Strategic Differences in the Coordination of Different Views of Space

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

Abstract

Participants were trained to use one of two different strategies in an orientation task, which were based on verbal reports from participants in another experiment. The data provide support for the conclusion that participants in the two training conditions searched the screen differently to complete the task, but that neither group used mental transformations like image rotation. These results have implications for research in this area as well as for conceptualizing how individuals perform such tasks. A comparison of the results from the two strategy conditions is made based on an ACT-R model of one of them. Small differences in how information on the screen is scanned can produce the observed differences in performance.

Introduction

The coordination of different views of space is a fundamental task in human functioning. An everyday example of it involves determining which way to turn at an intersection by using a map. The visual scene presents one view of the space (egocentric), while the map presents an alternative representation (allocentric). In order to accurately decide which way to go, it is necessary to bring these two views of the space into correspondence. Of course, with a physical map it may be possible to actually rotate it to align it with your own orientation. In other situations, mental transformations may need to be done in order to coordinate these views to make accurate decisions.

On a continuum of reasoning about orientation within a space, deciding whether the correct turn is left or right is a fairly straightforward task. Still, research on this issue has shown that it becomes increasingly difficult to perform as a function of the difference in orientation between the two views of space (Shepard and Hurwitz, 1984). The phenomenon bears a strong resemblance to findings in the mental rotation literature (Shepard and Metzler, 1971) where the time needed to determine that two objects are identical increases linearly as a function of the angular disparity between them. These findings have been used to support the conclusion that performance in orientation tasks involves analog mental rotation of mental images. Note, however, that the task of coordinating views of space adds a layer of complexity to the traditional mental rotation task. In a spatial orientation task, the information is presented in two different formats. Thus, deciding whether the visual scene matches the information on the map requires additional reasoning beyond the image transformation.

In an important series of experiments, Hintzman, O'Dell, and Arndt (1981) had participants perform orientation tasks in a variety of ways. In the basic task, participants had to indicate the direction of a target relative to a given orientation. Figure 1 shows the orientation task used in the experiment presented here. In this figure, the left side represents the target field as viewed from a camera (on a plane above the field) and the darkened circle indicates the target. The right side represents a map-view with the target field at the center. The arrow on that side shows the camera's orientation for viewing the target field. Participants are asked to indicate in which cardinal direction the target is located relative to the center of the target field. In the sample trial in Figure 1, the correct response is South. The general finding is that decisions for targets in line with the assumed orientation are made more rapidly, and response times for other targets increase as they depart from the nearest point immediately in front of the viewpoint. Although not explicitly addressed by Hintzman, et al., this increase in response time is not strictly linear. In addition, no evidence was presented in their study about how participants claimed to be performing the task.

In order to investigate what factors influence performance on this task, we asked participants to complete the task and then questioned them as to the manner in which they solved it. While we will not go into detail about this experiment, the data are presented below and bear a strong resemblance to results from similar studies, including Hintzman, et al. (1981). However, by questioning participants after they had completed the experiment, we discovered that participants were using at least two distinct strategies to do the task. Some participants claimed to be implementing a strategy that incorporated imagery and mental rotation to determine correct responses.

However, other participants indicated that they used a different strategy altogether, one that did not depend on mental imagery or mental rotation at all.

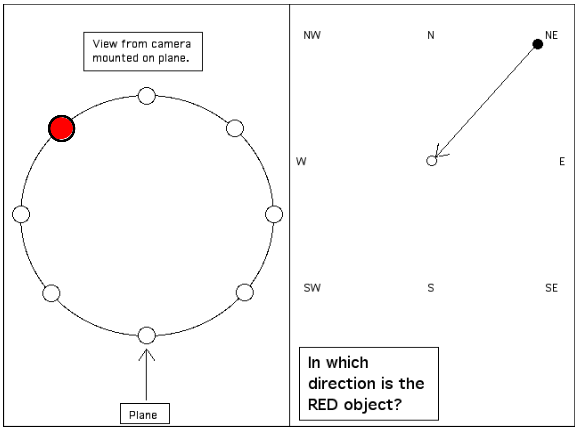

Figure 1: Sample trial for the orientation task.

