract. We then sketch a different view by applying our neurally inspired model of spatial working memory—the Dynamic Field Theory (DFT)—to the issue of how people activate and use spatial information in linguistic tasks. Our analysis suggests that linguistic and non-linguistic behavior can arise from a single, integrated system that has a representational format different from what previous researchers have claimed. We then test some novel implications of our model to highlight the fact that a process-based view offers new ways to probe the linguistic/non-linguistic connection.

6.4 Missing the connection: the challenges of focusing on representation in the abstract

6.4.1 *Hayward & Tarr (1995): shared linguistic and perceptual representations of space*

To explore the possible connections between linguistic and sensorimotor representations of space, Hayward & Tarr (1995) examined how object relations are linguistically and visually encoded. Participants were presented with a visual scene depicting a referent object and a target object that appeared in varying locations. Participants were 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 declined as target positions deviated from the respective axes.

Next, Hayward & Tarr extended 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 more graded quantification and metric manipulation of linguistic representations beyond the standard gross linguistic output (e.g. 'above'/'not above'). It therefore provides a means of empirically bridging the gap between metric, dynamic sensorimotor representations and discrete 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 & 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 & Tarr examined performance on location memory and same-different discrimination tasks. Importantly, they found that the areas of highest spatial recall accuracy were 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 sensorimotor systems with spatial prototypes along the cardinal axes. Such prototypes lead to high linguistic ratings and a high degree of accuracy in sensorimotor tasks for targets aligned with the axes.

6.4.2 Crawford, Regier, and Huttenlocher (2000): distinct linguistic and perceptual representations of space

Results from Crawford et al. (2000) present a different picture. Like Hayward & Tarr, these researchers probed both linguistic and non-linguistic representations of space by analyzing 'above' ratings as well as spatial memory performance. Results showed an 'above' ratings gradient aligned with the vertical axis similar to that of Hayward & Tarr (1995). Counter to the claims of representational similarity, however, Crawford et al. also found that location memory responses were biased *away* from the vertical axis when participants had to recall the locations of targets to the left and right of this axis. To account for these data, Crawford and colleagues proposed that the cardinal axes function as *prototypes* in the linguistic task (see Figure 6.1a) but serve as *category boundaries* in the spatial memory task (Figure 6.1b). Moreover, the diagonal

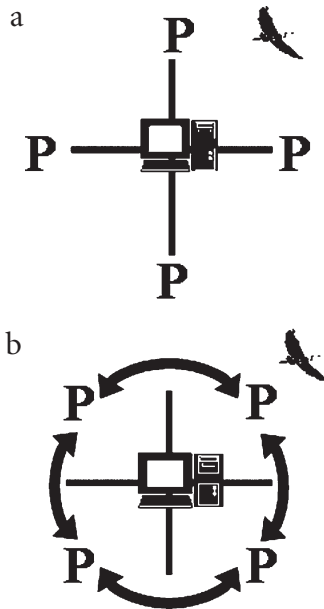


FIGURE 6.1 (a) Proposed layout of spatial prototypes (P) relative to a reference object (computer) and a target object (bird) in the linguistic task from Hayward & Tarr (1995) and Crawford et al. (2000). According to Hayward & Tarr, the same figure captures spatial prototypes in the non-linguistic task. (b) Proposed layout of spatial prototypes in non-linguistic tasks according to Crawford et al. Arrows in (b) indicate direction of bias in the spatial recall task. Lines in (b) indicate location of category boundaries.

axes in the task space, while serving no particular function in the linguistic task, serve as prototypes for spatial memory (Figure 6.1b) (Engebretson & Huttenlocher 1996; Huttenlocher, Hedges, & Duncan 1991). Thus, while both linguistic and non-linguistic spatial representations use the cardinal axes, these axes serve functionally distinct representational roles in the two tasks. It appears, therefore, that linguistic and non-linguistic representations of space differ in critical ways.

6.4.3 A prototypical debate

Results of the studies described above suggest that the cardinal axes serve as prototypical locations for spatial prepositions like ‘above’. At issue, however, is what accounts for performance in the non-linguistic tasks—prototypes along the cardinal axes (Figure 6.1a) or prototypes along the diagonals (Figure 6.1b)? Both sets of researchers present standard evidence of prototype effects—graded performance around some special spatial locations. The challenge is that there appear to be two sets of special locations. Specifically, recall accuracy is highest when targets are near the cardinal axes and declines systematically as the target object is moved away from the axes, while at the same time bias is largest near the cardinal axes, declining systematically as one moves closer to the diagonal axes. Given these two sets of special locations—near the cardinal axes and near the diagonal axes—how do we know which layout of prototypes is correct?

Crawford et al. seem to present a compelling case by focusing on a critical issue: what goes into making a recall response in these tasks? In particular, Crawford and colleagues used their Category Adjustment (CA) model to explain why adults’ responses are biased away from cardinal axes and toward the diagonals. According to this model, people encode two types of spatial information in recall tasks: *fine-grained* information about the target location (e.g. angular deviation) and the region or *category* in which the target is located. Data from a variety of studies suggest that adults tend to subdivide space using vertical and horizontal axes (Engebretson & Huttenlocher 1996; Huttenlocher et al. 1991; Nelson & Chaiklin 1980). This places prototypes at the centers of these regions, that is, along the diagonals of the task space. At recall, fine-grained and categorical information are combined to produce a response. Importantly, these two types of information can be weighted differently. If, for example, fine-grained information is uncertain (as is the case after short-term delays), categorical information can be weighted more heavily, resulting in a bias toward the prototype of the category. This accounts for the bias toward the diagonals in Crawford et al. (2000). It also accounts for the improved accuracy along the cardinal axes, because recall of targets

aligned with a category boundary can be quite accurate (see Huttenlocher et al. 1991).

Given that Crawford et al. grounded their account of spatial memory biases in a formal theory of spatial recall that does not use prototypes along the cardinal axes, it appears that there are important differences in the representations underlying linguistic and non-linguistic performance. However, there are two limitations to this story. The first has to do with constraints provided by the CA model. Although this model can explain the biases that arise in recall tasks once one has specified the location of category boundaries, prototypes, and the certainty of fine-grained and categorical information, we do not know the processes that specify these things. That is, we do not know the factors that determine where category boundaries should go, what factors influence the certainty of spatial information, and so on. More recent work has documented some of these factors (Hund, Plumert, & Benney 2002; Plumert & Hund 2001; Spencer & Hund 2003; Spencer et al. 2007) but these details are not specified *a priori* by the CA model (for a recent modification of the CA model in this direction, see Hund & Plumert 2002; 2003).

