8-2004

Spatial Orientation Using Map Displays: A Model of the Influence of Target Location

Glenn Gunzelmann
Air Force Research Laboratory

John R. Anderson
Carnegie Mellon University

Spatial Orientation Using Map Displays: A Model of the Influence of Target Location

Glenn Gunzelmann (glenn.gunzelmann@mesa.afmc.af.mil)
National Research Council Research Associate
Air Force Research Laboratory
6030 South Kent St
Mesa, AZ 85212-6561

John R. Anderson (ja+@cmu.edu)
Department of Psychology, Baker Hall 342-C
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

This paper presents a model of human spatial orientation, using a task that involves locating targets on a map of the space The model uses a hierarchical solution process that was reported by many of the participants in an empirical study It encodes the location of the target in the visual scene by identifying a cluster of objects that contains the target and then encoding the position of the target within that cluster. By applying this description of the target's location to the representation on the map of the space, the model is able to correctly identify the target Using this general strategy, it reproduces all of the major trends in the empirical data

Introduction

The relative ease with which people are able to navigate through familiar and unfamiliar environments is a human ability that is not well understood. This process requires the integration of multiple sources of information, since immediate visual perception rarely provides a complete representation of a space To make informed decisions, generally additional information is necessary When the space is familiar, this information may be available in memory (e g., a cognitive map) In other cases, however, people often use external maps of a space to facilitate their decision-making.

When external maps are used in conjunction with visual perception to make spatial judgments, one source of difficulty is the difference in how spatial information is represented in the two views of the space. In visual perception, spatial information is available in egocentric terms (e g, Klatzky, 1998). That is, the locations of objects in the space are encoded in terms of their distance from the viewer and their bearing relative to the viewer So, the viewer serves as the origin and the direction the viewer is facing defines the orientation In contrast, external maps identify the orientation and origin within an allocentric frame of reference These representations are commonly oriented according to cardinal directions, with north at the top

When the frames of reference in two representations of a space are different, they must be brought into correspondence before they can be used together to facilitate decision-making (Levine, Jankovic, and Palij, 1982). This process requires the ability to identify a common point in both views of the space, along with another piece of information (a second point or a reference direction) to align the orientations. Once this is done, information can be shared between the two views to provide more complete information about the space. Orientation tasks require individuals to establish correspondence between two views of a space. Often, participants are shown a target in one view of a space and are asked to locate it in the other view. Research has shown that the difficulty of this kind of task depends on a number of factors, including the location of the target object and the difference in the orientations (misalignment) between the two views of the space (e g., Easton and Sholl, 1995; Hintzman, O'Dell, and Arndt, 1981; Rieser, 1989).

The cognitive model that is presented here illustrates a perspective for understanding how the results obtained in studies of orientation tasks arise. The model was developed in the context of the ACT-R architecture. The remainder of this paper presents a brief description of the empirical work on which the model is based, followed by a description of the model and its performance

Experiment

Participants were shown 2 views of a space containing 10 objects On the left was a visual scene showing the 10 objects, one of which was highlighted to identify it as the target. On the right side was a map of the space, indicating the locations of the 10 objects as well as the viewer's position. Participants were asked to click on the object on the map that corresponded to the target that was indicated in the visual scene Figure 1 shows a sample trial

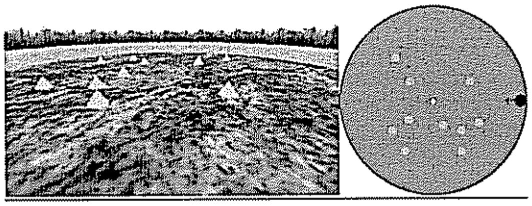

Figure 1: Sample trial for the orientation task.

Method

The spaces were created using the Unreal Tournament game engine (2001), which allows users to create their own 3-D worlds. The 10 objects in the space for each trial were placed in clusters, which were centered around one of 8 positions in the space. In each trial there were four clusters, containing one, two, three, and four objects. The positioning of the clusters was such that on some trials, there were two clusters directly in front of the viewer (one nearby and one farther away), a cluster on the left, and a cluster on the right. In the other trials, there were two clusters on each side of the space relative to the viewer, one nearby and the other farther away. The sample trial in Figure 1 shows the latter case. The configurations and locations of the clusters (in terms of the number of objects in each one) relative to the viewer were counterbalanced. This design resulted in spaces where the objects were not arranged in a well-defined pattern, making it less likely that participants would use strategies that have been described for similar tasks in the past (Gunzelmann and Anderson, 2002).

