Spatial Memory of Immediate Environments

Holger Schultheis (schulth@informatik.uni-bremen.de)
University of Bremen
Bremen Spatial Cognition Center
Institute for Artificial Intelligence
Am Fallturm 1, 28359 Bremen, Germany

Abstract

Memorizing and retrieving information about the spatial layout of one's surrounding is of crucial importance for humans. We propose a new theory of spatial memory of immediate environments and develop a corresponding computational realization. We detail how the theory explains key findings on human spatial memory (use) and show that the computational realization accounts well for human behavior from three pertinent experiments. One implication of the theory's success is that enduring spatial memory representations may best be conceptualized as flexible combinations of representation structures and reference frames.

Keywords: spatial memory; spatial reference frames; interference; perspective taking; computational modeling

Introduction

Knowledge about the location of and relation between objects in the immediate environment is crucial for everyday life. Such spatial knowledge can be obtained by perception, but many everyday activities crucially rely on memory representations of spatial information. For example, spatial memory enables avoiding collisions with objects currently outside the perceptual field (e.g., a chair slightly behind oneself when moving around the table), anticipating and planning movements from positions one has not yet reached (e.g., planning the movement to place a plate once one is next to the appropriate spot on the table), and navigating towards objects which are not directly perceivable (e.g., approaching the appropriate cupboard to retrieve the plates contained in it).

In line with its importance, spatial memory of immediate environments has received considerable attention by research in the cognitive sciences and a substantial number of theories of spatial memory has been proposed (Avraamides & Kelly, 2008; Byrne, Becker, & Burgess, 2007; Mou, McNamara, Valiquette, & Rump, 2004; Sholl, 2001; Wang, 2017). A prominent approach to investigating spatial memory of immediate environments have been *perspective taking* (PT) studies, in which people have to judge spatial relations of a previously learned object layout from imaginal perspectives. From these studies a number of main findings have emerged (May, 2007), which can be assumed to characterize key aspects of the structures and processes involved in spatial memory.

We propose a new theory of spatial memory that offers explanations for all main findings. We first describe the PT paradigm and the main findings arising from it. Subsequently, we expound our theory, how it explains the findings, and a computational realization of the theory. After briefly considering related theories, we close with a discussion of the implications of our work.

Main Findings

In typical PT studies on spatial memory of immediate environments people are first asked to memorize the location of objects in their surrounding. After learning the object layout, people are deprived of perceptual access to their surrounding (e.g., by blindfolding) and tested for their knowledge of the spatial relations between objects. Two common forms of testing spatial relations are *judgment of relative direction* and *egocentric pointing*. In a judgment of relative direction task, people are asked to point to obj_1 as if they were standing at obj_2 facing obj_3 (where obj_i are three objects of the previously learned object layout). In an egocentric pointing task, people are asked to point to obj_1 as if they were standing at or facing obj_2. The object to point to is called the *target object*. In particular, the to-be-imagined perspective (e.g., facing obj_2) is usually different from the actual bodily perspective of the participants. The imaginal perspective can differ from the bodily perspective by *rotation* (i.e., the locations of bodily and imaginal perspective coincide, but orientations of the perspectives differ), *translation* (i.e., the orientations of the perspectives coincide, but locations differ), or both (often the case in judgment of relative direction tasks).

By using such a PT approach, existing studies have uncovered many intriguing phenomena of spatial memory organization and access. In the following, we will focus on a set of phenomena, which can be considered the main findings of existing research (May, 2007):

- Taking an imaginal perspective different from the bodily perspective is hard. Indicating the direction to the target object from the imaginal perspective takes more time and is more error prone than from the bodily perspective.
- Imaginal perspectives involving rotations are harder (slower, more error prone) than imaginal perspectives involving only translations.
- The difficulty of pointing to the target object increases with increasing angular disparity between the pointing direction from the imaginal perspective and the pointing direction from the bodily perspective.

- The difficulty of responding from the imaginal perspective can be reduced, if observers are ignorant of the actual spatial relation of their body to the object layout (e.g., when being disoriented).
- Differences between the orientation of the imaginal perspective and salient orientations in the environment (e.g., orientation of axes of symmetry or orientation of learning perspective) may lead to extra processing costs.
- If people are allowed to move their body such that the bodily perspective coincides with the tested perspective, the above mentioned difficulties are reduced notably and sometimes even eliminated.