Why are these details important? In the context of the linguistic/non-linguistic debate, this issue is central because both spatial language and spatial memory use the cardinal axes in some way. Specifying precisely what these axes do in both cases and how these axes are linked up to the representational states in question is critical if we are to evaluate the different claims. Put differently, we contend that it is important to specify the process that links the sensorimotor (e.g. perception of the cardinal and diagonal symmetry axes) and the cognitive (e.g. spatial prototypes). Note that this critique of the CA model does not indicate that this model is incorrect. Rather, we think the time is ripe to move the ideas captured by this model to the next level, that is, to the level of process.

A second limitation of the Crawford et al. story is that it fails to specify what is happening on the linguistic side: neither Crawford et al. nor Hayward and Tarr provided a formalized theory of spatial language performance. A recent model proposed by Regier and Carlson (2001)—the AVS model—specifies how prototypicality effects might arise in ratings tasks. Interestingly, this model can account for prototypicality effects without using prototypes *per se*. Rather, this model scales ratings by the difference between an attentionally weighted vector from the reference object to the target object and the cardinal axes in question (e.g. the vertical axis in the case of ‘above’ ratings). Thus, this model moves closer to explaining how ratings performance arises from processes that link the cardinal axes to representations of the target location. Unfortunately, this model says nothing about spatial recall performance. As such, it is not possible to directly compare the CA account of spatial memory biases and the AVS account of ratings performance.

In the sections that follow, we describe a new theory of spatial working memory that we contend can overcome both limitations described above. In particular, this model overcomes the limitation of the CA model by specifying how perception of symmetry axes is linked to the representational states associated with target locations in spatial recall tasks. The critical insight here is that we can account for both accuracy along the cardinal axes and bias away from the cardinal axes without postulating category boundaries and prototypes; rather, such effects arise due to the coupling between perception of reference axes in the task space—visible edges and axes of symmetry—with working memory processes that serve to actively maintain location information. With regard to the second limitation—the absence of a formal model of both spatial recall and spatial preposition use—we sketch an extension of our model that can account for prototypicality effects in linguistic ratings tasks. Although this extension requires further development (which we point toward in the conclusions section), it is generative enough at present to produce novel predictions which we test empirically.

6.5 A process approach to the linguistic/non-linguistic connection

6.5.1 *The Dynamic Field Theory: a process account of spatial working memory*

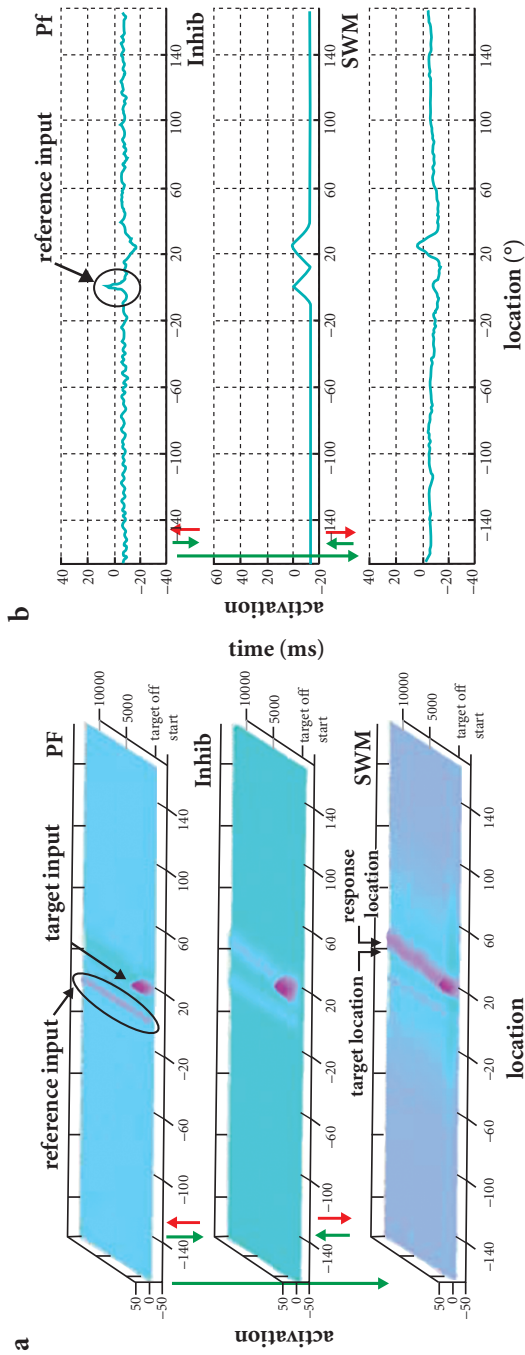
Data from Hayward and Tarr (1995) and Crawford et al. (2000) point toward two types of prototypicality effects in spatial memory—higher accuracy and greater bias near cardinal axes. Although the CA model explains these biases using two types of representation—boundaries and prototypes—our Dynamic Field Theory (DFT) suggests that both effects actually arise from the interaction of perceived reference frames and information actively maintained in spatial working memory (Simmering, Schutte, & Schöner 2008; Spencer & Schöner 2003; Spencer et al. 2007). That is, the DFT provides a formalized process account of spatial memory bias away from reference axes *without positing prototypes*.

The DFT is a dynamic systems approach to spatial cognition instantiated in a particular type of neural network called a dynamic neural field. The DFT accounts for the spatial recall performance of younger children (2–3 years), older children (6 and 11 years), and adults (see Spencer et al. 2007). A simulation of the DFT performing a single spatial recall trial is shown in Plate 1. The model is made up of several layers (or fields) of neurons. In each layer, the neurons are lined up along the x-axis according to their ‘preferred’ location, that is, the location in space that produces maximal activation of each neuron. The activation of each neuron is plotted along the y-axis, and time is on the

z-axis. The top layer in each panel is the perceptual field, PF. This field captures perceived events in the task space, such as the appearance of a target, as well as any stable perceptual cues in the task space, such as the midline symmetry axis probed in many studies of spatial recall. This layer sends excitation to both of the other layers (see green arrows). The third layer, SWM, is the working memory field. This field receives weak input from perceived events in the task space and stronger input from the perceptual field. The SWM field is primarily responsible for maintaining a memory of the target location through self-sustaining activation—a neurally plausible mechanism for the maintenance of task-relevant information in populations of neurons (Amari 1989; Amari & Arbib 1977; Compte, Brunel, Goldman-Rakic, & Wang 2000; Trappenberg, Dorris, Munoz, & Klein 2001). The second layer, Inhib, is an inhibitory layer that receives input from and projects inhibition broadly back to both the perceptual field and the working memory field. Note that the layered structure shown in Figure 6.2 was inspired by the cytoarchitecture of visual cortex (see Douglas & Martin 1998). Note also that our full theory of spatial cognition includes longer-term memory layers that we will not consider here because they do not affect the hypotheses tested below (for an overview of the full model, see Spencer et al. 2007).

