Qualitative and Quantitative Reasoning and Instance-Based Learning in Spatial Orientation

Glenn Gunzelmann (glenn.gunzelmann@mesa.afmc.af.mil)
Air Force Research Laboratory, 6030 South Kent Street
Mesa, AZ 85212 USA

Don R. Lyon (don.lyon@mesa.afmc.af.mil)
L3 Communications at Air Force Research Laboratory, 6030 South Kent Street
Mesa, AZ 85212 USA

Abstract

This paper describes an experiment and a computational cognitive model involving a spatial orientation task. The experiment tests participants' ability to identify their location on a map, given a view of the space. The model performs this task by applying an instantiation of a general strategy that has been shown to be effective in other spatial orientation tasks. It uses perceptual grouping to organize the space into recognizable elements (clusters), combined with qualitative (left versus right) and quantitative (egocentric bearing) information about those clusters to determine its response. The model is compared to the empirical data, showing good overall correspondence to human performance, including response times and errors.

Introduction

There are countless applications of using a map, from finding an attraction at an amusement park, to coordinating complex military operations in unfamiliar territory. Most of these applications require the map user to answer a very basic question: 'How does the map correspond to the surrounding visible environment?' Frequently, individuals attempt to answer this question by trying to determine their position on the map. This involves determining both the appropriate location on the map and the direction the individual is facing (i.e., orientation). Once these two pieces of information are known, it is possible to effectively use the map to guide decision-making. Without this information, it is impossible to reason effectively to make appropriate spatial and navigational decisions.

A large body of research has accumulated around the topic of establishing correspondence between a map (both physical and cognitive) and a visible space. Much of it has concerned the relationship between the alignment of the map and the orientation of the user. This research has shown repeatedly that when the map's vertical axis is misaligned relative to the user's orientation, performance is impaired. Studies have demonstrated this in a variety of contexts, involving familiar and novel spaces (Thorndyke & Hayes-Roth, 1982), large geographic regions to small-scale room-sized environments (Glicksohn, 1994; Hintzman, O'Dell, & Arndt, 1981), and simple left-right judgments to complex locating tasks (Shepard & Hurwitz, 1984; Sholl, 1995). They have shown that increasing misalignment results in more errors and longer response times.

The existing research has established a map misalignment effect, but it does not provide a process-level understanding of map-use behavior. For this purpose, we need a valid model of the cognitive processes that people employ when they try to orient themselves using maps. We need to understand how and why people make errors, and what factors affect response times. The remainder of this paper describes research that is targeted at producing detailed, quantitative explanations of human performance on this kind of spatial task. We have developed a computational model that performs a version of a map-scene orientation task (described below). It uses a strategy that is based on previous work using a similar task (Gunzelmann & Anderson, 2004; 2006). The model produces response times and errors that are similar in important ways to data from human participants performing the task.

Experiment

The task used here required participants to determine their location on a map, based on information available in a visual scene (see Figure 1). It involved a circular space containing 10 identical objects. Each trial displayed an egocentric perspective from a point on the edge of the space, along with a map. All 10 objects were visible in both views, and the center of the space is identified with a light (green) dot on each view. The viewer was always facing toward the center of the space (the light green dot visible in both views), which allowed us to focus on understanding how individuals determine their location in a space, without the added complexity of an unknown orientation.

Surely, the task in Figure 1 represents a simplification of naturalistic tasks, but it captures an important component of them. Specifically, our task requires that local cues in the environment be used to identify the current location on a map. Therefore, this task provides an opportunity to look at how people identify their location on a map in a controlled way. This represents the first step in the process of using a map to guide spatial decision making, a difficult process that causes some trouble for many individuals.

Figure 1: Sample trial. Participants click on the map to indicate where they believe the viewer is standing.

Method

There were 8 participants in this study (6 males and 2 females), with a mean age of 28.5 years. In each trial, a space containing 10 objects was presented, as illustrated in Figure 1. The locations of the objects were constrained by dividing the space into quadrants and requiring that quadrants opposite to each other contain equal numbers of objects. For half of the trials, quadrants contained 1 or 4 objects (Figure 1); for the remaining trials, each quadrant contained 2 or 3 objects. The location of each object within the quadrant was random. To examine how the particular locations of object clusters impacts performance, the axes used to divide the space into quadrants on the map was varied as well. By rotating those axes in 15° increments, it was possible to create 12 unique orientations of the quadrant boundaries (a rotation of 180 degrees produces a space with an identical quadrant configuration as an unrotated space).

