

# Saliency Based Hierarchical Spatial Representations

Hongbin Wang (Hongbin.Wang@uth.tmc.edu)

Yanlong Sun (Yanlong.Sun@uth.tmc.edu)

School of Health Information Sciences, University of Texas at Houston  
7000 Fannin Suite 600, Houston, TX 77030

## Abstract

How people represent space is a central issue in spatial cognition research and has profound implications on human performance in spatial tasks. In this article, we describe a theory of human spatial representations, which claims that space is represented in the brain and in the mind not once but multiple times, each being a map of saliency with a distinctive frame of reference, and that human performance is determined by the interaction among multiple such representations. An experiment is reported to test the theory's claim on saliency based hierarchical spatial representations.

## Introduction

People live in a 3D world and perform various spatial tasks in every waking minute of their lives. Searching for an object in a visual scene, walking to your office from the parking lot, navigating in a new city, and surfing Google Earth, to name a few, all involve spatial information processing. Some tasks are easy while others are quite challenging. How and why is this so?

Decades of interdisciplinary research on human spatial cognition has resulted in a range of theories, with each focusing on distinctive aspects (Freksa, Habel, & Wender, 1998; Golledge, 1999; Newcombe, 2002; O'Keefe & Nadel, 1978; Tolman, 1948). However, the role of spatial representations on human spatial performance has been almost universally emphasized (Klatzky, 1998; H. Wang, Johnson, Sun, & Zhang, 2005; R. F. Wang & Spelke, 2002). While it is generally accepted that any physical space has to be somehow digested and encoded in a psychological space in order to be cognitively useful, a large body of evidence has convincingly shown that a psychological space is not an exact copy of the corresponding physical space. On the one hand, a physical space appears perfectly three-dimensional, absolute, unified, continuous, and Euclidean. On the other hand, it is well documented that the psychological space is often segmented, relative, partial, distorted, and non-Euclidean (e.g., Halligan, Fink, Marshall, & Vallar, 2003; Hunt & Waller, 1999; McDonald & Pellegrino, 1993; Newcombe & Huttenlocher, 2000; Tversky, 2000). Because of this, if a spatial task requires a certain piece of information and if that piece of information is represented in a form that is not suitable for the task or not represented at all, then the task becomes more difficult. In contrast, if a task requires a piece of information that is properly represented and readily available, then the task often becomes easier.

Unfortunately, different opinions exist about how a psychological space is encoded. In this article, we briefly outline an integrated theory of human spatial representations that is built on solid neuropsychological evidence and is able to accommodate several seemingly contradictory previous findings. The theory essentially states that a psychological space consists of multiple representations, each with a distinctive frame of reference and each being a partial selection (anchor-based) of all possible spatial information. The theory lays out a set of principles about what information is readily encoded in each representation and how different representations interact to solve spatial problems, which often lead to tested predictions in spatial performance.

This article consists of three parts. In the first part, we will briefly review several key concepts and claims of the theory, including frame of reference and saliency. Several major theoretical debates in the field are highlighted. In the second part, we report an experiment that was designed to test theory's claim on saliency based hierarchical spatial representations. In the final part, the implications of the theory are discussed in the broad context of modeling human spatial cognition.

## A Theory of Spatial Representations

Any theory describing the representation of space has to solve at least two critical problems – how to represent and what to represent. The how problem has to do with the concept of frame of reference (FOR). In a certain sense, to represent a space is to encode the space in a frame of reference (FOR). Though different taxonomies exist (e.g., Brewer & Pears, 1993; Garnham, 1989; Klatzky, 1998; Levinson, 1996; Logan & Sadler, 1996; Palmer, 2002; Talmy, 1983), three classes of frame of reference can be generally distinguished – egocentric (where the FOR is centered on the observer or her body parts), allocentric (where the FOR is centered on an object external to the observer), and intrinsic. An intrinsic FOR is a blend of an egocentric frame and an allocentric frame in that in an intrinsic FOR, while spatial information is centered on a specific object other than the observer (similar to an allocentric frame of reference), it is coded in egocentric terms as if the observer is located at the same place and with the same heading as the object at the origin. An example is “John is sitting on Mary’s right”.