In the imagery-based strategy, participants reported forming an angle connecting the camera to the target in the camera view with the vertex at the center of the target field (a 135 degree angle in Figure 1). This angle was then mentally transformed to line it up with the position of the camera on the map view. A second group of the participants simply counted around the target field to the target on the camera view (3 in Figure 1), and then counted that number of steps from the camera's position in the appropriate direction on the map view.

Both strategies are equally effective and valid for doing the task, but one depends on mental imagery while the other does not. In addition, the verbal reports indicate that few of the participants treated the task as an orientation task. Rather, the strategies they reported suggest that they treated the task more like a traditional mental rotation experiment. That is, participants effectively eliminated the added level of complexity introduced by having different representations of the information by selecting strategies that bypassed the need to consider them. This finding casts some doubt on some previous explanations for performance on similar tasks. The experiment presented in detail below was conducted to further investigate the implications of these strategies, by training participants to perform the task using either the imagery-based "angle" strategy or the more analytic "counting" strategy.

Experiment

Previous research aimed at addressing performance on tasks similar to the one presented here have based explanations largely on mental imagery and mental rotation (Shepard and Hurwitz, 1984; Hintzman, et al., 1981), though Hintzman et al. do consider a sequential scanning explanation. However, all these explanations fail to account for some of the more subtle aspects of the data and ignore the potential for different approaches to the task. The experiment presented here examines strategic differences in an orientation task that is similar to those used by Hintzman, et al (1981).

If the strategies were to be implemented according to the descriptions provided to participants, there should be clear differences in performance between the two strategy conditions. For the counting strategy, the position of the target relative to the camera should greatly influence response times. That is, response time should increase linearly as a function of the amount of the counting that needs to be done. However, the location of the camera in the map view should have no impact on performance, since the strategy can be implemented identically regardless of the camera's position in the map view. The angle strategy makes the opposite set of predictions. Response time should be unaffected by the target's location, since the angle to be formed is similar in complexity regardless of the target's position in the camera view. However, the degree of rotation that needs to be done depends on the camera's position in the map view, suggesting that response times should increase linearly as a function of the camera's position relative to the bottom of the screen.

Method

The experimental task was based both on the experimental task used by Hintzman, et al (1981) as well as on an unmanned air vehicle (UAV) flight simulator used by the Air Force for training UAV pilots (see Gugerty, et al, 2000; Figure 1). The display consisted of two static views, an egocentric "camera" view of a target field, and an allocentric "map" view. The target field was in the center of the map view, and the perspective of the camera was identified with an arrow (the right half of Figure 1). The target field was a circle, containing eight objects equally spaced at 45 degree intervals on the circle (the left portion of Figure 1), with one of them highlighted in red to identify it as the target. Participants were asked to indicate in which cardinal direction the target was located relative to the target field's center. Responses were made using the number pad on the keyboard.

After being introduced to the experimental task, participants were trained to complete the task using either the angle or counting strategy (n=16 per condition). They first read a brief description of the strategy, and then were shown how the strategy applied to a sample trial. After that, participants completed 16 paper-based practice trials in random order. In these practice trials, participants were asked to explicitly demonstrate use of the strategy they had been taught by labeling them appropriately based on the strategy they had been taught. In the counting strategy condition, participants were taught to use positive numbers for targets on the left (clockwise from the camera), and negative numbers for targets on the right (counterclockwise). Participants in angle strategy condition were instructed to note the direction in which the angle "opened". Feedback was given on each of the practice trials by the experimenter.