In addition to the different object configurations, the viewer's position was varied systematically to manipulate the degree of misalignment between the two views. The viewer's location was varied in 15 degree intervals as well, beginning directly at the bottom of the map (a total of 24 possible locations). We used this many locations to make it difficult to determine that a discrete set of viewer locations was being used. The final design, then, comprises 576 trials (2 different distributions of objects within quadrants, 12 rotations of the quadrant orientation on the map, and 24 possible viewer locations in each of these cases).

For each trial, a unique space was generated that met the constraints specified by the combination of levels for the three factors. The experiment incorporated a drop-out procedure, such that when a participant made an error on a particular trial, that trial was presented later in the experiment. The constraint on this was that the same trial could not be presented twice in a row (unless it was the last remaining trial in the experiment). As a result, participants had to respond correctly to all 576 possible trials in order to complete the experiment. Each time a trial was presented, a new and unique space was generated. Thus, participants never saw the same configuration of objects twice.

To complete each trial, participants were asked to click on the location on the map where they believed the viewer was standing based on the visual scene on the left. Clicks not falling on the dark ring around the outside of the map were ignored. Responses that fell within 15 degrees of the correct location were counted as correct. All other responses were errors. The experiment was broken into blocks of twenty trials, and participants were allowed to take a short break between those blocks. After each block, a window appeared indicating the number of trials completed, and how many of them were correct. Progress through the experiment was indicated after every 100 trials by showing what percentage of the total trials had been successfully completed. The experiment was further divided into two-hour sessions. Participants required from two to four sessions to complete the experiment, and were paid $10/hour for their efforts.

Results

The data from this experiment provide a rich source of information about the difficulty of the task. However, because of limited space, this paper will only address a subset of these results. We will not discuss effects of object clustering (i.e., groups of 2 & 3 versus 1 & 4), nor of the orientation of the quadrant axes on the map. We will instead focus on the influence of misalignment on performance. Although previous research found that greater misalignment tends to result in more frequent errors, that was not the case here (Figure 2; all figures show standard errors for human data). There was no significant effect of misalignment on the proportion of errors, $F(12,84)=0.71$, $p>.50$. Also, the slope of the best-fitting line for this effect (-0.0007) was not significantly different from 0, $t(7)=0.07$, $p>.90$.

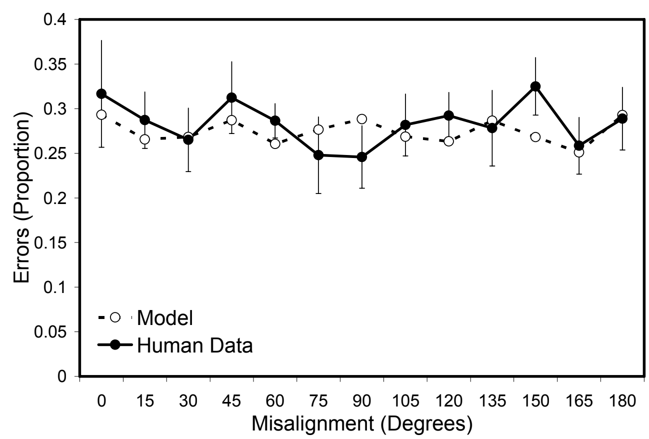

Figure 2: Error rate (proportion) as a function of the misalignment between the two views.

Though misalignment did not affect errors, it had a substantial effect on response times (Figure 3). The average response time for correct responses across the entire experiment was 13.55 s. However, when the two views were aligned, the average response time was 10.50 s, increasing to 15.28 s when the two views were maximally misaligned (180°). This effect was highly significant, $F(23,161)=4.75$, $p<.01$. In addition, when the data are averaged over left and right misalignments (as shown in Figure 3), there is evidence of a linear trend, F(1,7)=13.36, p<.01, in line with previous research.