A further finding that we will consider is the influence of perspective preparation on PT performance. If people are given information about the tested perspective before they are informed about the target object, they may be able to prepare the to-be-taken perspective such that they can respond with less difficulty once the target object is presented. Several studies have investigated the influence of preparation, because preparation effects can help reveal how access to spatial memory is organized. Although preparation has been found to generally reduce processing times associated with PT (Brockmole & Wang, 2003; May, 2004), it seems hard to prepare for certain difficulties (e.g., increase of processing costs with increasing disparity May, 2004; Wang, 2005).

A Theory of Spatial Memory

As virtually all previous theories of spatial memory (Avraamides & Kelly, 2008; Byrne et al., 2007; Mou et al., 2004; Sholl, 2001; Wang, 2017), our theory assumes that one component of spatial memory is what we will call the *sensorimotor representation*. It represents self-to-object relations for (certain) objects in the immediate environment. If any movement of one's body is perceived (through vision, proprioception, etc.) these relations are updated accordingly. In this sense the sensorimotor representation is dynamic and transient. Access to this representation is quick and automatic and it serves as the default basis for motor actions.

In addition, our theory assumes a more enduring representation of the environment as a second component. We will call this component the *LTM representation*. It represents object-to-object relations between the objects in the immediate environment. One's own body can be one of the objects in the LTM representation. The LTM representation is orientation-free and not inextricably linked to some spatial *reference frame* (RF). However, the representation may be associated with a RF in the same sense as items in long-term memory are usually assumed to be associated with each other.

Because the LTM representation is orientation-free, it will be of limited use without further additions. Consider the two object layouts in Fig. 1. Both layouts yield identical representations in an orientation-free representation, but for acting on or within the layout (e.g., approaching object A) it makes a difference which situation is represented. To create the necessary correspondence between the LTM representation and the real world, that is, to anchor the representation in the real world, it has to be oriented. We argue that this is achieved by imposing a spatial RF onto the LTM representation and our theory assumes that any access to the LTM representation involves such an imposition of a RF. A common RF that people will likely employ to access the LTM representation is the *bodily* RF arising from the sensorimotor representation (i.e., a RF that is oriented as the actual body). Other RFs may be RFs associated with the LTM representation (e.g., RFs salient during encoding of the spatial layout) or an *imaginal* RF that allows assessing the spatial layout from a vantage point differing from the current bodily vantage point (e.g., when trying to identify the seats with the best view on the stage in a theater without first walking through the whole theater).

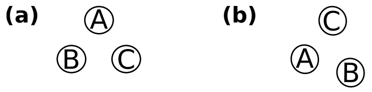

Figure 1: Two spatial layouts (a) and (b) that yield identical orientation-free object-to-object representations.

Notably, differing RFs may concurrently be available for accessing the LTM representation. Accordingly, we propose that accessing the LTM representation requires RF selection and depending on which frames are available this selection may be competitive and effortful. Specifically, our theory assumes that selection probability and effort depend on the conflict between the available frames, where conflict is a function of the salience and the (mis)alignment of the available frames (see further detail below).

According to our theory, taking an imaginal perspective involves the following steps (see also Fig. 2): First, a RF has to be selected and imposed onto the LTM representation. To perform the PT task successfully, the selected RF needs to be the one corresponding to the to-be-taken perspective or a different but aligned RF. Second, once the RF has been imposed, the LTM representation can be accessed to determine the direction towards the target object. Third, the determined target direction is used to activate a pointing movement towards the target. If the imaginal and bodily perspective differ from each other, the determined pointing direction is in conflict with the pointing direction to the target given by the sensorimotor representation. Because access to the sensorimotor representation is automatic, the disagreeing movement directions give rise to motor interference. The strength of this interference is assumed to depend on the dissimilarity of the two movements: The more the two movements' differ in direction, the stronger the interference. Note that in this process, activation of the motor response can only start after LTM access is completed. On the other hand, nothing precludes RF selection and LTM access to happen before the target direction is known. Accordingly, we propose that LTM access may start before the target object is known.

