The Effects of Communication Medium Upon Collaborative Orientation Task Performance

Laura M. D'Andrea (dandrea1@illinois.edu) & Wai-Tat Fu (wfu@illinois.edu)
Applied Cognitive Science Lab, Human Factors Division & Beckman Institute
405 N. Mathews Ave., Urbana, IL 61801

Abstract

Pairs of dispersed individuals are often forced to solve orientation tasks collaboratively. The present study examines how collaborative orientation tasks are solved when pairs of individuals complete the tasks using one of three computer-based communications (text, audio, and video). Both simple and complex tasks were presented. Pairs in the audio condition outperformed those in the text and video condition overall, and specifically on complex tasks, despite the fact that the video condition allows for the greatest amount of information to be communicated. Analysis of conversations between pairs indicates that those in the video condition had different conversational behavior. Results suggest that social effects of video communication may impair collaborative orientation task performance.

Keywords: Spatial Cognition; Interactive Behavior

Introduction

An understanding of spatial relationships is critical for successful interaction with the world around us, impacting our ability to complete tasks as simple as reaching for a pencil and complex as maneuvering an environment (Taylor & Tversky, 1996). The ability to orient oneself in an environment with the use of a navigational aid, such as a map, is a particularly interesting spatial task. This situation demands that the map-reader's personal perception/view of the environment be aligned with the map's view, a frame of the same environment from a perspective entirely independent of, and depicting locations external to, the reader (Gunzelmann, Anderson, & Douglass, 2004; Klatzky, 1998). These differing perspectives of the same environment are referred to, respectively, as egocentric and allocentric frames of reference (Klatzky, 1998). To orient oneself in an environment, one must recognize how the two frames of reference correspond and depict the same environment; in other words, the reference frames must be aligned. The act of aligning reference frames likely requires somewhat more complex processing than a mental rotation of the two perspectives, because the frames are two distinct formats of information (egocentric vs. allocentric) (Gunzelmann, Anderson & Douglass, 2004). Though various processing strategies are used to mentally coordinate the different perspectives of a scene, including array rotation and viewer rotation, all accomplish orientation within the environment through the same overall strategy - alignment of the differing reference frame types (Gunzelmann, Anderson & Douglass, 2004).

The current ubiquity of communication-oriented technologies has, in some ways, added complexity to the process of orienting oneself in an environment. The aforementioned work on orientation was done with regard to a single individual. However, a lost driver can now easily call a friend for directions rather than look at a map. An astronaut repairing a broken device in space can receive instruction from ground control on how to repair it if the instruction manual is outdated. In these and countless similar situations the orientation task is distributed across multiple geographically distributed individuals, each with information that is crucial to solving the task but insufficient on its own, and each with a different frame of reference.

Disparate, communicating individuals presumably must orient themselves in the same general manner as is done by an individual – by aligning the available egocentric and allocentric reference frames (Gunzelmann, Anderson, & Douglass, 2004). In distributed orientation tasks, however, reference frames cannot be aligned by examining the egocentric and allocentric frames and physically aligning them (as would be possible if an individual were lost and had a map in hand). Therefore, it seems that communicative partners can only overcome the disparity in their reference frames by actively discussing pertinent spatial relationships within the environment, until they are able to align each other's perspectives. The role of communication in distributed collaborative orientation tasks is critical, and the fact that the individuals are not co-located introduces challenges to their ability to effectively communicate. For instance, compared to face-to-face collaborators, dispersed collaborators have been found to often have different understandings of the information/task at hand and of the meaning of their partner's actions (e.g., silence during a conversation), and show a reduced ability to establish and maintain a common understanding of each others' knowledge of the situation or task at hand (Cramton, 2001; Diamant, Fussell, & Lo, 2008). A major contributing factor to these challenges is the inability of computer-mediated communication tools to allow for the same access to social and contextual cues that are visible during face-to-face interaction (Cramton, 2001; Diament, Fusell, & Lo, 2008).