The working memory field, SWM, is able to maintain an activation pattern because of the way the neurons interact with each other. Specifically, neurons that are activated excite neurons that code for locations that are close by, and—through the Inhib layer—inhibit neurons that code for locations that are far away. The result is an emergent form of local excitation/lateral inhibition which sustains activation in working memory in the absence of inputs from the perceptual layer (see Amari 1989; Amari & Arbib 1977; Compte et al. 2000 for neural network models that use similar dynamics).

Considered together, the layers in Plate 1 capture the real-time processes that underlie performance on a single spatial recall trial. At the start of the trial, there is activation in the perceptual field associated with perceived reference axes in the task space (see reference input arrow in Plate 1a), for instance, visible edges and axes of symmetry (Palmer & Hemenway 1978; Wenderoth & van der Zwan 1991). This is a weak input and is not strong enough to generate a self-sustaining peak in the SWM field, though it does create an activation peak in perceptual field. Next, the target turns on and creates a strong peak in PF which drives up activation at associated sites in the SWM field (see target input arrow in Plate 1a). When the target turns off, the target activation in PF dies out, but the target-related peak of activation remains stable in SWM. In addition, activation associated with the reference axis continues to influence the PF because the reference axis is supported by readily available perceptual cues (see peak in PF during the delay).



1. A simulation of the Dynamic Field Theory performing a single spatial recall trial. Panels in (a) represent: perceptual field [PF]; inhibitory field [Inhib]; spatial working memory field [SWM]. Arrows represent interaction between fields. Green arrows represent excitatory connections and red arrows represent inhibitory connections. In each field, location is represented along the x-axis (with midline at location 0), activation along the y-axis, and time along the z-axis. The trial begins at the front of the figure and moves toward the back. (b) Time slices through PF, Inhib, and SWM at the end of the delay for the model shown in (a). See text for additional details.

Central to the recall biases reported by Huttenlocher and colleagues (1991) is how reference-related perceptual input affects neurons in the working memory field during the delay. Figure 6.2b shows a time slice of the SWM field at the end of the delay. As can be seen in the figure, the working memory peak has slightly lower activation on the left side. This lower activation is due to the strong inhibition around midline created by the reference-related peak in PF (see circle in Plate 1b). The greater inhibition on the left side of the peak in SWM effectively ‘pushes’ the peak away from midline during the delay. Note that working memory peaks are not always dominated by inhibition as in Figure 6.2b. For instance, if the working memory peak were positioned very close to or aligned with midline (location *o*), it would be either attracted toward or stabilized by the excitatory reference input. This explains why spatial recall performance is quite accurate for targets aligned with the cardinal axes (for related results, see Engebretson & Huttenlocher 1996; Hund & Spencer 2003; Spencer & Hund 2003).

In summary, the DFT provides a process-based alternative to the CA model. Critically, the DFT links spatial memory biases to a process that integrates remembered information in working memory with perceived reference frames—the cardinal axes of the task space—the same reference frames implicated in linguistic performance. As a result, Crawford et al.’s central argument against Hayward & 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 possibility that there is a direct link between spatial memory and spatial language.

6.5.2 *Connecting the Dynamic Field Theory and spatial language*

Given that we have proposed a process-based account of the link between cardinal axes and spatial memory, we can ask whether this proposed link between the sensorimotor and the cognitive can be extended to the case of spatial language. The central issue this raises is: what is the connection between the representational states associated with space captured by our theory and the representational states underlying words? A simple way to conceptualize this link is depicted in Figure 6.2, which captures the use of the spatial preposition ‘above’ to describe a target presented at -20° . This figure shows the working memory field depicted in Figure 6.2b reciprocally coupled to a linguistic node that represents the label ‘above’. The -20° target location is captured by the Mexican-hat-shaped activation distribution which arises from the locally excitatory interactions among neurons in the SWM layer and lateral inhibition from activation in the Inhib layer. The

forward projection from SWM to the ‘above’ node is spatially structured by systematically varying the connection strengths (captured by the Gaussian distribution of connection lengths) around the vertical axis. In particular, neurons in SWM associated with the vertical axis (location *o*) project activation most strongly onto the ‘above’ node, while neurons to the far left and right of the excitatory field project activation quite weakly onto this node. These variations in synaptic strength are meant to reflect the long-term statistical probabilities of spatial preposition use. In particular, we hypothesize that over the course of development, ‘above’ is used most often when referring to cases where a target object is close to a vertical axis and less often when a target object is to the far left and right of a vertical axis. This is consistent with findings from Hayward & Tarr (1995) showing that spontaneous use of prepositions like ‘above’ and ‘over’ declines as target objects diverge systematically from a vertical or ‘midline’ axis (see also Franklin & Henkel 1995). Note that the strength of the projection gradient depicted in Figure 6.2 is somewhat arbitrary: the gradient does not have to be very strong for our account of spatial language performance to work (see below). Note also that we only consider the forward projection from SWM to the ‘above’ node in this chapter. We view the coupling between spatial memory and spatial language as reciprocal in nature; thus, the vectors in Figure 6.2 go in both directions. The details of this reciprocal coupling, however, are beyond the scope of the present chapter (see Lipinski, Spencer, & Samuelson 2009b for an empirical probe of the reciprocal nature of these connections).

How can the model depicted in Figure 6.2 be applied to capture performance in spatial language tasks? Essentially, this model provides an account for why some locations might be perceived to be better examples of ‘above’ than others. In particular, a target-related peak of activation in SWM close to the vertical axis (e.g. the activation peak in Figure 6.2 to the left of location *o*) would strongly activate the ‘above’ node. By contrast, a target-related peak in SWM far from the axis would weakly activate the ‘above’ node. We turn these activations into a linguistic ‘above’ rating by scaling the amount of activation to the magnitude of the rating. Concretely, ratings should be highest when targets are aligned with the vertical axis, and should fall off systematically as peaks of activation in SWM are shifted to the left or right. This is similar to the approach adopted by Regier & Carlson’s AVS model (2001). Although this account of ratings performance is, admittedly, simplistic, we contend that it has a clear strength: it grounds linguistic performance in the real-time events that occur in spatial language tasks, and places primary emphasis on the real-time activation of lexical representational states that are reciprocally coupled to spatial working memory. In the next section, we empirically demonstrate