After training, participants completed 4 blocks of trials on the computer. Each block included all 64 possible trials in random order. A dropout procedure was used such that if an error was made on one of the trials it was presented again later in the block. During this phase of the experiment feedback was still given after each trial, including what the correct answer was in cases where participants made an error.

Results

The results for the original experiment and the two training conditions in this experiment are presented in Figures 2 and 3. In Figure 2, response time is plotted as a function of the target's clockwise deviation from the camera. The numbers correspond to the measure of the clockwise arc from the camera position to the target on the target field in the camera view. In Figure 3, the data are presented as a function of the location of the camera relative to the target field in the map view. In the sample trial shown in Figure 1, the target angle is 135 and the camera's location is NE. One aspect of the data that should be immediately apparent from these graphs is that performance was symmetrical in terms of left and right positions of both the camera and the target. In addition, response times were somewhat faster in this experiment than in the first one. This may be a result of the training given in this experiment, which participants in the first experiment did not receive.

Finally, the training conditions used in this experiment seem to separate out two components of the data from the first experiment in terms of the effect of the target's position. Specifically, data produced by participants using the counting strategy increase linearly with the target's angular deviation from the camera. In contrast, the data produced by participants using the angle strategy show a scalloped effect, with no difference between 45 and 90 degrees (or 315 and 270 degrees). The data from the original experiment show evidence of a combination of both trends. This suggests that averaging data over all participants may not provide a complete story of the effects in this task.

At the highest level of abstraction, there was no main effect of strategy condition in average response time, $\underline{F}(1,210)=0.233, p=.63$, suggesting that at a global level both strategies were equally effective for completing the task. One has to be struck by the overall similarity of the results between the two strategy conditions and their close relation to the results from the first study, given that participants were taught quite different ways of doing the task. Despite the overall similarity, there was a significant interaction between the strategies and the particular target angle, $F(7,210)=3.534, p<.02$, as well as between the strategies and the camera angle, $\underline{F}(7,210)=3.810, p=.01$. Looking at Figure 2, response times were higher for participants using the angle strategy when the target was directly in front of the camera or when it was 45 degrees to the right or left. In terms of camera angle, Figure 3 shows that participants trained to use the angle strategy exhibit relatively longer latencies when the camera is located in a northerly position.

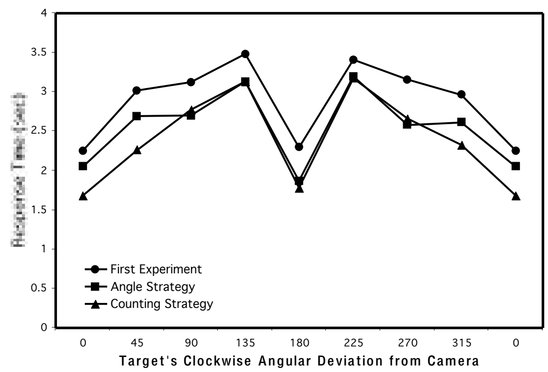

Figure 2: Response time (sec) as a function of the target's position.

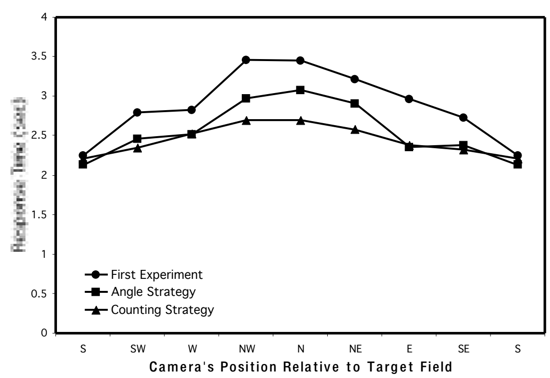

Figure 3: Response time (sec) as a function of the camera's position.