Although different reference systems are theoretically equivalent in locating objects in a space, psychologically people acquire and utilize frames of reference quite differently. In which frame of reference psychological space

in general is represented is the holy grail of the field. Following the seminal work of Tolman (1948) and O'Keefe & Nadel (1978), it seems to some researchers that people represent space in terms of allocentric based cognitive maps. However, others have argued that people maintain egocentric based spatial representations (e.g., Tarr, Williams, Hayward, & Gauthier, 1998; R. F. Wang & Spelke, 2000, 2002). This latter view has gained much support from neuropsychological studies, especially in hemispatial neglect patients. These patients typically lose the perception of some part of space, and this part of space is often defined in an egocentric frame of reference (e.g., one patient may not be able to see things on their left side of the body; see (Halligan, Fink, Marshall, & Vallar, 2003; Vallar, 1998)). More recently, the importance of spatial representations in intrinsic reference systems in human spatial behavior has been emphasized (e.g., McNamara, 2003; Mou & McNamara, 2002; H. Wang, Sun, Johnson, & Yuan, 2005).

While the debate continues (R. F. Wang & Spelke, 2002), it is becoming increasingly clear that people often maintain and use different types of frames of reference simultaneously. One set of evidence comes from studies on child development and human evolution. The key idea is that while the egocentric representations are acquired before the allocentric representations in both phylogeny and ontogeny, they tend to co-exist in normal human cognition. For example, Piaget and Inhelder (1956) suggest that there are three stages in children's development of frames of reference, from egocentric, to fixed relative to external objects, and to absolute coordinates. Another set of evidence comes from cognitive neuroscience research on how the brain represents space. At least two brain systems, the parietal cortex and the hippocampal system, are found to be especially critical for spatial representations, with the parietal cortex representing space in multiple egocentric FORs, and the hippocampal system, via place cells (in rats) or spatial view cells (in primates), for the construction of allocentric representations of space. Therefore, it seems that the issue at stake is not which one frame of reference humans adopt, but how different available representations work together to produce spatial performance (e.g., Burgess, Jeffery, & O'Keefe, 1999; Colby & Goldberg, 1999; McNamara, 2003; H. Wang, Johnson, & Zhang, 2001).

The what problem of spatial representations has to do with spatial selection. A natural space is typically crowded by objects and events. The large amount of spatial information presented by these entities, including their locations and spatial relationships, either egocentric or allocentric, either static or dynamic, seriously challenges the limited human cognitive capacity. Apparently, not all possible spatial information in the physical environment can be explicitly represented in the psychological space and a selection is necessary.

A large body of evidence has shown that salience, a quality of "standing-out" that is determined by both the behavioral significance and bottom-up perceptual distinctiveness of relevant information, plays an important role in spatial selection (Colby & Goldberg, 1999; Itti & Koch, 2001; Li, 2002). That is, while all possible spatial

information is available physically only a subset of salient spatial information is selected by spatial cognition for further processing and consequently explicitly represented in psychological space. In a certain sense, a psychological space is a partial selection of the corresponding physical space, and this selection process is guided by the salience of relevant stimuli.

For example, Gottlieb and colleagues (1998) investigated the firing patterns of neurons in monkey's lateral intraparietal area. They found that neurons show strong responses to stimuli appearing abruptly in their receptive fields. However, these neurons show little or no response to stimuli brought into their receptive field by saccades, unless the stimuli were made behaviorally significant. The authors suggest that the entire visual space is only weakly represented, with the most salient or behaviorally relevant objects being strongly represented. Similarly, Li (2002) proposed that there is a salience map in primary visual cortex such that "firing rates of V1's output neurons increase monotonically with the salience [scalar] values of the visual input..." (p. 9).