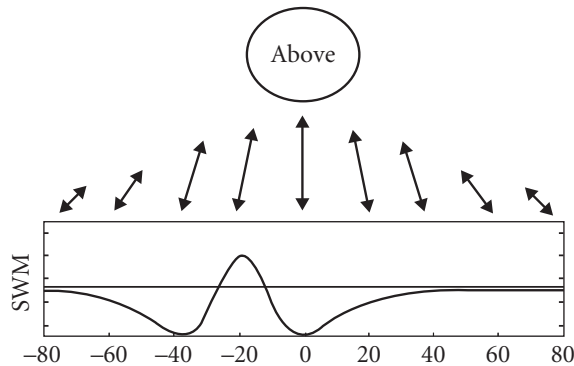


FIGURE 6.2. Proposed reciprocal coupling between the working memory field in Plate 1 and a linguistic node representing the label ‘above’. The projections between SWM and this node are spatially structured by systematically varying the connection strengths (captured here by the Gaussian distribution of connection lengths; see dashed line) around the vertical axis (location 0). The -20° target location is captured by the Mexican-hat-shaped activation distribution. See text for additional details.

that this emphasis on process can shed new light on what is happening in spatial language tasks.

6.6 An empirical test of the DFT approach to spatial language

Inspired by the model sketched in Figure 6.2, we recently conducted a study designed to investigate whether linguistic and non-linguistic processes are temporally connected in spatial tasks. In particular, we asked whether the processes that create delay-dependent spatial drift in spatial working memory might also leave empirical signatures in a spatial language task. Toward this end, we used the ratings task from Hayward & Tarr (1995), given its capacity to reveal quantifiable metric effects and its centrality in the spatial language literature (e.g. Crawford et al. 2000; Hayward & Tarr 1995; Logan & Sadler 1996; Regier & Carlson 2001). We predicted that if spatial language and spatial memory are coupled together as shown in Figure 6.2, then ‘above’ ratings should become systematically lower for targets to the left and right of the vertical axis as memory delays increase—that is, the ‘above’ node should become systematically less active as peaks of activation in SWM drift away from the vertical axis. Furthermore, the variability of ratings performance should increase over delays and be systematically lower when participants rate targets aligned with the cardinal axes. These predictions regarding response variability mirror

effects we have reported in our previous studies of spatial recall (e.g. Spencer & Hund 2002).

6.6.1 Design and methods

To test this prediction, we used a variant of the basic ‘spaceship’ task used in our previous spatial working memory studies (e.g. Schutte & Spencer 2002; Spencer & Hund 2002; 2003). Participants were seated at a large (0.921×1.194 m), opaque, homogeneous tabletop. Experimental sessions were conducted in a dimly lit room with black curtains covering all external landmarks. In addition, a curved border was added to occlude the corners of the table, thereby occluding the diagonal symmetry axes. Thus, visible reference cues included the edges of the table and its axes of symmetry as well as the objects included in our visual displays (see below).

On each trial, a single reference disk appeared along the midline (i.e. ‘vertical’) symmetry axis, 30 cm in front of the participant. This disk remained visible throughout each trial. Next, the participant moved a computer mouse on top of this disk and a random number between 100 and 500 appeared in the center of the table. Participants were instructed to count backwards by 1’s from this number until the computer prompted them to make a response. This counting task occupied verbal working memory, preventing participants from verbally encoding and maintaining the position of the spaceship on trials with a memory delay. This was important because we wanted to examine whether verbal performance would show evidence of delay-dependent ‘drift’. This also took care of a potentially important experimental confound in Hayward & Tarr (1995) and Crawford et al. (2000). In both of these studies, the verbal responses could be formulated while the target was visible; spatial recall responses, on the other hand, were given after a memory delay. Thus, any differences between spatial language and spatial memory performance might be simply due to processing in the absence of a memory delay in one task and processing following a delay in the other.

After participants started counting, a small, spaceship-shaped target appeared on the table for 2 sec. Next, participants gave a response based on one of two prompts spoken by the computer. For *spatial memory trials*, participants moved the mouse cursor to the remembered target location when the computer said ‘Ready-Set-Go’. For *spatial language rating trials*, participants gave a verbal rating when the computer said ‘Please give your “Above” rating.’ The computer prompts were both 1,500 msec. in duration. On ratings trials, participants rated on a scale of 1 (‘definitely not above’) to 9 (‘definitely above’) the extent to which the sentence ‘The ship is ABOVE the dot’ described the spaceship’s location relative to the reference disk. Ratings and recall trials were randomly

intermixed, and responses were generated following a 0 sec. or 10 sec. delay. In particular, in the No Delay condition, the end of the computer prompt coincided with the disappearance of the target, while in the Delay condition, the prompt ended 10 sec. after the disappearance of the target. Targets appeared at a constant radius of 15 cm relative to the reference disk and at 19 different locations relative to the midline axis (0°): every 10° from -70° to $+70^\circ$ as well as $\pm 90^\circ$ and $\pm 110^\circ$.

6.6.2 Results and discussion

Figure 6.3a shows mean directional errors on the memory trials across target locations and delays. Positive errors indicate clockwise errors relative to midline (vertical), while negative errors indicate counterclockwise errors. As can be seen in the figure, participants' responses were quite accurate in the No Delay condition. After 10 sec., however, responses to targets to the left and right of midline were systematically biased away from this axis (see also Spencer & Hund 2002). This bias gradually increased and then decreased as targets moved away from midline, reducing considerably at the horizontal or left-to-right axis (i.e. $\pm 90^\circ$). These data were analyzed in an ANOVA with Target and Delay as within-subject factors. This analysis revealed a significant main effect of Target, $F(18\ 234) = 20.6, p < .001$, as well as a significant Delay by Target interaction, $F(18\ 234) = 19.4, p < .001$. This interaction is clearly evident in Figure 6.3a.

Similar results were obtained in analyses of response variability (standard deviations of performance to each target at each delay; see Figure 6.3b). There were significant main effects of Delay, $F(1\ 13) = 172.3, p < .001$, and Target, $F(18\ 234) = 5.4, p < .001$, as well as a significant Delay by Target interaction, $F(18\ 234) = 3.4, p < .001$. As can be seen in Figure 6.3b, variability was higher in the 10 sec. delay condition, and responses to targets to the left and right of midline were more variable than responses to the targets aligned with the cardinal axes. These results are consistent with predictions of the DFT that memory for locations aligned with symmetry axes is more stable than memories for targets that show delay-dependent drift.