This experiment varied several factors to closely examine how they impact human performance on orientation tasks. First, the target was located in one of the clusters on each trial. As a result, the target could have been positioned in any of eight general locations relative to the viewer on a given trial. In addition, the target was located in a cluster that contained from one to four objects. So, the target was located in the vicinity of zero to three nearby distractors. Finally, this experiment involved a manipulation of the degree of misalignment between the two views of the space. The two perspectives were either aligned, misaligned by 90 degrees (clockwise or counterclockwise), or misaligned by 180 degrees (maximally misaligned).

There were 20 participants in the experiment. Ten participants completed one half of the possible trials in the experiment, while the other ten completed the other half. When they finished the experiment, participants completed a version of the Vandenberg and Kuse Mental Rotations Test, an assessment that measures spatial ability (Vandenberg and Kuse, 1978). Participants were ranked based upon their scores on this task. Using these rankings, participants were matched between the two conditions and their data were combined to create "meta-participants". The data from these meta-subjects were used in the analyses described here, although the general conclusions remain the same if the data from the two conditions are analyzed independently. Finally, the experiment utilized a drop-out procedure, such that if a participant made an error, the trial was repeated later in the experiment. More complete details concerning the methodology are available in Gunzelmann and Anderson (submitted).

Results

Response times and accuracy were recorded for each trial in the experiment. Overall, accuracy was quite high (96%), and the pattern of errors was quite similar to the response time data ($r = .83$). This suggests that the results were not due to a speed-accuracy trade-off. As a result of the high degree of accuracy, the data presented here consider the response times for correct responses that participants produced as they performed the experiment.

There are a number of important findings in this experiment, which are summarized in Figures 2, 3, and 4 below. First, in terms of misalignment, the data correspond well with previous research (Figures 2 and 4). As the misalignment between the two views increases, response times increase as well, $F_1(3,27)=38.62$, $p<.001$. Next, the local distractors placed around the target also had an impact on performance (Figures 2 and 3). These data show that as more local distractors were present, participants took longer to identify the correct object on the map, $F_1(3,27)=60.67$, $p<.001$. The magnitude of this effect, however, depended on the degree of misalignment between the two views (Figure 2). Specifically, the impact of the number of local distractors increased as misalignment between the two views increased, $F_1(9,81)=8.79$, $p<.001$.

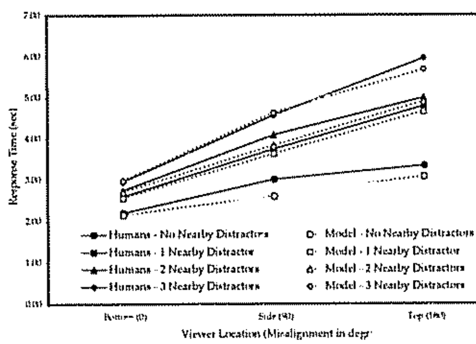

Figure 2: Response times (sec) as a function of misalignment and the number of distractors nearby to the target

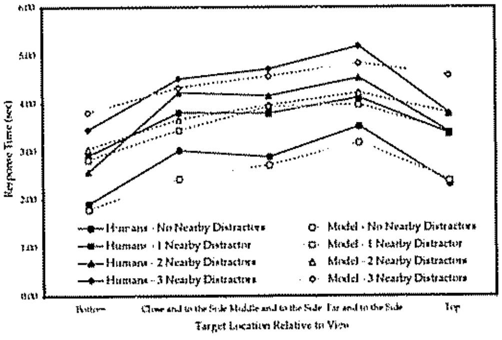

Figure 3: Response times (sec) as a function of the target's location relative to the viewer and the number of distractors nearby to the target

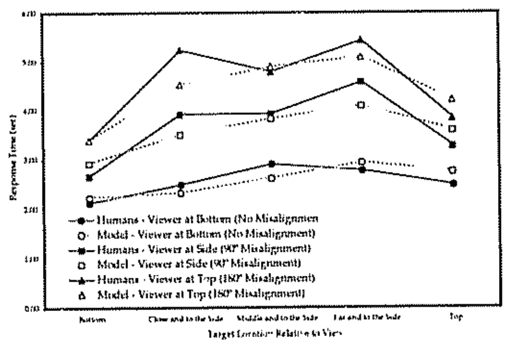

Figure 4: Response times (sec) as a function of the target's location relative to the viewer and the misalignment between the two views