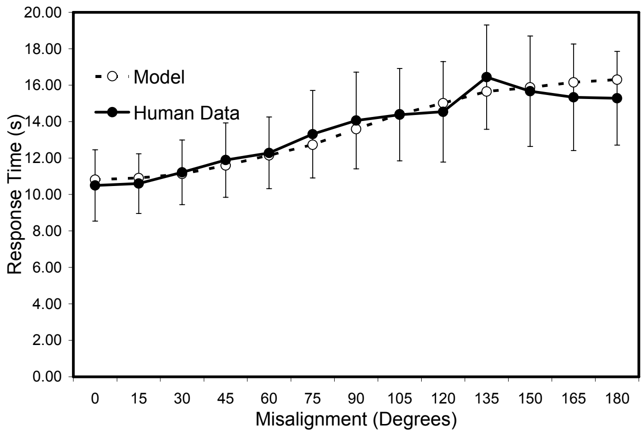

Figure 3: Response time (s) as a function of the misalignment between the two views.

Discussion

The results discussed above illustrate that misalignment has an impact on performance, but only in terms of response times. This result provides important information about human performance on this kind of task. It suggests that misalignment causes participants to take longer to establish correspondence between elements in the visual scene and elements on the map, which is a necessary first step in making comparisons between the elements in the two views. In the next section, we describe a computational cognitive model that provides a quantitative account of human performance on this task, including errors. In addition to the data presented so far, other details of human performance are explored to evaluate the mechanisms in the model in more detail.

Computational Model

The model presented here was developed in the ACT-R cognitive architecture (Atomic Components of Thought, Rational; Anderson et al., 2004). ACT-R is a unified theory of cognition that has been implemented as a running simulation. In ACT-R, a fundamental distinction is made between declarative and procedural knowledge. Declarative memory stores facts and information in the form of chunks with various slot values, while procedural memory stores actions and operators as productions.

In addition to the declarative and procedural components of ACT-R, there are other modules that have been integrated. For instance, there are perceptual and motor modules, which process stimuli and elicit actions, using realistic timing constraints for these processes. These modules allow ACT-R to interact directly with software-based tasks, which is essential for tasks like the one used here. Each module has one or more buffers, which hold the results of processing from within the module.

The overall architecture is a production system, where the current state, defined by the contents of the buffers, is used to select an appropriate action (production). This production is executed (fired), which serves to create a new state by modifying the buffers or making requests of the modules. For instance, if a production makes a retrieval request of the declarative module, the module will replace the contents of the retrieval buffer with the result of the request, allowing the production system access to it. In addition to the symbolic level consisting of chunks and productions, ACT-R's behavior is influenced by subsymbolic equations that govern quantities like declarative activation and production utility. These values can be learned through experience, based upon the statistics of a model's interaction with the environment. Some of these mechanisms are described next.

Relevant Architectural Mechanisms

The model described here is influenced to a large extent by information in declarative memory. Productions are able to make retrieval requests of the declarative module, which may include constraints on what information is desired. When a retrieval request is made, the probability that a particular chunk, *i*, will be retrieved (i.e., deposited into the retrieval buffer) is governed by the equation:

$$\mathrm{Prob}_i = \frac{e^{M_{ip}/t}}{\sum_j e^{M_{ip}/t}}$$

where M_{ip} is the "match score" of chunk *i* in the context of the slot values indicated in the request in production *p*, and the summation is over all chunks, *j*, of the appropriate type in declarative memory (e.g., numbers). The parameter *t* is the temperature, which represents the amount of noise in the system. The match score of a chunk is defined as:

$$M_{ip} = A_i - D_{ip} + \varepsilon$$

In this equation, A_i is the activation of chunk *i*, and D_{ip} is the degree of mismatch between chunk *i* and the chunk requested in production *p*. Using this match score as the basis for chunk selection makes it possible that a chunk with different slot values from the requested chunk will be retrieved from memory. The addition of noise, *ε*, to this calculation means that it is not always the best-matching chunk that is retrieved. The activation of a chunk, *i*, is a combination of the base-level activation of the chunk, plus an associative component that allows the current context to influence the level of activation. Base level activation is affected by experience – it is higher for chunks that are used (retrieved) more often and more recently.

Finally, the degree of mismatch between two chunks, *i* and *j*, is a measure of how different they are. In ACT-R, this value is reflected in a “similarity” value. This mechanism is important in this model in the context of numbers. At various points, the model encodes numerical quantities, which need to be compared to information stored in declarative memory. Following previous research on this topic (Lebiere, 2005), similarity between numbers in this model decreases exponentially as a function of the difference between them. The equation is:

$$Sim_{ij} = -1 + \frac{1}{1 + |x - y|}$$

Here, x and y are the numerical values represented by chunks i and j. This equation means that no penalty is assessed when the numbers match (sim_{ij} = 0). However, as the difference between the numbers increases, the mismatch penalty increases. Similarity is also relevant to the model’s decision about when to make a response. This aspect of the model is described in the next section, which focuses on the details of the model we have developed.