Discussion

Based upon the data, it is clear that participants were not executing the strategies precisely according to the instructions provided. In fact, only one of the predictions is clearly borne out in the data. Specifically, response times increased linearly as a function of the extent of counting for participants trained to use the counting strategy. However, these participants still showed a small effect of the camera's location. In addition, data from participants trained to use the angle strategy showed a discontinuous effect of both the camera's position and the target's relative position.

The most curious result is the effect of the target's position relative to the camera in the angle strategy. That is, the description of the angle strategy predicts no increase in response time as a function of the target's location. However, an increase does occur, and it is complicated by the discontinuity at 90 and 270 degrees. This finding, in particular, casts doubt on the claim that participants were using mental rotation at all in performing this task. In particular, it is hard to imagine how an imagery-based strategy can account for this particular effect without resorting to specialized mechanisms relating to imagining and/or manipulating 90-degree angles. Research does suggest that cognitive representations of space tend to distort angles to be closer to 90 degrees (Glicksohn, 1984), and also indicates that horizontal and vertical lines are preferred in visual perception (45 and 135 degree angles involve oblique angles; Cecala and Garner, 1986). Still, it is not clear how this should have such a large impact on the ability to manipulate or create mental images of angles of various sizes. A more likely explanation is that the differences in performance between the two strategy conditions arise from small differences in how the screen was scanned by participants as a result of their training, rather than because of differences in higher-level cognitive operations on the information.

In the counting strategy, the linearity of the target-position effect suggests that participants were indeed counting from the camera's position to find the target. The small effect of the camera's position, however, indicates that the strategy was not being implemented exactly according to the instructions. We believe that participants encoded the location of the target as being to the "left" or "right" of the camera, rather than as "clockwise" or "counterclockwise". While this is a small difference in encoding, it does have implications for locating the target on the map view. If a target location is encoded as clockwise, the map view can be scanned in a clockwise direction regardless of where the camera is located. However, if the location of the target is represented as "left" instead, the correct scanning direction is "right" when scanning from NW, N, or NE. So, whenever participants search the screen from one of these locations, extra cognitive steps are needed to make sure that the screen is scanned in the appropriate direction.

An example should clarify how we believe the counting strategy was implemented by participants. For this purpose, consider the trial presented in Figure 1. We believe that counting participants would begin this trial by locating the target on the camera view and encoding it as "3-left". At that point, they would find the camera's location on the map view. Since the camera is located at NE, the correct search direction is actually "right", so an extra operation is needed to convert the direction of scanning. Then, the screen can be scanned to locate East, and the count can then be incremented. Then, Southeast can be found, and the count incremented again, followed by South and the final increment in the count sequence. At this point, participants have located the answer and can issue their response by pressing the "2" key on the number pad (keys were assigned to correspond to the layout of cardinal locations on the screen).

Given that participants using this strategy produced data that were largely in line with predictions and the results were similar to the other condition, we decided to develop a model for the counting strategy. This is a first step to an overall model for the task, which will involve some mixture of strategies.

ACT-R Model

The ACT-R theory (Anderson and Lebiere, 1998) provides an architecture in which the proposed mechanisms can be implemented to determine how well they fit with the data. In addition, ACT-R now incorporates a theory of perceptual-motor action, allowing it to interact directly with the experimental software (Byrne and Anderson, 1998). In this way, an ACT-R model can participate in the experiment exactly as though it were a participant by gathering information from the screen using visual perception, operating on that information within its cognitive system, and issuing a response by sending commands to its motor module. This integration means that all aspects of performance are considered in the model's performance.

Model Design

There is certainly a large degree of overlap between the two strategy conditions. In particular, the details of gathering information and issuing responses in the task are assumed to be largely the same for both strategies. Thus, by understanding how participants executed one of the strategies it will be easier to understand how participants in the other condition may have performed the task. Toward that end, a model of the counting strategy has been implemented and is described next. In the conclusion, we will describe how we believe the behavior of participants trained to use the angle strategy may have differed to produce the observed results.