The impact of the target's location relative to the viewer is shown in Figures 3 and 4 The overall effect was significant, $F_1(7,63)=11.39$, $p<.002$ This result contains two major components Firstly, response times were fastest when the target was located more directly in front of the viewer (the first and last points on each line in Figures 3 and 4) Secondly, difficulty tended to increase as the target was located farther from the viewer on one side of the visual scene or the other (the remaining points in Figures 3 and 4) Importantly, these trends were produced regardless of the number of distractors that were located near to the target (Figure 3), and this interaction was not significant, $F1(21,189)=1.79$, $p>.15$ In contrast the target's location did have an influence on the impact of misalignment (Figure 4) This result shows that the impact of misalignment was diminished when the target was in line with the viewpoint (the first and last points on each line in Figure 4) This interaction was significant in these data, $F1(21,189)=3.78$, $p<.02$

In addition to the response time data, retrospective verbal reports from participants provided evidence about how they did the task In general, participants indicated that they engaged in a two-step process to find the answer The first step involved identifying a cluster of objects that contained the target so that the cluster could be found on the map Once the cluster was identified, participants determined which of the objects in the cluster was the correct response

Discussion

The results of this experiment highlighted several factors that contribute to difficulty in orientation tasks First, misalignment between the two views of the space impacted difficulty similarly to the results of previous studies (e g , Gunzelmann and Anderson, 2002; Hintzman, et al , 1981; Rieser, 1989; Shepard and Hurwitz, 1984) Also, the findings show that the location of the target relative to the viewer within a space influences how difficult it will be to locate it on a map Unlike previous research, this result is demonstrated without using a highly organized configuration of objects in the space As the target was positioned farther from the viewer, and when the target was less directly in front of the viewer, difficulty increased These findings suggest that participants were using the viewer's location in the space as a key reference feature to help them determine the location of the target Response times were faster when the target was in a location that could be encoded more easily with respect to the viewer's position in the space

This experiment also showed that difficulty increased as more objects were located in the vicinity of the target, a factor that previous research has not addressed This result suggests that participants were considering only a portion of the space when trying to locate the target, since the total number of objects was the same for all trials In addition, this effect did not vary as a function of the particular location of the target This outcome suggests that the location of the cluster does not impact how the target's location within that cluster was encoded

The hierarchical solution process reported by participants illustrates how they were able to limit their search to a subset of the items in the space and shows why more local distractors would result in longer response times The presence of more objects near to the target requires more, or more complex, transformations to bring the information in the two views into correspondence, which should take more time (e g , Bethell-Fox and Shepard, 1988) It appears that one can view the process of solving these tasks as developing a description of the target's location, which then has to be transformed to apply to the map This description could be verbal, or could involve the creation of a mental image

ACT-R Model

ACT-R is a general theory of cognition that has been implemented as a running simulation (Anderson and Lebiere, 1998). It operates as a production system with several core assumptions related to its operation. First, there is a division between declarative and procedural knowledge. Declarative memory contains information in the form of chunks, while procedural knowledge is composed of productions, which contain information about transforming one state into another. The latest version of ACT-R is composed of a set of modules for perceptual, motor, and cognitive aspects of human performance. Information is processed independently within these modules, allowing them to operate in parallel. There are buffers associated with each of the modules that essentially represent working memory. The contents of these buffers are what drive the production system. Productions match against the contents of the buffers, and it is only the contents of those buffers that can be directly accessed. It is at the level of production selection and execution that the system operates in a serial manner.

Because ACT-R includes perceptual and motor modules, it is able to interact with experimental software under realistic constraints. Although the perceptual module currently is not sophisticated enough to parse the visual scene shown in Figure 1, it does contribute important timing information to the model's performance. The motor module adds additional constraints to the mouse movements and clicks that the model executes. The parameters that control these aspects of ACT-R's behavior are based on largely on the EPIC theory (Kieras and Meyer, 1997). At a general level, the model was implemented within this architecture to perform the task based on the two-step process described by the participants in their verbal reports. There are, however, a number of details that are important to the model's performance as it goes through this general process. These are described next.

Model Design

The model begins each trial by locating the target in the visual scene. Once this location has been identified, the model finds other objects that are in the vicinity of the target. It counts those objects, and encodes the overall location of the cluster as being in the left, right, or central portion of the visual scene. Then, to encode the location of the target in the cluster, the model revisits the items in the cluster, and encodes the target's position relative to the near-far and left-right axes. So, the model develops a representation of the target's location in the visual scene that would be something like "the leftmost object in the cluster of two on the right of the visual scene." This is the target's position in the sample trial in Figure 1.