One interesting notion about salient spatial information is that a spatial location can become salient, or behaviorally significant, without being occupied by an object. A location where an interesting object is going to appear or has appeared before can be salient and needs to be represented. Corbetta et al (2000), using normal subjects and event-related fMRI, show that the intraparietal area becomes active as soon as a location is made relevant (salient) and attended before the presentation of stimuli. Similarly, Duhamel et al (1992) asked their patient, who suffered right frontoparietal damage, to visually track a sequence of quickly appearing-then-disappearing targets. The patient's systematic failure in this simple task indicated that she could not maintain an updated (salience) map induced by previously-presented stimuli.

These results support the notion that that a psychological space, instead of a genuine copy of the corresponding physical space, is a map of salience. In such a map, only a few salient locations/relations, defined by a combination of perceptual distinction and behavioral significance, are strongly represented, and all other locations/relations are only weakly represented or "do not exist" (i.e., not represented at all). While this claim readily explains why psychological space is a distorted and segmented representation of physical space, it is important to note that salience is an empirically defined subjective variable and that there may not be a universally accepted formula to define salience. This is especially the case for the goal-related aspect of salience, where the determination of salience has to be task specific and context specific.

## Experiment

In this section, we report an experiment we have conducted to examine the salience aspect of spatial representations. Our investigation is particularly related to a long-standing suggestion that spatial knowledge is organized hierarchically. Huttenlocher, Hedges, and Duncan (1991) showed the participants on a piece of paper a circle with a dot in it. They then asked the participants to reproduce the

dot position based on memory. They found that the reproduced positions systematically deviate from their original positions. Specifically, if the circle is divided into horizontal-vertical and radial slices, the reproduced dots were often displaced toward the center of the slices in which they fall. This result suggests that a psychological space might be hierarchically represented and there is a central tendency at each hierarchical level (e.g., the slice) which all other spatial information is anchored upon. As a result of this hierarchical organization, the anchors become salient, and the distance between an anchor and the locations around the anchor is distorted and becomes smaller (see also Hirtle & Jonides, 1985). As yet another example of the hierarchical organization of spatial knowledge, consider the encoding of the United States, Texas, Houston, and Houston Medical Center. Psychologically each space is a representation of map with a central tendency towards its anchors (e.g., the capital, the downtown, or other significant landmarks). Together they form a hierarchy. As we go deeper in the hierarchy, the anchoring effect at the higher levels diminishes and new anchors (and more details) at the lower levels become available. In other words, it appears some spatial information does not exist (or is not represented explicitly) at the higher levels of hierarchy due to the central tendency (or anchoring) effect.

Our theory predicts that people organize spatial knowledge in various hierarchical structures, each of which is anchored around some salient objects and encoded in a distinctive frame of reference. One implication of this

prediction is that if we present to subjects an environment, in which some objects are made salient, either through top-down instructions (e.g., emphasizing their importance) or bottom-up distinction (e.g., perceptually standing out from other objects), and ask subjects to remember layout of the environment, then we would expect the encoding of the environment would be anchored around these salient objects. As a result, if we measure the time it takes to retrieve the spatial relations from the memory, the reaction times might show a symmetrical pattern surrounding the anchoring object (hence referred to as “landmark”). Thus, reaction time is not always proportional to the corresponding physical distance. Rather, it is determined by the psychological distance in the hierarchical spatial representation.

**Design and Procedure.** To test this hypothesis, we presented to subjects an array of objects on a computer screen. In one condition, one object in the array was perceptually distinct from all others (Figure 1A), and in another condition, a non-distinctive object was used instead (Figure 1B). Subjects were asked to remember the spatial layout of these objects (self-paced). After the encoding, we tested subjects' memory by presenting pairs of objects on the screen and asking subjects to decide if the pair was in its originally studied spatial relations. The reaction time data were recorded.