When a new trial is presented to the model, its first action is to search for the location of a red object on the left side of the screen. Its location is encoded as being left or right of the camera and as an integer value from 0 to 4 to define its distance from the camera. Then, the model finds the location of the camera on the map view and shifts its attention to that location.

Since it is hypothesized that the location of the target is encoded as left or right rather than clockwise or counterclockwise, the model needs to alter its scanning direction when the camera is in the NW, N, or NE positions. Once the appropriate scanning direction is selected, the model finds the nearest cardinal direction to the camera and increments its count. This process is repeated until it has incremented the count the prescribed number of times. At that point, the current cardinal location is encoded and mapped to a response on the number pad. Finally, the model issues a response by sending a command to press the correct key.

Based on verbal reports from participants, there were a couple of exceptions to this operation. First, when the target was located in line with the camera, participants reported that they did not bother to count. Rather, for target positions of 0 degrees they simply responded with whatever position the camera was in, and for target positions of 180 degrees they responded with the cardinal direction directly opposite the camera's position. The other instance where the strategies were not used was when the position of the camera was S. In this case, participants reported that they went directly from the target's location on the camera view to a response. In response to these verbal reports, these special cases were implemented in the model. These reports also correspond to data presented in previous studies (e.g., Hintzman, et al., 1981).

Model Performance

The model's performance using the counting strategy compared to the data in the two conditions is presented in Figure 4. As can be seen, it makes accurate predictions for response times for both conditions in both aggregations of the data (correlation = .98, mean deviation = .11 seconds). The model performs the task in exactly the way we believe participants are doing the task. That is, the model incorporates all of the perceptual, motor, and cognitive steps that humans would need to go through to do the task. Based on this completeness, we feel that the model captures all of the relevant aspects of participant performance.

The linear increase in response time as a function of target location is produced in the model by the simple act of counting and scanning cardinal locations sequentially. For target angles of 45 and 315 degrees, one step is counted, for 90 and 270 degrees this is done twice, and for 135 and 225 degrees there are three cycles. The small effect of the camera's position results from the left/right encoding of the target position. As described above, this creates the need to perform extra cognitive operations to switch the scanning direction at any point when searching from NW, N, and NE, thus increasing response times in trials where those situations arise. This evaluation occurs at each step in the search process. So, each time the model searches for the next cardinal location, it determines whether or not the encoded direction of the target is the correct search direction, and then alters the search direction when necessary.

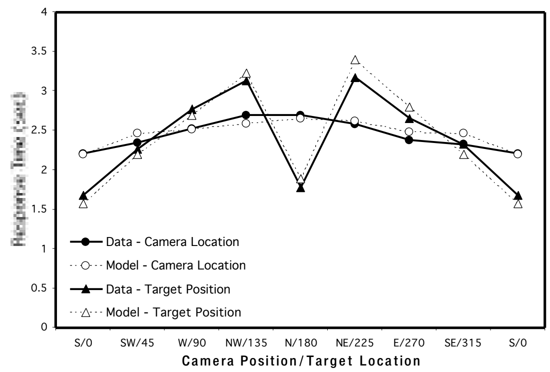

Figure 4. performance of the model of the counting strategy compared to participants' data.

In terms of the overall qualitative pattern of data, the performance of the model is parameter-free. By constructing a model that really does the task, its performance is highly constrained at this level. The parameters serve only to adjust the magnitude of the effects. First, retrievals from memory are an important aspect of the model's operation. The model retrieves various facts from memory as it performs the task, including counting sequences for the counting process, associations between cardinal directions and number keys for making responses, and information about cardinal directions for guiding the search and problem solving process. In this model, the time to perform these retrievals was set to .05 (seconds). The only other parameter that was explicitly set in this model is the execution time for the production that encodes the target's location on the camera view. This value was set to .7 (seconds) and impacts all conditions similarly. The remaining parameters all reflect default perceptual-motor parameters in ACT-R/PM (Byrne and Anderson, 1998). The model's source code is available online at http://act.psy.cmu.edu/.