With a representation of the target's location in the visual scene, the model shifts its attention to the map of the space, beginning by locating the viewer's position, which is indicated. The next step is to find the correct cluster. If the cluster was encoded as being in the middle of the visual scene, the model searches straight out from the viewer's location to find it. However, when the cluster was positioned to one side or the other, spatial updating is required when the two views are misaligned so that the correct portion of the map is searched. This updating consists of a remapping of "left" and "right" to the corresponding directions on the map relative to the viewer's orientation. For instance, in the sample trial in Figure 1, the right portion of the visual scene corresponds to the top half of the map.

The updating process is a source of difficulty which requires extra time. In addition, when the two views of space are maximally misaligned (the viewer is at the top of the map), the updated values for the map directly conflict with the egocentric values. This adds a second source of difficulty to the updating process, which adds additional time to its execution. Once the values are updated, the model is able to search the appropriate portion of the map for the correct cluster. To perform this search, the model begins near to the viewer's location and searches outward until it finds an object that is in a cluster of the appropriate size.

Once the first step of finding the appropriate cluster is completed, the model needs to determine which of the objects within the cluster on the map is the target. Like the cluster location, the encoding of the target's position is based in the egocentric coordinate system from the visual scene. So, when the two views are misaligned, updates to this information are needed to match the rotated coordinate system of the map. These updates are similar to those described above. Misalignment is one source of difficulty, while direct conflict between the two reference frames is another.

One detail of this process requires some explanation. In the model, the amount of updating done in the second step depends on the number of objects in the cluster. When there are no nearby distractors, this step is skipped. In this case, when the "cluster of one" is found, the model is immediately able to respond by clicking on that object. In cases where there are 2 or 3 objects in the cluster, there is a simple encoding of the target's position within the cluster that requires only one axis. With 2 objects, the target is always on the left or on the right, and is also always the closest or farthest object in the cluster. When the cluster has three objects, the target can also be the one in the middle on each of the axes. In the model, this possibility is represented by having the model update only one of the axes in order to locate the target within the cluster.

When the cluster has 4 objects, the encoding necessarily becomes more complex. This is represented by having the model update both axes when the cluster has four objects in it Once again, there is the potential for direct conflict between the two frames of reference in these updates, which adds to the difficulty of this operation The basic idea is that the complexity of the description needed to encode the target's location increases as the number of objects in the cluster increases As a result, the difficulty of transforming this description so that it applies appropriately to the map increases as well This notion is supported by past research, which has demonstrated that more complex figures take longer for individuals to mentally rotate (Bethell-Fox and Shepard, 1988)

The model's performance is modulated by several parameters First, as noted above ACT-R's perceptual mechanisms are not currently sophisticated enough to process a raw image like the one shown in Figure 1 Thus, as a simplification, the model is presented with a 2-D, egocentrially-oriented representation of the visual scene, which essentially is another map So, the model implicitly embodies the assumption that participants extract a 2-D representation from the visual scene as they encode the information from it A constant of 25 seconds was added to the model on each trial to represent the cost of extracting such information from the visual scene

The second parameter in the model was the retrieval time, which was set to 11 seconds As the model does the task, it requires some declarative information (mostly related to directional information) Each time a chunk is retrieved from declarative memory, it takes 11 seconds However, most of the model's performance is driven by the information on the screen, so this parameter does not play a large role in determining the model's predictions

The only other parameter that was manipulated in the model controls how long it takes to perform the operations needed to update the directions (left, right, up, and down) that it uses to locate the target on the map The parameter was set so that each of the updates requires .60 seconds This value applies to each operation that is necessary, and is also applied when direct conflict arises between the allocentric frame of reference and the original egocentric reference frame This means that if one axis needs to be updated, it will take .60 seconds If two axes need to be updated, it will take 1 20 seconds However, if in addition to the update there is direct conflict between the two frames of reference, these updates take 1 20 and 1 80 seconds respectively These costs apply to the updates needed to locate the cluster and to identify the target within the cluster When an update is needed to identify the portion of the map where the cluster is located, it involves updating a single axis (left-right) When the cluster has been located and the search begins for the target, the update involves one axis when there are one or two nearby distractors, and two axes when there are three nearby distractors Each of these updates may or may not involve direct conflict that needs to be resolved in addition to the update All of the other parameters in the model were given their default ACT-R values