The type of communicative technology being used also introduces potential issues. By nature, different types of computer-mediated technologies (e.g., audio conference, video conference, text communication) convey different levels of cues about one's partner, and can differentially impact how an individual feels about his partner and their task performance (Diamant, Fussell, & Lo, 2008). Diament, Fussell, & Lo (2008) evaluated the impact of three communication mediums – Text, Audio, and Video – to examine in relation to one another, and found that technology type interacted with the culture of individuals to predict their attributions of performance (Diament, Fussell, & Lo, 2008). Diament, Fussell & Lo's (2008) findings indicate that the affordances of a technology determine the way in which it influences attributions of performance. Perhaps the affordances of those technologies also have differential impacts upon how individuals work together to complete collaborative orientation tasks, tasks in which communication is critical to solving the task. We expect the video condition to generate the best performance, the audio to allow the second best, and the text to result in the worst. This prediction is based upon the amount of information that each communication type provide (i.e., the video condition allowing individuals to not only speak but also use gesture to help describe spatial relationships within their view of an environment. We tested this possibility with an experiment studying the impact of different communication mediums upon collaborative orientation task performance.

Method

Overview

A collaborative orientation task is a type of spatial task that can only be solved when multiple individuals work together, combining their knowledge to deduce the solution. The impact of communication medium type upon collaborative orientation task performance was studied by requiring pairs of individuals to work together to solve spatial tasks while communicating through one of three communication mediums: text, audio, or video chat. Each of our collaborative orientation tasks included two unique displays of task-relevant information, one for either participant in the pair. Both displays contained solution-critical information, and the task demanded that pair members communicate their information in order to ultimately deduce the cardinal direction of the target.

Participants

Participants were 48 adults over the age of 18, recruited from Champaign, IL and paid for their participation. Participants were randomly assigned to a pair. One pair of participants (Audio condition) failed to perform the tasks, and their data was not included in the analysis. Of the remaining 46 adults (mean age=24.6; mean years of education=15.7), 29 were female and 17 were male. Participants were screened for normal or corrected-to-normal vision.

Measures

Spatial abilities were measured with two paper-pencil tasks. Participants' ability to mentally rotate objects was measured with the Mental Rotation Test (MRT) (Vandenberg and Kuse, 1978). Perspective taking ability (i.e., the ability to imagine how a scene looks from a different location in space) was assessed with the Perspective Taking/Spatial Orientation Task (PTSOT) (Hegarty & Waller, 2004; Kozhevnikov & Hegarty 2001).

Collaborative Orientation Task Stimuli

Our stimuli were adapted from those used by Gunzelmann, Anderson, & Douglass (2004). In their study, a single task contained two separate displays of information that needed to be reconciled to solve the task; individuals completed the tasks alone. In our study, a single task contained the two separate displays of information. However, we gave only one display to either member of the pair (one for the Responder and one for the Instructor). We also slightly modified the appearance of the tasks. The Responder was presented with a 2D array of seven images and one target icon, all located in one of the eight cardinal directions (*North, South, East, West, Northeast, Northwest, Southeast, Southwest*). The Instructor was given a display showing two of the seven images seen by the Responder, as well as an arrow indicating North (relative to the center of their screen). In either pair members' display, the icons maintained identical spatial relationships with other icons. However, the entire array was rotated to some degree (rotations of 90° increments), so the Instructor and Responder's displays were not identical. See Figure 1 for examples of each display, as they would appear in an actual task.

For each task, pairs' goal was to deduce the cardinal direction in which the target was located, relative to the X in the center of the Responder's screen. The Responder was ultimately responsible for reporting the direction of the target. Because the Responder was given information regarding the target's location relative to other images, and the Instructor was given the cardinal directions of certain images, pairs needed to discuss their displays (e.g., images, directions of images, relationships between images, etc.) in order to align their perspectives of the displays and deduce the target's direction. Stimuli of two levels of complexity (simple, complex) were displayed. In simple tasks, each icon on the Responder's display was unique. In complex tasks, the Responder's display contained multiples of the two icons that were present in the Instructor's display.

Figure 1. Sample trial displays for a simple task. The left is a display seen by a Responder; the right is a display seen by an Instructor (correct response=Southwest).