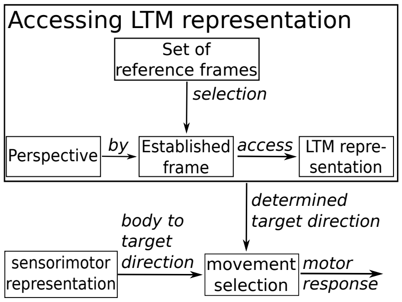

Figure 2: Processing steps in imaginal perspective taking.

Explanation of Main Findings

Our theory explains the main findings mentioned above as follows:

- Imaginal PT is difficult, because it requires RF selection to access the LTM representation. Selecting a frame when (at least) the bodily frame and the imaginal frame are in conflict takes more time than when the frames are aligned. Furthermore, an incorrect frame (e.g., the bodily frame) may be selected, which will result in an erroneous pointing response. Additional difficulty arises from motor interference such that interference leads to slower and more error prone response execution.
- Perspectives involving rotations create a misalignment between the orientation of the bodily and imaginal RF. Perspectives involving only translations do not lead to such a misalignment. Consequently, accessing the LTM representation is harder for rotations than for translations
- With increasing disparity, the difference between movement direction to the target from the bodily and the imaginal perspective increase. This leads to increased motor interference, which results in slower and more error prone responding.
- Lacking a sensorimotor representation has two effects: First, the sensorimotor representation does *not* give rise to a RF, which may otherwise have lead to conflict during LTM access. Second, motor interference is reduced or even eliminated. Accordingly, taking an imaginal perspective different from the bodily perspective can be easier without a sensorimotor representation.
- If a RF is associated with the LTM representation, it may be co-activated with the LTM representation. If this associated RF differs from the imaginal RF, it creates conflict during accessing the LTM representation. As a result, a disagreement between imaginal perspective and, for example, the learning perspective renders PT more difficult.
- Bodily movements towards the to-be-imagined perspective will lead to an accordingly updated sensorimotor representation. This means that the bodily and the imaginal RFs will be aligned and there will be little or no motor interference. Consequently, PT difficulty will be greatly reduced.
- Access to the LTM representation may proceed during preparation and thus reduce the overall processing time. However, motor interference arises only after the target object has been determined and, consequently, cannot be reduced by preparation. This explains why parts but not all of the processing costs of PT can be reduced by preparation.

Formalization

Our theory's ability to provide explanations for the main effects lends support for its assumptions. To allow comparing the behavior predicted by the theory to human behavior in more detail we formalized the theory as a computational model and applied the model to two pertinent PT studies. As a first step, we decided to use a formalization that captures the main assumptions of the theory while remaining as simple as possible. This has the advantage that any successes or failures of the model can be more directly attributed to the theory and its assumptions instead of being a result of implementation-specific detail (see, e.g., Cooper & Guest, 2014). An implementation of the theory that provides more detail on the possible mechanisms is discussed below.

Because establishing a RF and motor interference are the main factors in driving PT difficulty, the model focuses on these two aspects.

RF Selection. Establishing a RF is formalized as follows: Each of the available reference frames RF_i is assumed to have a salience sal_i such that the salience of all available reference frames sums to one. Following Botvinick, Braver, Barch, Carter, and Cohen (2001), we define the strength of conflict (cV) of RF_i with RF_j as

$$cV(i,j) = \delta * sal_i * sal_j * (1 - jConf),$$

where $jConf$ is the conflict of RF_j to all other RF (i.e., $RF_k, k \neq i$) and

$$\delta = \begin{cases} -1, & \text{for } RF_i \text{and } RF_j \text{aligned,} \\ 1, & \text{for } RF_i \text{and } RF_j \text{misaligned.} \end{cases}$$

The overall conflict of RF_i is given as the sum over all pairwise conflict values across all other RF:

$$cV_i = \sum_{j, j \neq i} cV(i,j).$$

The salience and the conflict of each frame are combined to yield an impact score imp_i. Specifically, the frame's salience is scaled based on its conflict value such that higher conflict leads to a lower impact score and $imp_i \in [0.5 * sal_i, 1.5 * sal_i]$. Probability of a frame being selected sp_i and the speed with which it can be selected st_i are proportional to imp_i:

$$sp_i = \frac{imp_i}{\sum_j imp_j},$$

$$st_i = A * (impMax - imp_i),$$

where $impMax$ is the maximum possible impact score and A serves to scale the response time to the order of magnitude of the human data.