**Subjects.** 14 graduate students in the Texas Medical Center area were paid to participate in the experiment.

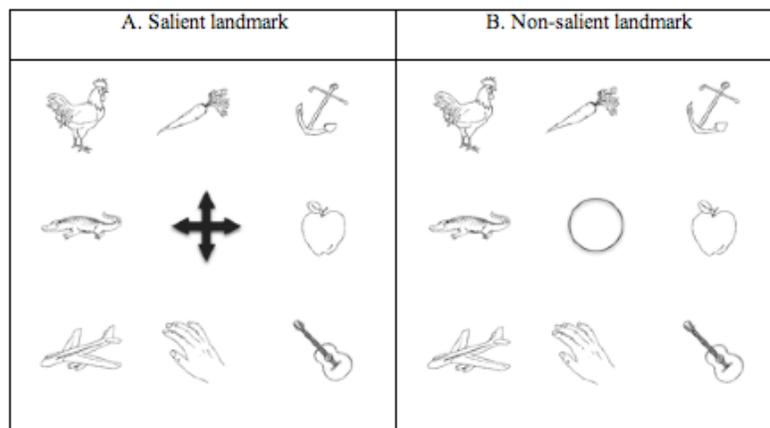


Figure 1. The object layout in the experiment. One key difference lies in the center object in each layout. The center object (a cross) in layout A is in blue color, which is different from other objects (including the center object in layout B, a circle), which are all in black-and-white. All objects (except the center objects) are selected, with controlled familiarity and frequency, from the database developed by Snodgrass and Vanderwart (1980) and randomly assigned to each location.

**Results.** Among the 28 total number of possible object pairs in each array, we classified them into 7 groups based on two variables (see Table 1). One was whether the pair involved any landmark, which could be a landmark-object relation (e.g., the chicken and the blue landmark in Figure 1A) or a landmark-linked relation (e.g., the chicken and the violin in Figure 1A). The other one was the distance, which could be 1, 1.414, 2, and 2.818 (arbitrary unit).

The average accuracy was 95.6%, with a standard deviation of 5.27%. The subjects' reaction time data are shown in Figure 2, separated based on the salience conditions and the group number. The overall statistical analyses showed that both the salience effect (mean difference = 150.3 ms,  $F(1, 13) = 14.45$ ,  $p < .01$ , estimated effect size = .526) and the group effect ( $F(6, 78) = 17.66$ ,  $p < .01$ , estimated effect size = .576) were significant, as well as their interaction ( $F(6, 78) = 3.20$ ,  $p < .01$ ). Further analyses showed that while a subset of pair groups (1, 3, 5, and 6) showed the salience effect (with a minimal mean difference of 142.4 ms and standard error of 53.7 ms), the other groups (2, 4, and 7) did not. This was in general consistent with our predictions in that those pairs in those significant groups (except for group 6) all involved the landmark, suggesting that the spatial representations might be organized around the salient landmarks in a hierarchical fashion. This result was especially significant given that we did not emphasize to the subjects the importance of the central landmark in the encoding and subjects were free to choose their own encoding strategy.

Table 1. The classification of object pairs

Group	# of pairs	Landmark-Linked	Landmark-Object	Distance	Example (see Fig1A)
1	4	no	yes	1	landmark-carrot
2	8	no	no	1	chicken-carrot
3	4	no	yes	1.414	landmark-chicken
4	4	no	no	1.414	crocodile-carrot
5	2	yes	no	2	carrot-hand
6	4	no	no	2	chicken-anchor
7	2	yes	no	2.818	chicken-violin