Model Implementation

The strategy implemented in the model was derived from the strategy described in Gunzelmann & Anderson (2004) for performing a somewhat different orientation task. The main principles in that model, including hierarchical encoding and a focus on groups, or clusters, or objects, were used to generate a strategy for the task used here. The strategy relies on using clusters of objects as a means of organizing the space, which constrains how the model solves of the task.

The model uses clusters to identify the area of the map where it believes the viewer is located. This is accomplished in several steps, which serve to progressively narrow the potential response area. The first and second passes utilize qualitative spatial relations between the viewer and groups of objects in the visual scene. For each, the model identifies a group of objects in the visual scene, and encodes whether it is in the left, right, or center of the field of view. The model then locates the corresponding group on the map, and identifies a portion of the edge where the same qualitative relation exists between the edge and the cluster.

The result of going through this process with two distinct clusters is to identify a portion of the edge of the map that satisfies the constraints imposed by the locations of both clusters (the “overlap” of the areas). Much of the time, this region will be small enough to allow for an accurate response simply by clicking in the center of it. Indeed, one of the options available to the model is to respond in this manner. However, some of the time these constraints do not sufficiently narrow the response region, and responding at the center of the overlap area will result in an error. If the model chooses to further refine its estimate, it makes a third pass, beginning with a quantitative estimate of the bearing from the viewer to one of the clusters in the visual scene. The model encodes this value and then estimates the bearing to the corresponding cluster on the map from a location in the overlap region on the edge. When the model finds a location where the difference between these two estimates is small enough, it responds by eliciting a mouse click at that location.

Instance-Based Learning There are two critical decision points in the model’s solution process. The first is deciding whether to further refine the response using quantitative estimates of bearing. The second is deciding whether each estimated response is close enough. In both of these cases, the decision is guided by an ACT-R implementation of instance-based learning (cf. Gonzalez, Lerch, & Lebiere, 2003). This mechanism involves storing information about the context and outcomes of past experiences, and using those instances to guide current decision making.

Instance-based learning allows the model to become a bit more accurate as it accumulates experience. This is because each trial adds knowledge about the relationship between the context and the outcome. The instances retrieved on a given trial should come to more closely reflect the current situation as the number of instances increases. This, in turn, should lead to choices that more often reflect accurate decision-making. Of course, noisy perceptual processes, the noise included in the subsymbolic components of ACT-R and variations in the stimuli mean that this learning never leads to perfect performance.

At each decision point, the model essentially asks the question: ‘In previous situations like this, what did I do and what happened?’ For the first decision, retrieval of an instance of a correct response where no refinement was done is evidence that responding at the midpoint of the overlap arc is a good strategy. Other retrievals suggest that further refinement would be helpful. The critical piece of information used to guide this retrieval is the size (in degrees) of the overlap area on the map. This is a numerical value that allows the similarities between the numbers to influence which instance is actually retrieved (see above).

In the refinement stage, the model encodes an estimate of the bearing to one of the clusters in the visual scene and estimates the bearing from a potential response location to the corresponding cluster on the map. Both of these values are noisy, with the noise value randomly sampled from a normal distribution with a mean of 0 and a standard deviation that increases as a function of the bearing, to reflect biases and error in perceptual encoding (e.g., Appelle, 1972). The model retrieves an instance from memory based upon the difference between these two noisy values. Again, this is a numerical value. Chunks in declarative memory with more similar values are more likely to be retrieved.

If the previous instance that is retrieved was correct, the model responds at the estimated response location. If an instance of an incorrect response is retrieved, the model revises its estimate a little in the direction that serves to reduce the difference, estimates a new bearing, and retrieves a new instance based upon the resulting difference using the updated information. This process repeats until the model retrieves a previous correct response from memory.