Motor Interference. Processing time arising from motor interference is assumed to be directly proportional to the disparity $disp$ between the directions of the two interfering movements: $B * disp$, where B is a scaling factor analogous to A above. Error is determined distinguishing two cases: First, if the imaginal RF (or a frame aligned with it) is used to access the LTM representation, error is also assumed to be proportional to disparity: $C * disp$. If a frame misaligned with the imaginal frame is selected, the error will amount to the angular difference of the selected frame's and the imaginal frame's orientation.

Example. To illustrate the workings of the model, we will consider a situation, in which a person is located in the middle of a previously learned configuration of objects and asked to point to one of the objects as if facing one of the others. Let us assume that the imagined facing direction differs from the actual bodily facing direction and that the learning view coincides with the actual bodily orientation.

In such a situation, the model computes response time and error for each frame individually. The overall response time and error is given as a weighted average of all individual terms: each individual frame's time and error are weighed by the probability of selecting the frame and the resulting values are summed. To obtain the individual frame's values, the model computes the impact score of all three involved frames. Based on the impact scores, the selection probability and selection time of each frame is computed. Given any individual frame RF_i, the model computes the time from motor interference as a linear scaling of the disparity between pointing from the actual bodily perspective and the imaginal perspective given RF_i. If the selected frame is the imaginal frame, error is computed as a linear scaling analogously to the scaling of time. If any of the other two frames is selected, the error equals the orientation difference between the imaginal frame and the bodily/environmental frame. The ultimate output of the model are its prediction of response time and error in the given situation.

Given that humans are generally well able to perform PT tasks, we assumed that the imaginal frame has the strongest salience and set this salience to 0.6. The remaining salience amount of $1 - 0.6 = 0.4$ (saliences of all RF sum to 1) was distributed uniformly across all other RF. This left the scaling factors A, B, C as the only free parameters of the model.

Simulations

The first simulation addressed Experiments 2 and 3 of May (2004). These experiments provide a rich dataset of response times and pointing errors across 24 experimental conditions and also exhibit several of the main findings mentioned above.

Experiments 2 & 3 of May (2004). Participants had to perform an egocentric pointing task with to-be-imagined perspectives being either rotations or translations. Across different blocks, participants either had 1 s, 3 s, or 5 s to prepare their perspective before the target object was presented. Furthermore, the disparity between the target direction from the bodily perspective and the target direction from the imaginal perspective was systematically varied to yield levels of increasing disparity: 22.5°, 67.5°, 112.5°, 157.5°. Both experiments revealed that (a) rotations were slower and more error prone than translations, (b) response time and error increased with increasing disparity, and (c) that overall processing time but not the disparity effect decreased with increasing preparation time.

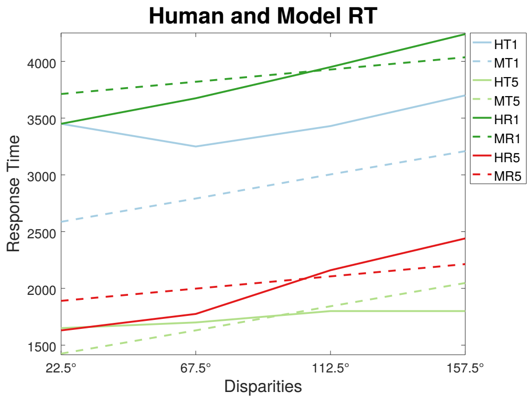

Figure 3: Human (solid lines) and model (dashed lines) response times for Experiment 2 of May (2004). Shown are translation and rotation times for two preparation durations (1 s and 5 s) each. H = Human; M = Model; T = Translation; R = Rotation; 1 and 5 indicate the duration of the preparation interval.