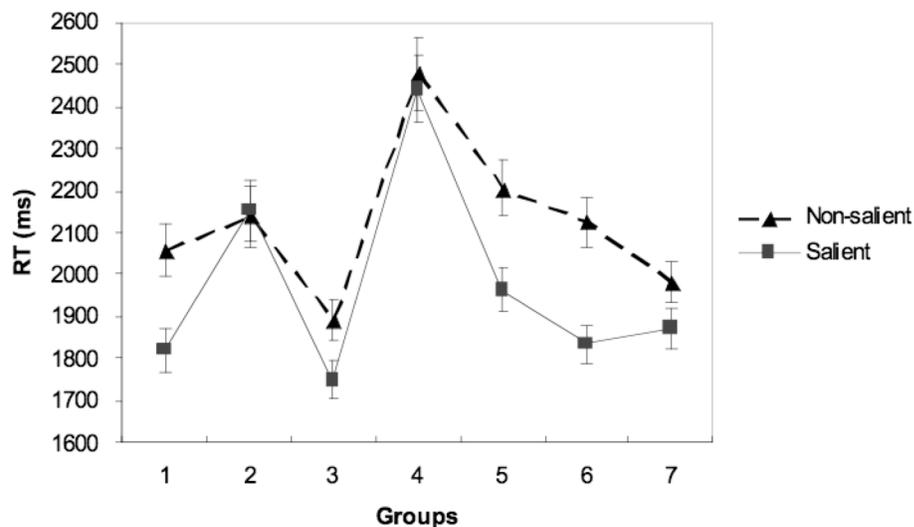


Figure 2. The reaction time data based on salience and group variables. The error bars are standard errors.

## Discussions

People live in a 3D world containing objects that are either stationary or moving. Being able to maintain oriented in the space around us, including knowing the location of

important objects and their relations, is crucial for our everyday life. Decades of studies in the broad area of psychology and neuroscience have produced large amount of data and theoretical hypotheses. Unfortunately, some fundamental issues related to spatial representations remain

controversial. In this article, we present a theory of human spatial representations. It claims that space is represented in the brain and in the mind not once but multiple times, each being a map of salience with a distinctive frame of reference. Multiple such representations exist simultaneously and interactively determine spatial performance. Theoretical analyses suggest that the theory is consistent with various previous proposals that only emphasize a single type of representations and is able to make novel and testable predictions.

The theory has been empirically evaluated in a series of psychological experiments FORMS (H. Wang, Johnson, & Bao, 2005; H. Wang, Johnson, Sun, & Zhang, 2005; H. Wang, Sun, Johnson, & Yuan, 2005). These experiments have provided a systematic examination of the spatial representations and operations underlying typical spatial tasks, including the relationship between visual features and spatial locations, how multiple maps of salience interact, and how spatial updating works with multiple representations. The experiment reported here represents yet another test on the theory. It supports the idea that spatial representations are maps of salience, which typically consist of only a set of salient spatial information and is often organized around salient objects in hierarchical structures. However, we recognize that to fully test the theory, studies involving more complex spatial environments that go beyond the simple laboratory setting are necessary.

We have attempted to extend the finding of the current study through more experiments. For example, in one experiment we presented multiple salient landmarks in a more complex environment (e.g., align two Figure 1A horizontally in a single environment, with the two landmarks are in different colors: blue and red). In this setup, a hierarchical organization was more clearly present in the sense that some landmarks' salience effect became apparent only when we went deep into the lower levels of the hierarchy. We have found similar results, supporting the idea that salience-based spatial organization in spatial representations.

Yet another way to evaluate the theory is through computational modeling. The theory has been implemented in the ACT-R cognitive architecture (Anderson et al., 2004; Anderson & Lebiere, 1998) as an extension module (H. Wang, 2005). By doing so, we not only show that the theory is computationally feasible but also are able to explore its capacity and limitations in a more rigid way. Similar to other ACT-R modules, the spatial module basically acts as an interface between ACT-R and the external environment that specializes in spatial information processing. Multiple maps of salience are implemented through multiple ACT-R buffer structures. These buffers allow the spatial module to communicate with ACT-R and its other modules, such as vision and motor. In addition, because these buffers provide different representations for the same environment, it is possible that some information is readily available in one representation but not available at all in another, thus leading to testable predictions.

## Acknowledgments

This work is supported by grants from the Office of Naval Research Cognitive Science Program (Grant Nos. N00014-01-1-0074 & N00014-04-1-0132). We thank Drs. Todd R. Johnson and Jiajie Zhang for their assistance in this work.