The first critical question was whether delay-dependent spatial drift would be evident in participants' ratings performance. Figure 6.4 shows that this was indeed the case. Overall, 'above' ratings in the spaceship task followed a gradient similar to that obtained by Hayward & Tarr (1995) and Crawford et al. (2000); however, ratings were systematically lower for targets to the left and right of midline after the delay (see Figure 6.4a). An ANOVA on these ratings data with Target and Delay as within-subjects factors revealed a significant main effect of Target, $F(18, 234) = 240.2, p < .001$. More importantly, there was a significant decrease in ratings over Delay, $F(1, 13) = 12.5, p = .004$, as well as a trend toward a Delay

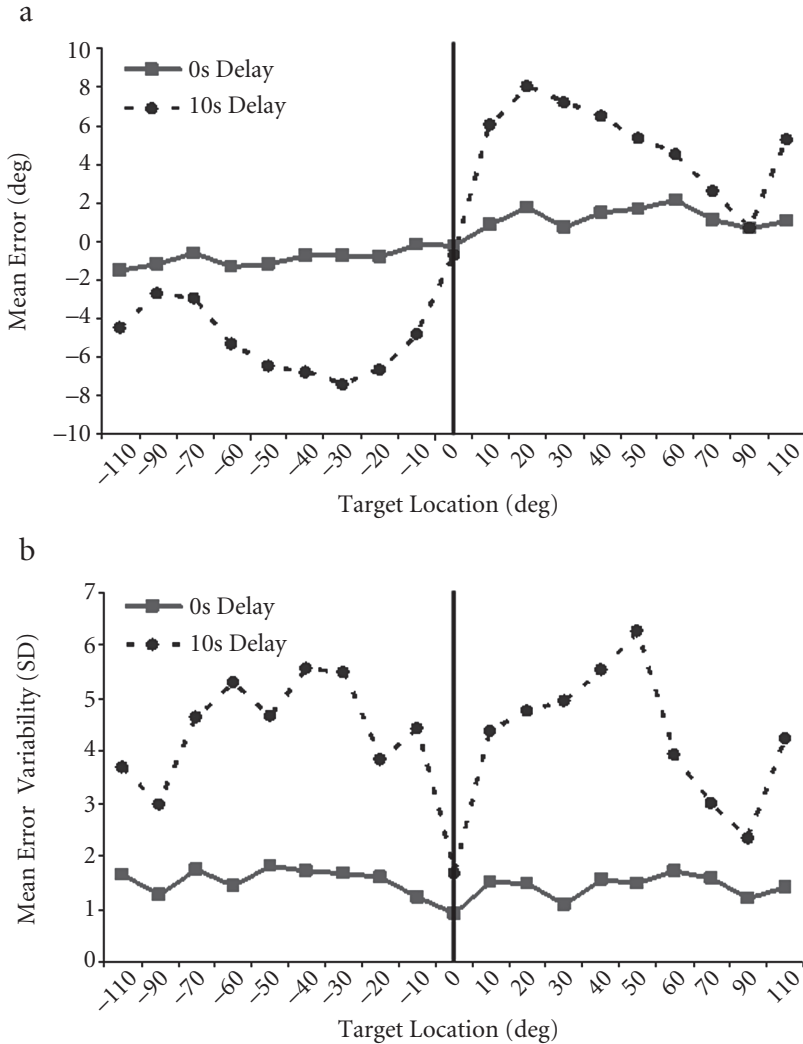


FIGURE 6.3 (a) Mean directional error across target locations for No Delay (0 sec.; solid line) and Delay (10 sec.; dashed line) location memory trials. Positive errors indicate clockwise errors and negative errors indicate counter-clockwise errors. (b) Mean error variability (SDs) for No Delay (0 sec.; solid line) and Delay (10 sec.; dashed line) location memory trials. Solid vertical line in each panel marks the midline of the task space.

by Target interaction, $F(18, 234) = 1.5, p < .10$. This systematic decrease in ratings responses as a function of delay—particularly for targets to the left and right of the reference axis—is consistent with the proposal that there is a shared representational process used in both the spatial memory and spatial language tasks.

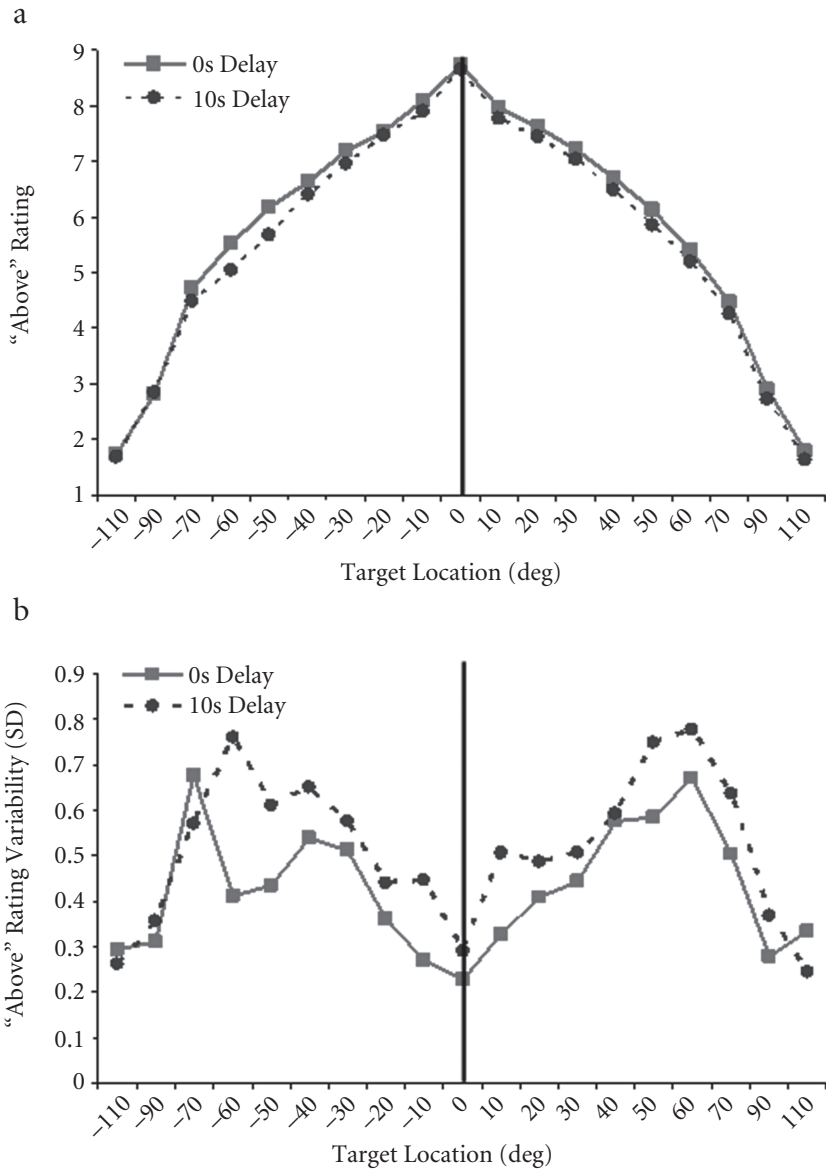


FIGURE 6.4 (a) Mean 'Above' ratings across target locations for No Delay (0 sec.; solid line) and Delay (10 sec.; dashed line) trials where '9' indicates the target is 'definitely above' the reference dot and '1' indicates the target is 'definitely NOT above' the reference dot. (b) Mean 'Above' ratings variability (SDs) for No Delay (0 sec.; solid line) and Delay (10 sec.; dashed line) trials. Solid vertical line in each panel marks the midline of the task space.