Experience with the task teaches the model that smaller overlap areas and smaller discrepancies between bearing estimates are more likely to result in correct responses. This allows the model to show a small increase in accuracy over the course of the experiment (Figure 4). This application of instance-based learning makes the mechanisms governing the retrieval of chunks from declarative memory critical in determining the model's performance, particularly similarity values. The more similar two values are, the more likely they are to be confused. So, the model will tend to retrieve instances that are more similar to the current context, but often not exactly the same.

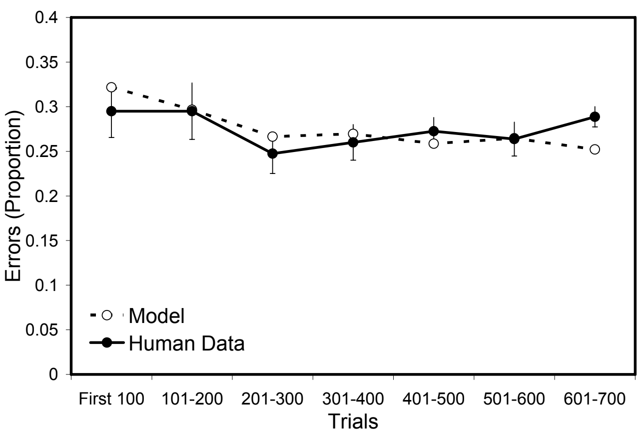

Figure 4: Error rate (proportion) as a function of practice.

Parameters and Details The previous section illustrated the importance of similarity between numbers in influencing the model's performance. There is one other similarity value that plays an important role in the model's performance. This is the similarity between *correct* and *incorrect*, which influences the decisions of whether to refine its response and when to respond. Essentially, the similarity between these two values provides the resolution to the speed-accuracy tradeoff for the model. A greater focus on retrieving a previous correct response reflects an emphasis on speed, whereas focusing on previous errors places more weight on accuracy. This parameter may provide a useful way of understanding some of the individual differences in performance on this task. The model here is moderately biased toward accuracy. That is, at the decision points in the model, it attempts to retrieve a previous error from memory, with the similarity between *correct* and *incorrect* set to -0.58 (about the same as the similarity between two numbers that are 1 unit apart). This parameter was estimated for the current data set.

The only other ACT-R parameter that was explicitly manipulated for this model was the execution time for one of the productions. The default execution time in ACT-R is 50 ms. In the model, there is a production that performs mental rotation to align the perspective of the viewer with an estimated location on the map. This production was given an execution time of 200 ms, to reflect the cognitive effort required to perform this spatial transformation. The need for such a parameter illustrates that the ACT-R architecture lacks a strong theory of mental imagery and performing spatial transformations. This is a focus of current research. Hopefully, future versions of the model will have a more elaborate mental image manipulation system with mechanisms based on research in the area.

Model Performance

Although the trial drop-out procedure should have motivated participants to be accurate, overall accuracy was only 71.7%, ranging from 61.7% to 80.8% for individual participants. While the model showed a smaller range of accuracy (from 69.7% to 75.5%), it corresponds well in terms of overall accuracy (72.7%). It would be possible to capture the range of performance by using different *correct*/*incorrect* similarity values for different model runs, but that was not the goal of this effort.

As shown in Figure 2, misalignment was not a key influence on accuracy. The slope of this effect in the model is 0.0004 (compared to -0.0007 for the human data), and its predictions are comparable to the human data (RMSD=0.027). In addition, Figure 4 illustrates that experience with the task did not provide much benefit in terms of accuracy. Although the model generally matched human accuracy throughout the experiment (r=.551; RMSD=0.020), it showed gradual improvement over the course of the experiment as a result of the instance-based learning mechanism, whereas the human participants did not.

Finally, it is possible to examine participants' responses as a function of how far off they were from the correct location. Recall from the task description that responses that were within 15 degrees of the viewer's actual location were counted as correct. If errors represent confusion about how the spaces correspond, then they should be randomly distributed in terms of discrepancy from the correct answer. In contrast to that view, Figure 5 shows that the vast majority of errors were relatively close to being correct. This finding suggests that participants were able to establish a relatively good qualitative sense of the viewer's location, but that they failed to get close enough in their quantitative estimate. The model produces the same pattern (r=.991; RMSD=0.010), which stems from the model's use of instance-based learning as well. Similar values tend to be confused more easily. Therefore, when the model is close to the correct answer, but not quite close enough, there is a higher probability that the instance retrieved from memory will reflect a previous correct response than in cases where the estimated location is further from the correct region.