Procedure

Within each pair, individuals were randomly assigned to their roles (Responder vs Instructor). Pairs were randomly assigned to one of the three communication conditions. 8

pairs participated in each condition; however, one pair in the Audio condition was not included in analysis due to a failure to perform the tasks. The orientation task portion of the experiment was conducted with both pair members present in the same room, but seated at computers separated by enough distance/barrier so that participants were out of each other's sight and hearing range during the task. Before beginning the collaborative orientation tasks, individuals completed demographic and spatial tasks. They were then shown to their computers and instructed as to what their communication medium would be. On each computer, an instruction screen was presented which explained the tasks and offered a sample display representative of what each individual would view, depending upon their role (Responder vs. Instructor). The pair then performed one practice task before beginning a set of 20 tasks, 10 complex and 10 simple. In each condition, the task workspace took up half of the computer screen; the other half contained the communication tool (*Text condition:* an IM chat box; *Video:* Skype video chat interface (see Figure 2); *Audio*: Skype audio chat interface). Accuracy of task performance and conversations between pair members were recorded.

Across all pairs, the practice trial was identical. However, each of the 20 actual trials was randomly generated for each pair. This randomization included: the 7 icons that appeared on the Responder's display (randomly selected from a master set of 18 icons); the target's location on the Responder's display; the direction of North on the Instructor's display; the relative locations of the two icons appearing on both the Instructor's and Responder's displays; the degree of disparity between the Instructor's and Responder's displays (90 increments); the distribution of complex/simple tasks throughout the 20 overall tasks.

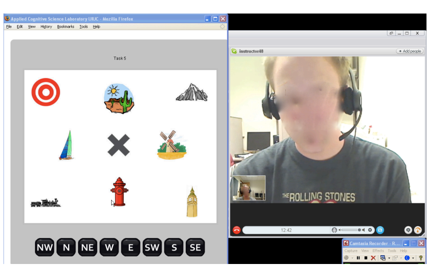


Figure 2. Screenshot from Video condition (Responder's computer screen)

Equipment

Stimuli were presented on each participant's computer screen; responses were made using the mouse. Camtasia screen capture software was used to record audio and video feeds in the Audio and Video conditions.

In the Text condition, participants communicated by typing to each other using AOL Instant Messenger. In the Audio condition, participants wore Logitech ClearChat headphones (with microphone) when performing the tasks. Auditory communication was enabled through the use of Skype's auditory calling feature. In the Video condition, participants wore Logitech ClearChat headphones (with microphone) when performing the task, as there is an auditory component to video chatting. Video chat communication was enabled through the use of Skype's video chat feature. Each computer was supplemented with Logitech Webcam Pro 9000 cameras in order to permit video chatting.

Results

Performance

We performed a two-way ANOVA examining the effect of communication media (video, audio, text) and task difficulty (simple, complex) on overall accuracies. There main effect of communication media was not significant ($p=0.11$), but the main effect of task complexity was significant ($F[1,21]=4.22$, $p<0.05$). There was also a significant communication by complexity interaction ($F[2,21]=3.55$, $p<0.05$). Simple effect analysis between media conditions in simple tasks showed no significant difference, but the difference was significant in complex tasks ($F(2,21)=5.86$, $p<0.05$). Posthoc tests (Fisher's LSD) showed that the audio condition was significantly better than the text condition ($t(6)=4.8$, $p<0.01$) and the video condition ($t(6)=3.1$, $p<0.01$), but the difference between text and video was not significant (see Figure 3). This demonstrated that pairs in the Audio condition performed the tasks better than pairs in either the Video or Text conditions only in the complex tasks.

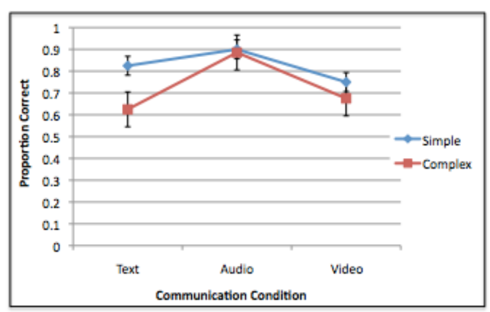


Figure 3. Average proportion of trials correct by communication medium, as a function of trial complexity.