Conclusions

The experiment and model presented here provide an alternative view of findings in the area of spatial cognition concerning how participants perform orientation tasks. There are two basic questions to answer. First, are participants actually performing an orientation task in these studies? The participants in this experiment were clearly not treating this task as a traditional orientation task where two distinct representations of spatial information are brought into correspondence. Rather, much of the complexity was eliminated by implementing strategies that avoided this aspect of the task. It is unclear whether similar strategic choices can achieve the same effect in more realistic orientation tasks (e.g., Gugerty, et al., 2000).

The other basic question to ask based upon these results is whether participants use mental imagery in performing the task. If they do, it is important to investigate how such cognitive abilities are applied in these tasks. If not, the question becomes what mechanisms are responsible for participant performance on these sorts of tasks. Based on the data presented here, it appears that participants assumed a more analytic approach to the task, simply scanning the screen in a systematic way to determine the correct answer. These findings also illustrate that there is variability in how participants approach virtually any task, and these variations have implications for performance.

The model shows that we can reproduce much of the qualitative form of the results in this task by implementing a strategy that involves systematically scanning the information on the screen. Moreover, this strategy corresponds to what some participants spontaneously report. However, what about the other participants who spontaneously report an angle strategy? We believe that they may be just engaging in a variant of the implemented scanning strategy, which explains why their behavior is so similar to the participants who were counting. More specifically, we believe that implementing the angle strategy involves such differences as looking at more of the information on the camera view but not systematically looking at the intermediate points between the camera and target on the map view. Both of these differences could be produced by the different training conditions in the experiment. We are currently implementing a model which incorporates such a variant of the scanning strategy and doing an eye movement study to see if we can find evidence for the hypothesized scanning patterns.

Basically our proposal is that participants prefer to process the information given on the screen rather than transform an internal image of this information. This aversion for mental transformations is consistent with the results of Kirsh & Maglio (1994) who found that people prefer to rotate objects on the Tetris screen rather than rotate them in their head. We suggest that some results attributed to mental rotation like those in this task may reflect the operation of some other process like the scanning in the counting strategy that we have implemented. While Hintzman, et al. (1981) considered sequential scanning as an alternative explanation to mental rotation, they did not consider the possibility of strategic differences in the scanning process. The results presented here demonstrate that such strategic differences exist and that some scanning strategies can result in data that approximately match predictions based on imagery and mental rotation. In addition, participants trained to use mental imagery produced data that does not fit with the imagery account. An evaluation of the model for the counting strategy suggests that small differences in encoding and visual scanning can account for the differences found in the angle strategy. These findings suggest that mental rotation may not provide a full account of human performance in orientation tasks.

Acknowledgements

The research reported here was supported by grant number F49620-99-1-0086 from AFOSR.

References

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

Byrne, M. D. & Anderson, J. R. (1998). Perception and action. In J. R. Anderson, & C. Lebiere (Eds.). *The atomic components of thought* (167-200). Mahwah, NJ: Lawrence Erlbaum.

Cecala, A. J., & Garner, W. R. (1986). Internal frame of reference as a determinant of the oblique effect. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 314-323.

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

Gugerty, L., deBoom, D., Jenkins, J. C., & Morley, R. (2000). Keeping north in mind: How navigators reason about cardinal directions. In *Proceedings of the Human Factors and Ergonomics Society 2000 Congress* (pp. I148-I151). Santa Monica, CA: Human Factors and Ergonomics Society.

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

Huttenlocher, J., & Presson, C. C. (1979). The coding and transformation of spatial information. *Cognitive Psychology, 11,* 375-394.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science, 18,* 513-549.

Presson, C. C. (1982). Strategies in spatial reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 243-251.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three dimensional objects. *Science, 171,* 701-703.

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