## References

- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036-1060.
- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Press.
- Brewer, B., & Pears, J. (1993). Frames of reference. In R. Eilan, R. McCarthy & B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology* (pp. 25-30). Oxford: Blackwell.
- Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, *22*, 319-349.
- Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, *3*(3), 292-297.
- Duhamel, D. R., Goldberg, M. E., FitzGibbon, E. J., Sirigu, A., & Grafman, J. (1992). Saccadic dysmetria in a patient with a right frontoparietal lesion: The important of corollary discharge for accurate spatial behavior. *Brain*, *115*, 1387-1402.
- Freksa, C., Habel, C., & Wender, K. F. (Eds.). (1998). *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge*. New York: Springer-Verlag.
- Garnham, A. (1989). A unified theory of the meaning of some spatial relational terms. *Cognition*, *31*(1), 45-60.
- Golledge, R. G. (Ed.). (1999). *Wayfinding behavior: Cognitive mapping and other spatial processes*. Baltimore: The Johns Hopkins University Press.
- Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*, 481-484.
- Halligan, P. W., Fink, G. R., Marshall, J. C., & Vallar, G. (2003). Spatial cognition: evidence from visual neglect. *Trends in Cognitive Sciences*, *7*(3), 125-133.
- Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory & Cognition*, *13*, 208-217.
- Hunt, E., & Waller, D. (1999). *Orientation and wayfinding: A review*. Technical Report to ONR.
- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*, 352-376.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews*, *2*, 194-203.
- Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel & K. F. Wender (Eds.), *Spatial cognition: An interdisciplinary approach*

- to representing and processing spatial knowledge. New York: Springer-Verlag.
- Levinson, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel & M. Garrett (Eds.), *Language and space*. Cambridge, MA: MIT Press.
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6, 9-16.
- Logan, G. D., & Sadler, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, A. Peterson, L. Nadel & M. Garrett (Eds.), *Language and space*. Cambridge, MA: MIT Press.
- McDonald, T. P., & Pellegrino, J. W. (1993). Psychological perspectives on spatial cognition. In T. Garling & R. G. Golledge (Eds.), *Behavior and environment: Psychological and geographic perspectives*. North-Holland: Elsevier.
- McNamara, T. P. (2003). How are the locations of objects in the environment represented in memory? In C. Freksa, W. Brauer, C. Habel & K. Wender (Eds.), *Spatial cognition III: Routes and navigation, human memory and learning, spatial representation and spatial reasoning* (pp. 174-191). Berlin: Springer-Verlag.
- Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 162-170.
- Newcombe, N. S. (2002). Spatial cognition. In *Steven's Handbook of Experimental Psychology, Vol 2: Memory and Cognitive Processes* (3rd ed., pp. 113-163). New York: John Wiley & Sons.
- Newcombe, N. S., & Huttenlocher, J. (2000). *Making space: The development of spatial representation and reasoning*. Cambridge, MA: MIT Press.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. New York: Oxford University Press.
- Palmer, B. (2002). Linguistic frame of reference reconsidered. In *Proceedings of the 2002 Conference of the Australian Linguistics Society*.
- Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge & Kegan Paul.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, 6, 174-215.
- Talmy, L. (1983). How languages structure space. In H. Pick & L. Acredolo (Eds.), *Spatial orientation: theory, research and application*. New York: Plenum Press.
- Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, 1, 275-277.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55, 189-208.
- Tversky, B. (2000). Remembering spaces. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory*. New York: Oxford University Press.
- Vallar, G. (1998). Spatial hemineglect in humans. *Trends in Cognitive Sciences*, 2, 87-97.
- Wang, H. (2005). *Modeling spatial cognition*: Technical Report to ONR.
- Wang, H., Johnson, T. R., & Bao, R. (2005). What vs where: Which direction is faster? In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the Twenty-seventh Annual Meeting of the Cognitive Science Society* (pp. 2337-2341). Mahwah, NJ: Lawrence Erlbaum.
- Wang, H., Johnson, T. R., Sun, Y., & Zhang, J. (2005). Object-location memory: The interplay of multiple representations. *Memory & Cognition*, 33(7), 1147-1159.
- Wang, H., Sun, Y., Johnson, T. R., & Yuan, Y. (2005). Prioritized spatial updating in intrinsic frame of reference. *Spatial Cognition and Computation*, 5(1), 89-113.
- Wang, R. F., & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition*, 77, 215-250.
- Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in Cognitive Sciences*, 6(9), 376-382.