To assess whether these differences in task performance was linked to pairs' spatial aptitude, rather than the communication medium being used, spatial ability task scores were analyzed. Results indicate no significant difference in the abilities of individuals in the three communication conditions. A one-way ANOVA examining MRT scores of individuals in the different conditions (video, audio, text) revealed no significant difference in the MRT scores of individuals in the different communication conditions ($F[2,43]=0.70$, $p=0.50$). Similarly, a one-way ANOVA examining PTSOT scores across conditions (video, audio, text) revealed no significant difference in the PTSOT scores of individuals in the different communication conditions (F[2,41]=0.11, p=0.90).

We also examined the relationship between a pair's average spatial ability score and their collaborative orientation task performance. For each pair, a single score was generated for both the MRT and PSOT by averaging the scores of the two individuals within the pair. This score (denoted with "-P") was then correlated with task performance. The MRT-P score was significantly correlated with task performance in the Text condition (r=0.86, p<0.01, df=6) and the Video condition (r=0.77, p<0.05, df=6). However, MRT-P was not correlated with performance in the Audio condition (r= -0.38, p=0.40, df=5). The same trend followed with regard to the PTSOT-P scores. PTSOT-P was significantly correlated with performance in the Text (r= -0.87, p<0.01, df=6) and was marginally correlated with performance in the Video condition (r= -0.68, p=0.09, df=5), but was not correlated with performance in the Audio condition (r= -0.41, p=0.36, df=5). The lack of correlation between spatial ability and performance within the Audio condition may be due to the restricted range of performance observed within this group. Additionally, averaging spatial ability scores is not an optimal approach to examining abilities across conditions, as an average score can obscure potentially interesting information (e.g., relative abilities of the Responder and Instructor in each pair).

The effect of practice on task performance was also assessed. Average scores for each trial were correlated with trial number. Overall performance on the collaborative orientation tasks, regardless of communication medium, was significantly correlated with trial number (r=0.59, p<0.01). Trial number was not significantly correlated with performance in the Text condition (r=1.32, p=0.20). It was marginally correlated with performance in the Audio condition (r=1.98, p=0.06). Performance of pairs in the Video condition, was significantly correlated with trial number (r=2.46, p<0.05). This finding suggests differential effects of practice depending upon the communication medium being used; thus, the communication mediums, rather than the task itself, are impacting whether practice improves performance.

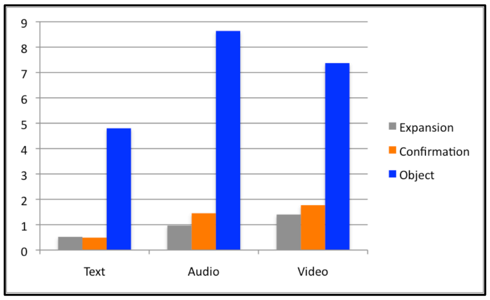
Conversational Analysis

To investigate underlying factors as to why the Audio condition allowed for superior performance, the conversations between each pair were transcribed and coded. A coding scheme was developed post-hoc and addressed four main categories of Utterance Type: Object Description, Revision/Repair, Request for Confirmation, and Request for Expansion. These types are rooted in Clark and Wilkes-Gibbs's (1986) work regarding the ways in which pairs of individuals, during conversation, collaborate to reach agreement on the noun phrase being referred to (the noun phrase being crucial to understanding what each other is trying to communicate).

Within each of these main Utterance Type categories were sub-categories regarding the contents of the utterance. In turn, each of these Utterance Content categories contained more-specific Statement Type categories. Each utterance was coded with regard to the main Utterance Type (e.g., revision/repair, request for expansion), and with respect to the different specific Statement Types within utterance content categories A, B, and C. Each utterance could receive multiple categorizations.

We were interested in whether there were differences across communication conditions in their use of the main Utterance Types, as well as whether the use of different main Utterance Types related directly to task performance. Therefore, a 4 (utterances types) X 3 (communication media) X 2 (correctness of response) ANOVAs with average frequencies of occurrences of utterances as dependent variable was conducted. Occurrences of each utterance type were normalized, taking the frequency of utterance in proportion to the number of trials that had occurred. Results indicated a significant three-way interaction (F(6,60)=4.33, p<0.001), a significant two-way interaction between correctness and media (F(2,20)=4.39, p<0.05), and significant main effects of utterance types (F(3,20)=43.24, p<0.001) and correctness (F(1,20)=20.42, p<0.001). Given that we observed interactions between media and correctness, the significant 2-way interaction and main effects were likely caused by the different number of correct and incorrect trials in each medium. We therefore focus on the further analyzing the 3-way interaction.