Given the effects of delay on response variability *and* bias in the spatial memory task, a second critical question is whether such variability effects would also emerge in our analyses of ratings performance. If ratings performance failed to reflect the same pattern of variability as that established in spatial memory (namely, lower variability for targets appearing along the vertical axis), it would indicate some difference in the underlying representational processes required for the linguistic and non-linguistic spatial tasks. If, on the other hand, the same general pattern is obtained, it bolsters our claim that both tasks rely on the same underlying representational process. Our analyses of ratings variability were consistent with the latter, showing significant main effects of both Target, $F(18\ 234) = 3.4, p < .001$, and Delay, $F(1\ 13) = 8.8, p = .01$. As can be seen in Figure 6.4b, the variability in ratings performance was lower for targets aligned with the cardinal axes, and systematically increased as targets moved away from midline. Moreover, variability increased systematically over delay. These findings are similar to results obtained in the spatial memory task.

Overall, the similar effects of delay and target for the spatial memory and spatial language tasks points toward a shared representational process for both tasks. However, in contrast to the large delay effects in the spatial memory task (see Figure 6.3a), the effect of delay on ratings means in Figure 6.4a appears small. Given this, it is important to ask whether the significant delay effect in the ratings task is, in fact, a *meaningful* effect. To address this question, we compared spatial memory and ratings responses directly by converting the ratings 'drift' apparent in Figure 6.4a into a spatial deviation measure. In particular, for each target within the range $\pm 60^\circ$,¹ we converted the ratings data in a two-step process. To illustrate this process, consider how we converted the data for the $+10^\circ$ target, the 'anchor' location in this example. First, we took the change in ratings in the No Delay condition between the anchor (10°) and the adjacent target moving away from midline (i.e. 20°) and divided this change by 10° —the separation between adjacent targets. This indicated the amount participants changed their rating in our baseline condition (i.e. No Delay) as we moved the anchor target 10° further from midline. Second, we scaled the change in rating over delay for the anchor target by this No Delay deviation measure (e.g. conversion score for the 10° target = (change in 10 s delay rating at 10°) * 10° / (change in 0 s delay rating between 10° and 20°)).

The converted ratings data for all targets within the $\pm 60^\circ$ range are plotted in conjunction with the recall data in Figure 6.5. If the drift underlying

¹ For targets greater than 70° away from midline, adjacent targets were 20° apart. Given this change in spatial separation, we only converted the ratings data from targets $\pm 60^\circ$ from midline.

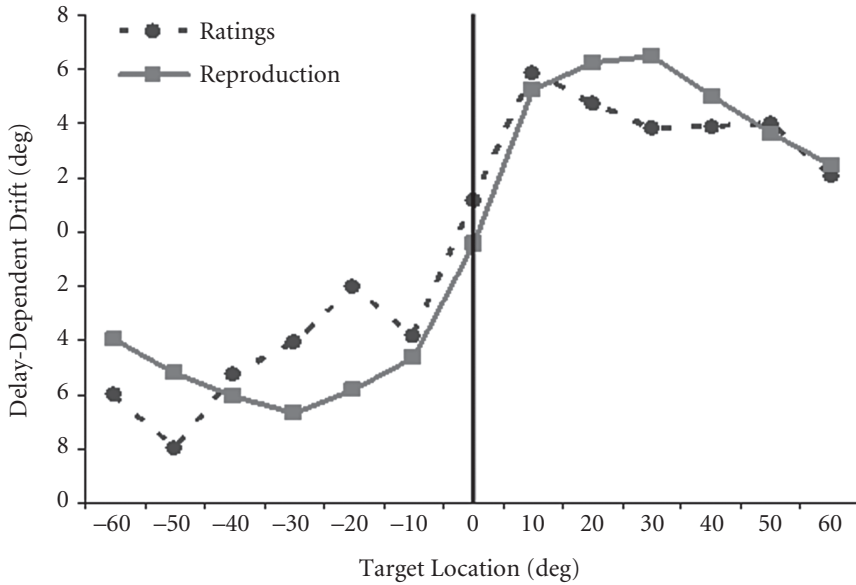


FIGURE 6.5. Comparison between location memory errors (solid line) and ratings drift converted to degrees (dashed line). Solid vertical line marks the midline of the task space. See text for details of ratings conversion method.

performance in the ratings task is produced by the same process that creates drift in the recall task, then these data should line up. Although differences in performance across tasks do exist, the converted ratings data show remarkable overlap with the recall data across target locations. This provides strong initial support for the prediction we generated from the modified dynamic field model shown in Plate 1, suggesting that a shared working memory process underlies performance in both tasks.

6.7 Conclusions

Understanding the relationship between linguistic and non-linguistic systems is a critical issue within cognitive science. Spatial language is of central importance here because it is an unambiguous case of these putatively different systems coming together. Although recent efforts have advanced our understanding of the link between spatial language and memory, we have argued in this chapter that previous approaches are limited in two related ways: these approaches have focused too narrowly on static representation and have led to under-constrained theories. To illustrate an alternative approach, we

presented an overview of our dynamic field theory of spatial working memory and applied this process model to the use of spatial prepositions. Moreover, we presented preliminary empirical findings that supported a novel prediction of this model—that linguistic ratings would show signatures of delay-dependent ‘drift’ in both changes in mean ratings over delay and response variability. We contend that these results demonstrate the utility of our approach and suggest that sensorimotor and linguistic systems are intricately linked. This supports the view presented by Hayward & Tarr (1995) and others (see also Barsalou 1999; Richardson et al. 2003; Spivey-Knowlton et al. 1998; Zwaan et al. 2004) that sensorimotor and linguistic representations overlap. Importantly, however, it builds on this perspective by grounding claims about representation in a formal model that specifies the time-dependent processes linking perception of reference frames to representational states in working memory.