For simulating these experiments, we assumed that the sensorimotor representation, the bodily RF, and the imaginal RF are always present. Given the realization of the learning phase and the spatial layout of the experimental environment (see Fig. 1 in May, 2004), we also assumed the existence of an associated RF, which was—with equal probability—either aligned with the bodily RF or 45° misaligned with the bodily frame. We estimated the 3 free parameters of the model using the Metropolis algorithm (Madras, 2002) by fitting the model to response times and errors of both experiments across all conditions. Since the purpose of the simulation was to investigate the model's ability to account for key effects in the observed behavior, the objective of estimation was to maximize correlations between model and human behavior for response time and error for each of the two experiments (i.e., 4 correlations).

Model response times and errors correlated strongly with human times and errors for both experiments: $\rho = 0.91, \rho = 0.95, \rho = 0.94, \rho = 0.86$ for times and errors of Experiments 2 and 3, respectively. Model behavior is shown alongside human behavior for Experiment 2[1] in Figs. 3 and 4. As can

---

[1] The fit to Experiment 3 was very similar. For the sake of clarity the data from Experiment 3 are not included in the plot.

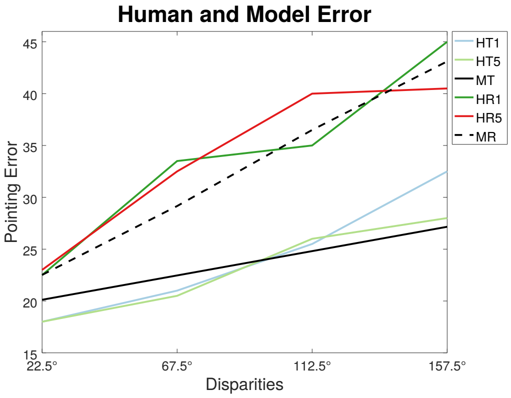

Figure 4: Human (colored lines) and model (black lines) pointing errors for Experiment 2 of May (2004). Shown are human errors for translation and rotation for two preparation durations (1 s and 5 s) each. Because model errors do not differ across different preparation durations, only rotations and translations are distinguished for model errors. H = Human; M = Model; T = Translation; R = Rotation; 1 and 5 indicate the duration of the preparation interval.

be seen from the plot, the model mirrors the main effects in the data well: (i) rotations are slower and more error prone than translations; (ii) response times and errors increase with disparity for both rotations and translations; (iii) preparation decreases the overall processing time, but does not substantially impact the disparity effect. Three further aspects of the simulation results seem noteworthy. First, humans show a stronger disparity effect for rotations than predicted by the model. Why humans should exhibit a stronger disparity effect for rotations than translations is currently unclear. Second, the model also captures that translations with short (1 s) preparation are slower than rotations with long (5 s) preparation. Third, the model correctly predicts that preparation time has no impact on error magnitude.

In sum, this first simulation lends further support to the assumptions of our theory in showing that a model based on the theory is able to closely mirror human behavior across a wide range of experimental conditions.

The second simulation will address a potential objection to our theory. Note that the theory makes no reference to mental transformations such as mental rotation or translation (e.g., Sholl, 2001). As a result the theory may seem to be at odds with findings indicating that PT time and error increases with the distance between the bodily and the imaginal perspective in translations (Easton & Sholl, 1995). Because no assumptions of our theory formulate an explicit relation between translation distance and PT performance, it seems an interesting question to what extent our theory can account for such a relationship. To address this question our second simulation models Experiment 1 of Easton and Sholl (1995).

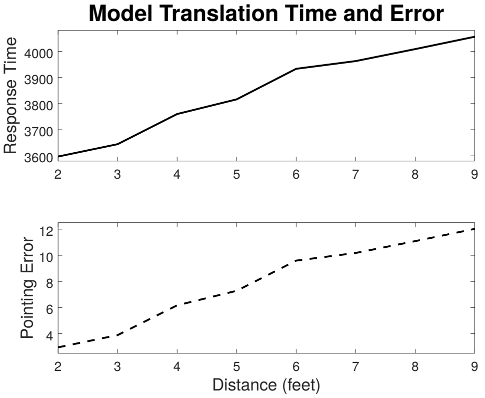

Figure 5: Model response times (solid line) and pointing errors (dashed line) when simulating Experiment 1 of Easton and Sholl (1995)

Experiment 1 of Easton and Sholl (1995). In this experiment, participants first memorized an object layout, in which 8 objects were placed with varying distance (2 – 9 ft) from the location of the observer (see Fig. 1B in Easton & Sholl, 1995). After learning the layout, participants were provided with a target object and then a to-be-taken perspective. They were asked to point to the target object as quickly as possible. The main finding was that for translations pointing time and error increased significantly with increasing distance.