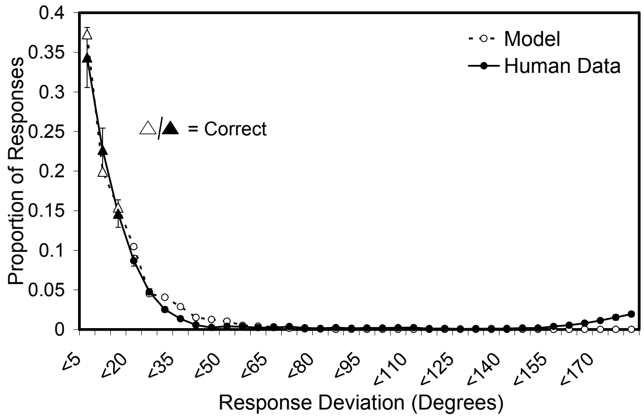

Figure 5: Response deviation (in degrees) from actual viewer location.

In addition to the accuracy data, there were some interesting results from the response time data. Participants took a considerable amount of time to make a correct response on some trials; up to 138 s! The average response time for the human participants was 13.55 s per trial. Interestingly, the model's average response time was 13.57 s per trial, with a maximum response time of 135.67 s. In contrast to the accuracy data, the response times were greatly affected by the extent of the misalignment between the views (Figure 3), and the model captures this trend (r=.968; RMSD=0.518 s). The model produces this result because it updates its frame of reference when it is processing the map. As noted above, however, more work is needed to increase the cognitive fidelity with which the model accomplishes this kind of transformation.

Conclusion

The strategy in the current model is derived from an existing model for a different task, completed by different participants, yet it still produces performance that captures major trends in the human data. In addition, this model represents a substantial increase in sophistication over the model described in Gunzelmann and Anderson (2004). This model incorporates an instance-based learning mechanism, which allows the model to both make errors and learn about the task. Consequently we believe that this model is a better representation of human performance in the task.

Finally, since the model gathers the information needed to perform the task by shifting its visual attention to the relevant areas of the display, it generates a set of predictions about the eye movements of individuals as they solve the task. In the experiment described above, eye movement data were collected from participants as they performed the task. These data provide a rich source of information about the moment-to-moment activities of individuals as they decide on a solution. The predictions of the model will be validated against these data, which will provide evidence about the appropriateness of the current model as well as additional constraints to be used in refining and extending the account of performance it represents.

Acknowledgments

This research was supported by the Air Force Research Laboratory, Warfighter Readiness Research Division, and by the Air Force Office of Scientific Research (AFOSR), Grant #02HE01COR. We would like to thank Sharon Lebeau for her assistance in data collection.

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.

Appelle S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin, 78*, 266–278.

Glicksohn, J. (1994). Rotation, orientation, and cognitive mapping. *American Journal of Psychology, 107*, 39-51.

Gonzalez, C., Lerch, F. J., & Lebiere, C. (2003). Instance-based learning in real-time dynamic decision making. *Cognitive Science, 27(4)*, 591-635.

Gunzelmann, G., & Anderson, J. R. (2006). Location matters: Why target location impacts performance in orientation tasks. *Memory & Cognition, 34(1)*, 41-59.

Gunzelmann, G., & Anderson, J. R. (2004). Spatial orientation using map displays: A model of the influence of target location. In K. Forbus, D. Gentner, and T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 517-522). Mahwah, NJ: Lawrence Erlbaum Associates.

Hintzman, D. L., O'Dell, C. S., & Arndt, D. R. (1981). Orientation in cognitive maps. *Cognitive Psychology, 13*, 149-206.

Lebiere, C. L. (2005). Constrained functionality: Application of the ACT-R cognitive architecture to the AMBR modeling comparison. In K. A. Gluck & R. W. Pew (Eds.), *Modeling human behavior with integrated cognitive architectures: Comparison, evaluation, and validation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Shepard, R. N., & Hurwitz, S. (1984). Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition, 18*, 161-193.

Sholl, M. (1995). The representation and retrieval of map and environment knowledge. *Geographical Systems, 2*, 177-195.

Thorndyke, P. & Hayes-Roth, B. (1982). Differences in spatial knowledge from maps and navigation. *Cognitive Psychology, 14*, 560-589.