Although the model and data we present in this chapter support our claim that process-based approaches can shed new light on the linguistic/non-linguistic connection, these are only first steps. Clearly, there is much more theoretical and empirical work to do to demonstrate that our approach can move beyond previous accounts toward a more theoretically constrained future. In this spirit, the sections below address three questions: what have we accomplished, what remains to be accomplished, and how does our model fit with other related models in the spatial memory and spatial language literatures?

6.7.1 *The DFT and spatial language: what have we accomplished?*

The model and data presented in this chapter are firmly positioned between arguments by Hayward & Tarr (1995) and Crawford et al. (2000). On one hand, our data show a time-dependent link between spatial language and spatial memory, consistent with the claim by Hayward & Tarr (1995) that linguistic and non-linguistic representations have considerable overlap within the spatial domain. On the other hand, our work also resonates with the move toward formal models by Crawford et al. (2000). In particular, our modeling work emerged from a focus on the question originally addressed by the Category Adjustment model: what goes into making a recall response (Huttenlocher et al. 1991)? By focusing on the processes that link cardinal axes to representational states in spatial working memory, the DFT provides a new answer to this question that does not have recourse to spatial prototypes. The absence of spatial prototypes in our model allowed us to reconsider the link between performance in spatial recall and ratings tasks. We proposed a new view that directly couples SWM and the activation of label nodes representing spatial terms like ‘above’. This new view moves beyond past approaches in two key ways: (1) it grounds both recall and ratings performance in

time-dependent perceptual and working memory processes, and (2) it provides a formal account of how people generate *both* types of responses.

Importantly, we also demonstrated in this chapter that the dynamic field approach is empirically productive. We generated a set of novel predictions that ratings of targets to the left and right of midline would be lower after a short-term delay, and that response variability in the ratings task would increase over delays and be lower for targets aligned with the cardinal axes. Analyses of both mean ratings and response variability were consistent with these predictions.

These results are not trivial because we predicted *lower* ratings over delay when other views appear to predict *higher* ratings. In the CA model, for example, people rely more on spatial prototypes after short-term delays. If the spatial prototypes for language lie along the cardinal axes, as both Hayward & Tarr (1995) and Crawford et al. (2000) contend, ratings should have drifted toward these prototypes over delay—that is, people should have rated a target close to midline as a better example of ‘above’ after a delay relative to the No Delay condition. As predicted by the DFT, however, we found the opposite result. Indeed, the converted ratings data showed a high degree of overlap with spatial recall biases, suggesting that a shared process generated both types of response.

This discussion makes it clear that our model did, in fact, generate a novel prediction. But wouldn’t *any* model in which sensorimotor and linguistic representations use the same underlying process make this prediction? We contend that the answer is ‘no’ because we predicted an entire suite of effects: a decrease in ratings over delays for targets to the left and right of the vertical axis; an increase in ratings response variability over delays; and lower ratings variability for targets aligned with the cardinal axes. It is important to note in this regard that our model provides a process-based account for both mean biases *and* response variability (see Schutte, Spencer, & Schöner 2003; Schutte & Spencer in press). This is rarely the case for models of spatial memory. For comparison, the CA model has not been used in the spatial domain to make predictions about response variability (although for a model that moves in a related direction see Huttenlocher, Hedges, & Vevea 2000).

Results showing signatures of delay-dependent spatial drift in both memory and ratings tasks are consistent with our predictions, but might these results be an artifact of how we structured the tasks? For instance, did we create an artificial link between spatial memory and spatial language by randomly intermixing recall and ratings trials? Perhaps in the face of this response uncertainty, participants prepared two responses in the delay conditions. This might have caused the two prepared responses to interact during the delay, leading to shared bias and shared response variability in the two tasks. Recent data suggest

that this is not the case. We conducted a second version of the experiment reported here with recall and ratings trials split across two sessions (Lipinski, Spencer, & Samuelson 2009a). The key result comes from the condition where participants did the ratings task in session 1 and the recall task in session 2. Critically, these participants had no knowledge of the recall task during their first session. Results replicated the findings reported in this chapter.

A related concern is whether we created an artificial link between spatial memory and spatial language by preventing participants from making a rating when both the target and reference object were visible. Recall that this was not the case in Hayward & Tarr (1995) and Crawford et al. (2000): in these studies, ratings could be prepared when the target was visible. In our task, therefore, people had to make a rating using their memory of the target location, in some sense forcing participants to link spatial memory and language. We certainly agree that the nature of our task requires that people use their memory of the target location in the ratings task. Importantly, however, the model we sketched in Figure 6.3 accounts for performance both with and without an imposed memory delay. More specifically, this model would generate a systematic shift in ratings of 'above' as visible targets were moved away from the vertical axis, and it would generate accurate pointing movements to visible target locations. Thus, even if we did create an artificial link between memory and language in our experiment, the model we proposed is still useful because it suggests how performance in multiple task contexts can be seamlessly woven together within a single framework. Moreover, we claim that, although our ratings task is certainly artificial, the processes at work in our 'delay' tasks are not. In particular, there are many naturalistic situations where we need to use our memory of objects' locations to generate spatial descriptions. Indeed, it is possible that spatial prepositions are used *more frequently* in cases where the objects in question are not visible. When two people are staring at the same visible objects, verbal communication is simple: 'hand me that' along with a pointing gesture will suffice. By contrast, when objects are not visible, 'hand me that' no longer works. In these situations, spatial prepositions are critical to effective communication.

6.7.2 *The DFT and spatial language: what still needs to be accomplished?*

Although the dynamic field model we sketched in this chapter provides a solid first step in a process-based direction, it is clearly overly simplistic. Nevertheless, the structure of the model provides useful constraints as we look to the future. In particular, we see five challenges that must be addressed within this theoretical framework. First, we must specify the process that aligns labels with particular reference locations in SWM. In Figure 6.3, we 'manually' aligned the 'above' node with location 0 in SWM. The challenge is that adults can do this

quite flexibly. Consider, for instance, what adults had to do in our task—they made ‘above’ ratings for targets presented in the horizontal plane. Although such judgements are not typical, participants had little difficulty adjusting to the task, and our results replicated the ratings gradient from studies that used a vertically oriented computer screen (Crawford et al. 2000; Hayward & Tarr 1995; Logan & Sadler 1996). The question is: what process accomplishes this flexible alignment? In our current model, we have an alignment process that matches perceived and remembered reference frames via a type of spatial correlation (Spencer et al. 2007). It is an open question, however, whether a related type of alignment process could work for the case of labels (for a robotic demonstration of this possibility, see Lipinski, Sandamirskaya, & Schöner 2009).