Model Performance

The model captures all of the major trends in the data First, the model reproduces the misalignment effect (Figures 2 and 4) As the misalignment between the two views increases, the model takes longer to respond In the model, this effect comes from the costs of updating the frame of reference to find the cluster on the map and to find the target within the cluster In addition, the costs associated with the second update depend on the size of the cluster, producing the interaction between misalignment and the number of nearby distractors shown in Figure 2 As the number of nearby distractors increases, the impact of misalignment increases This illustrates the idea that it is more difficult to update the descriptions of the target locations when those descriptions are more complex In the model, it is the extra cost associated with the spatial updating process as more nearby distractors are present that produces this interaction The mechanisms in the model capture the effect of both misalignment and the number of distractors well, with an overall correlation of 992 for the data shown in Figure 2 (RMSD= 187 seconds)

The model makes predictions about the difficulty of the task based upon the target's location in the visual scene as well These data are shown in Figures 3 and 4, along with the empirical results The model produces a good qualitative fit to the data, although the particular values are a little off in some instances The model's predictions arise because it begins its search for the cluster on the map from the viewer's position, moving outward until it locates an object in the cluster Thus, when the target is farther from the viewer, it takes the model longer to locate an object in the cluster In addition, the model produces an interaction between target location and misalignment (Figure 4) The effect of misalignment is smaller when the target is directly in front of the viewer (bottom and top target locations) because no spatial updating is necessary to find the cluster As noted above, the same effect appears in the empirical data, and the model captures the effect well (r= 954, RMSD= 325 seconds for the data in Figure 4)

Finally, there is no interaction in the model between the number of nearby distractors and the location of the target in the visual scene (Figure 3) This corresponds to the empirical results as well (r= 910, RMSD= 351) The result is because of the two-step process, reported by participants, that the model uses to do the task The target's location within the cluster is encoded without regard to the location of the cluster So, the impact of the target's location results from the search for the cluster. Similarly, the number of nearby distractors only impacts the solution process after the cluster has been located, when the correct target must be identified.

Conclusions

Overall, the model produces data that are in line with the performance of the human participants, which lends support to the conclusion that they were using the strategy they reported to do the task. The model produces all of the major trends, in most cases with data that are very close to the data from the human participants. In the model, misalignment impacts both the search for the cluster and the search for the target within the cluster. In contrast, the location of the target only influences the search for the cluster, both in terms of its distance from the viewer and whether or not it is in line with the viewer's position. The interaction of target location with misalignment in the model arises because no spatial updating is needed when the target is directly in front of the viewer. Finally, the number of nearby distractors only impacts the search for the target within the cluster, interacting with misalignment because of the different amounts of spatial updating required based on the number of objects in the cluster. Note that the location of the target does not interact with the number of nearby distractors, suggesting that they affect different aspects of the solution process. The performance of the model supports the conclusion that similar processes are being used by the participants.

In conclusion, this model provides a framework for understanding human performance on spatial tasks. It's most important characteristics relate to the hierarchical encoding of the target's location in the visual scene. This encoding allows the model to limit its search to a portion of the map, ignoring many of the objects in the space. In addition, the model's performance assumes that the two steps in the solution process are independent. As a result spatial updating that is performed for step 1 does not carry over to the execution of step 2. This contributes to the large effect of misalignment on the model's performance. Finally, the model also indicates that perceptual-motor aspects of performance are important factors in this kind of task. The time needed to execute shifts of visual attention contribute to many of the effects described here, especially the impact of the target's location and the impact of the number of local distractors. These issues deserve careful attention in future research.

Acknowledgements

This research was supported by AFOSR grant #F49620-99-1-0086 to Dr. John R. Anderson at Carnegie Mellon University and by NRSA training grant #5-T32-MH19983 from NIMH to Dr. Lynne Reder at Carnegie Mellon University.

References

Anderson, J. R., & Lebiere, C. L. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance. 14*, 12-23.

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*, 483-500.

Gunzelmann, G., & Anderson, J. R. (2002). Strategic differences in the coordination of different views of space. In W. D. Gray and C. D. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 387-392). Mahwah, NJ: Lawrence Erlbaum.

Gunzelmann, G., & Anderson, J. R. (submitted). Location matters: Why target location impacts performance in orientation tasks.

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

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction, 12*, 391-438.

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel, and K. F. Wender (Eds.). *Spatial cognition. An interdisciplinary approach to representing and processing spatial knowledge* (pp. 1-17). New York: Springer-Verlag.

Levine, M., Jankovic, I. N., & Palij, M. (1982). Principles of spatial problem solving. *Journal of Experimental Psychology. General, 111*, 157-175.

Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1157-1165.

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

Unreal Tournament [Computer software] (2001). Raleight, NC: Epic Games, Inc.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual & Motor Skills, 47*, 599-604.