We assumed the same RFs (bodily, imaginal, associated) as for the first simulation. Because the bodily orientation was identical for learning and testing, we assumed that all three RFs were aligned for translations. For each object that indicated the position of a translated perspective, we estimated the pointing disparity when pointing to each of the other objects from Fig. 1B of Easton and Sholl (1995). Response times and errors for one translation perspective were computed as the average times / errors across pointing to all other objects from this perspective. Because the purpose of this simulation was to assess the theory's general ability to account for increasing times and errors with increasing distance, we did not fit the model to the human data, but reused the parameters estimated in the first simulation.

As can be seen from Fig. 5, the model nicely accounts for the effects observed by Easton and Sholl (1995): Both times and errors increase with increasing translation distance. Given that our theory makes no reference to mental transformations or distances, it may not be immediately clear why the theory correctly predicts the observed human behavior. It turns out, however, that—at least in Experiment 1 of Easton and Sholl (1995)—the average pointing disparity systematically increases with increasing translation distance. Since our theory assumes increased PT effort with increasing disparity, it predicts the increased effort for increased distance observed in this experiment.

The Algorithmic Level

The model described above was designed to capture the gist of the theory with as little implementational overhead as possible. Consequently, the model remains somewhat abstract and does not provide much detail on the mechanisms and representations underlying the observed behavior. In this section, we propose a more mechanistic realization of our theory.

The object-to-object representation may be realized by the type of representation structure proposed by Schultheis, Bertel, and Barkowsky (2014). This circular representation structure preserves the neighborhood relations between directions, but requires a direction root (i.e., a RF) to ground it in the real world. We further suggest that RF selection may proceed as a leaky, competitive, accumulative process as in RF selection for spatial term use (Schultheis & Carlson, 2017). Finally, we think that activation of a pointing response and interference from the bodily pointing response can appropriately be captured by dynamic field theory (Schöner, Spencer, & DFT Research Group, 2015, with different pointing directions activating different parts of the dynamic field) .

Such a realization has the twofold advantage of promising to capture the main assumptions of the theory while, at the same time, constituting a computational instantiation that is more solidly grounded in previous cognitive theorizing than the above-described model. To what extent such a realization is able to mirror pertinent human behavior will be subject of our future research.

Related Theories

Several theories of spatial memory have previously been proposed and all of them have highlighted important properties of how humans represent and recall spatial information of immediate environments (e.g., Avraamides & Kelly, 2008; Byrne et al., 2007; Mou et al., 2004; Sholl, 2001; Waller & Hodgson, 2006; Wang, 2017). However, none of the existing theories provides an explanation of all of the main findings highlighted above. For some findings the theories do not offer any explanation and for others, the theories' assumptions seem to be in contradiction with the findings. Because space restrictions do not permit a detailed critical appraisal of all theories, we restrict ourselves to a brief exemplary discussion of two of the theories.

Sholl (2001) assumes an orientation-free object-to-object representation that is subject to access by two egocentric RFs: a motor and a cognitive RF. The two frames usually coincide but can be separated. In imaginal PT the cognitive frame is assumed to be mentally rotated / translated to an appropriate place in the object-to-object representation. PT effort is assumed to be driven by the effort to separate the two frames and to mentally transform the cognitive frame. This theory has difficulties, for example, explaining why translation effort increases with pointing disparity and why rotations are generally more effortful than translations.

Mou et al. (2004) also assume an object-to-object representation. In contrast to Sholl (2001) and our theory, however, this representation is assumed to be oriented (i.e., tightly coupled to a RF). If the imaginal frame is aligned with the representation's frame, information can be directly retrieved from memory. If the frames are misaligned, the relation has to be inferred. The mechanisms underlying this inference are sometimes declared outside the scope of the theory (Rump & McNamara, 2013) and sometimes characterized as being some form of mental transformation (Mou et al., 2004, p. 156). In either case, the theory does not offer a satisfactory explanation of some of the key findings.