Next, we need to specify the process that structures the projection from SWM to the ‘above’ node. Conceptually, this gradient reflects the statistics of ‘above’ usage over development, but we need to specify the process that accumulates this statistical information. In past work, we have used activation in long-term memory fields to accumulate a type of statistical information across trials (Schutte & Spencer 2007; Simmering, Schutte, & Spencer 2008; Spencer et al. 2007; Thelen et al. 2001). Such long-term memory fields implement a form of Hebbian learning. A related issue is how to accumulate information across contexts. For instance, when young children are first learning the semantics of ‘above’, what process integrates use of this term across the diversity of situations in which this term is used? Put differently, what process accounts for generalization across contexts?

A third central component of our dynamic field model that needs further development is the nature of the bi-directional coupling between SWM and the ‘above’ node. Conceptually, coupling means that the establishment of stable patterns of activation within one layer should contribute to stable patterns in the other. Similarly, instability and drift within one layer should contribute to instability and drift within the other layer. The data presented in this chapter are consistent with the proposed link from SWM to the ‘above’ node, but what about coupling in the other direction? Recent experiments have confirmed that activation of a spatial term can stabilize spatial memory in some cases and amplify drift in others (Lipinski, Spencer, & Samuelson 2009b). Importantly, these results shed light on how the activation of labels projects back onto SWM and the situations in which this occurs.

The fourth challenge presented by our model is to expand beyond ‘above’ to handle multiple spatial prepositions. This requires that the processes we develop to handle the challenges above should generalize to other spatial labels. In this sense, we need to develop a formal, general theory of the link between space and words (see Lipinski, Sandamirskaya, & Schöner 2009). Furthermore,

we need to expand the model to handle the labeling of locations with multiple spatial terms such as ‘above and to the right’ (see Franklin & Henkel 1995; Hayward & Tarr 1995). Such effects can be handled by neural connections among the nodes representing different labels; however, we must specify the process that structures these connections. In this context, it is useful to note that our treatment of spatial terms via the activation of individual label nodes is consistent with several recent models of categorization and category learning that treat labels as a single feature of objects (e.g. Love, Medin, & Gureckis 2004).

Consideration of multiple spatial prepositions leads to the final issue our approach must handle: the model must ultimately speak to issues central to language use, such as how the real-time processes of spatial memory and spatial language relate to symbolic capacities for syntax and type-token distinctions. These broader issues obviously present formidable challenges, but we contend that there is no easy way around such challenges if the goal is to provide a constrained, testable theory of the connection between linguistic and non-linguistic systems. Given the neurally inspired view proposed by Barsalou (1999), an intriguing possibility is that the dynamic field approach could offer a formal theoretical framework within which one could specify the details of a perceptual symbol system.

6.7.3 *Ties between our process-based approach and other models*

When discussing our dynamic field model, it is of course critical to consider alternative models that are moving in related process-based directions. Two models are relevant here. The first is Regier & Carlson’s (2001) AVS model. This model incorporates the role of attention in the apprehension of spatial relations (Logan 1994; 1995) as well as the role of the geometric structure of the reference object (Regier & Carlson 2001). As mentioned previously, there is conceptual overlap between our dynamic field approach and AVS, in that both models scale ratings for prepositions like ‘above’ by the deviation between a reference axis and the target object. The manner by which these two models arrive at this deviation measure differs, however. In our model, this deviation is reflected in activation differences of the ‘above’ node that are structured by the projection gradient from SWM to this node. In AVS, by contrast, this deviation reflects the difference between a vertical axis and an attentionally weighted vector sum. A critical question for the future is whether these differences lead to divergent predictions. It is also important to note that AVS says nothing about performance in spatial recall tasks. As such, this model is not well positioned to examine links between spatial language and spatial memory.

A second related model is O’Keefe’s (2003) Vector Grammar. This model is similar to AVS in that location vectors provide the link between the perceived structure

of the environment and the use of spatial prepositions. In contrast to AVS, however, these vectors are derived from a model of place cell receptive field activations (see Hartley, Burgess, Lever, Cacucci, & O'Keefe 2000). The Vector Grammar approach shares conceptual overlap with the model we sketched in this chapter. In particular, both the place cell model by Hartley et al. (2000) and our dynamic field approach (Amari 1977 1989; Amari & Arbib 1977; Bastian, Riehle, Erlhagen, & Schöner 1998; Bastian, Schöner, & Riehle 2003; Erlhagen, Bastian, Jancke, Riehle, & Schöner 1999) are grounded in neurophysiology. Moreover, there is a strong spatial memory component to O'Keefe's Vector Grammar approach in that it explicitly attempts to link 'Cognitive Maps' (Tolman 1948) with linguistic 'Narrative Maps' (O'Keefe 2003). Beyond these areas of overlap, it is not yet clear the extent to which linguistic and non-linguistic spatial representational states are truly coupled or simply analogous in the Vector Grammar model. It will be important to evaluate this linguistic/non-linguistic link in the future as both modeling frameworks are expanded.

6.7.4 *Summary: toward a more process-based future*

We end this chapter by reiterating three central themes. First, we contend that the linguistic/non-linguistic connection must remain a central focus in cognitive science. Although tackling this issue presents formidable challenges, we think that the time is ripe to revisit it afresh given recent advances in both empirical techniques—for instance, the eye-tracking methods pioneered by Tanenhaus and colleagues (Tanenhaus et al. 1995)—and formal theoretical approaches—for instance, the dynamic field framework presented here (e.g. Spencer et al. 2007). Second, although focusing on representations in the abstract appears to be a useful simplification of the linguistic/non-linguistic link, this approach is not a panacea. Instead, we contend that such efforts can lead to under-constrained theories, a point we illustrated using an example from the spatial preposition literature. Third, we think close ties between theory and experiment can move the spatial language literature toward a process-based and theoretically constrained future. A growing number of empirical studies have explored the real-time linkages between linguistic and non-linguistic systems (e.g. Richardson et al. 2003; Spivey-Knowlton et al. 1998; Tanenhaus et al. 1995). This exciting work provides an excellent foundation for the development of the formal, process-based approach we have sketched here. Clearly there is a long way to go in this regard, but efforts that link formal theory and empirical work in this domain are critical if we are to address one of the most vexing issues in cognitive science today—the connection between the sensorimotor and the linguistic.

Author Notes

We wish thank Gregor Schöner for helpful discussions during the development of the model sketched here. This research was supported by NIMH (RO1 MH62480) and NSF (BCS 00-91757) awarded to John P. Spencer. Additional funding was provided by the Obermann Center for Advanced Studies at the University of Iowa.