Conclusion

Our theory constitutes a promising account of spatial memory of immediate environments. As we have shown, the theory provides explanations for a wide range of key findings and a computational realization of the theory accounts well for human behavior in pertinent empirical studies. Moreover, the theory's view on spatial memory of immediate environments also fits well into frameworks of how larger-scale space representations are assembled as networks of more local representations of immediate environments (e.g., Chrastil & Warren, 2014; Meilinger, 2008).

Our theory suggests that enduring spatial memory representations may best be viewed as consisting of two main parts: a representation structure (e.g., the circular structure described above) and a RF. In particular, structure and RF may be flexibly combined such that the same structure / RF can yield different representations when combined with different RFs / structures. Such a view promises a more parsimonious account of spatial representations, because a comparatively small set of structures and frames may be sufficient to explain a wide range of spatial abilities. It also highlights an interesting possible connection between spatial language use and spatial reasoning through sharing RF selection mechanisms.

Future work will focus on refining the computational realization of the theory (see above) and on extending simulations to include further experiments.

Acknowledgments

The research reported in this paper has been partially supported by the German Research Foundation DFG, as part of Collaborative Research Center (Sonderforschungsbereich) 1320 EASE – Everyday Activity Science and Engineering (http://www.ease-crc.org/). The research was conducted in subproject P03 Spatial Reasoning in Everyday Activity.

References

Avraamides, M. N., & Kelly, J. W. (2008). Multiple systems of spatial memory and action. *Cognitive Processing*, *9*(2), 93–106.

Botvinick, M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624-652.

Brockmole, J. R., & Wang, R. F. (2003). Changing perspective within and across environments. *Cognition*, *87*, B59–B67.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, *114*, 340–375.

Chrastil, E. R., & Warren, W. H. (2014). From cognitive maps to cognitive graphs. *PLOS ONE*, *9*(11), 1-8.

Cooper, R. P., & Guest, O. (2014). Implementations are not specifications: Specification, replication and experimentation in computational cognitive modeling. *Cognitive Systems Research*, *27*, 42 - 49.

Easton, R. D., & Sholl, M. J. (1995). Object-array structure, frames of reference, and retrieval of spatial knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 483–500.

Madras, N. (2002). *Lectures on Monte Carlo Methods*. American Mathematical Society.

May, M. (2004). Imaginal perspective switches in remembered environments: Transformation versus interference accounts. *Cognitive Psychology*, *48*(2), 163 - 206.

May, M. (2007). Imaginal repositioning in everyday environments: effects of testing method and setting. *Psychological Research*, *71*(3), 277–287.

Meilinger, T. (2008). The network of reference frames theory: A synthesis of graphs and cognitive maps. In C. Freksa, N. S. Newcombe, P. Gärdenfors, & S. Wölfl (Eds.), *Spatial cognition vi. learning, reasoning, and talking about space* (pp. 344–360). Berlin: Springer.

Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 142–157.

Rump, B., & McNamara, T. P. (2013). Representations of interobject spatial relations in long-term memory. *Memory & Cognition*, *41*(2), 201–213.

Schöner, G., Spencer, J. P., & DFT Research Group, T. (2015). *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford University Press.

Schultheis, H., Bertel, S., & Barkowsky, T. (2014). Modeling mental spatial reasoning about cardinal directions. *Cognitive Science*, *38*(8), 1521-1561.

Schultheis, H., & Carlson, L. A. (2017). Mechanisms of reference frame selection in spatial term use: Computational and empirical studies. *Cognitive Science*, *41*(2), 276-325.

Sholl, M. J. (2001). The role of a self-reference system in spatial navigation. In D. R. Montello (Ed.), *Spatial information theory* (pp. 217–232). Berlin: Springer.

Waller, D., & Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 867–882.

Wang, R. F. (2005). Beyond imagination: Perspective change problems revisited. *Psicologica*, *26*, 25–38.

Wang, R. F. (2017). Spatial updating and common misinterpretations of spatial reference frames. *Spatial Cognition & Computation*, *17*(3), 222–249.