We performed separate 3 (media) x 2 (correctness) ANOVAs on each type of utterances. Results showed significant main effect of media for requests for expansion (F(2,20)=4.17, p<0.05). Post-hoc comparisons showed that Video had significant more requests for expansion than audio (t(7)=2.87, p<0.05) and text (t(7)=3.00, p<0.05) conditions. We found significant main effects of correctness and media for requests for confirmation (F(1,20)=16.15, p<0.05) and F(2,20)=3.67, p<0.05 respectively). Posthoc comparisons showed that Video had significant more requests for confirmation than text (t(7)=3.25, p<0.05). We found significant main effects of correctness and significant interaction between correctness and media for object description (F(1,20)=23.1, p<0.01) and F(2,20)=4.7, p<0.05 respectively). Posthoc comparisons showed that *only in incorrect trials,* Video had significant more requests for confirmation than audio (t(7)=2.47, p<0.05). There was no significant difference in any of the variables for Revision/repair. [See Figure 4].

4a)


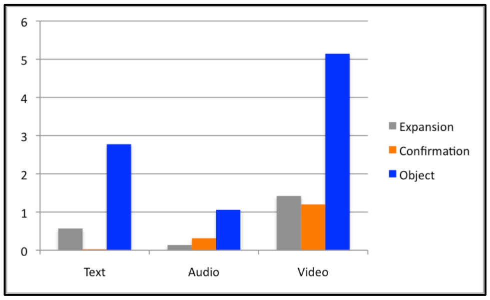
4b)


Figure 4. In each communication media, average utterance type per trial: 4a) Correct trials; 4b) Incorrect trials.

The Expansion and Confirmation request types are indicative of unsatisfactory or insufficient information communication between partners, indicating individuals' need/desire to gain more information from their communicative partner (Clark and Wilkes-Gibbs, 1986). Intuitively, one might think that the video condition would result in the fewest requests because it allows for greater amount of information to be communicated. However, participants in the video condition made more requests for expansion than individuals in text or audio, and more requests for confirmation than participants in text or audio (when incorrect performance resulted), indicating that the information communicated was more frequently deemed insufficient by pair members in the video condition than in either text or audio.

Discussion

This study's results were interesting in that, contrary to our predictions, the video condition did not induce the highest task performance level, but was in fact worse than audio and on the same level as text. The spatial abilities of individuals in each communication condition did not differ, indicating that our performance findings did not result from an overload in high or low ability individuals within specific conditions. Further research is needed to uncover the reasons behind observed performance trends, as our study was not geared to investigate several of the underlying factors that may have contributed to the performance differences. For instance, in the future we may match participants on spatial ability across conditions so as to better account for the abilities of pairs (rather than individuals) within conditions, and perhaps examine the impact of the relative ability of each pair member upon performance. However, current findings suggest an interesting social influence on cognition. The remainder of this discussion focuses upon tying together our performance measure and conversational analysis results.

We found evidence that pairs' performance on collaborative orientation tasks is impacted by the type of communication medium used during task solving. Contrary to our expectation that the Video condition would result in the highest level of task performance, pairs in the Audio condition outperformed pairs in the Video and Text conditions both overall and on complex tasks, while performance of simple tasks was no different across communication mediums. So, it appears that all three communication mediums allowed for good performance on easier tasks, but some aspect of the auditory communication medium allows its users to sustain their performance level when tasks increase in complexity. In addition, the finding that the joint measure of pairs' spatial ability (MRT-P, PTSOT-P) was not correlated with task performance in the audio condition, but was in text and video, suggests that some factor inherent to the audio communication medium was at the root of its optimality for solving collaborative orientation tasks.

Our conversational analysis, aimed towards investigating why pairs in the other communication mediums did not show the same performance, revealed differences in how pairs in different communication conditions actually communicated information to each other. Two of the main utterance types – Requests for Expansion and for Confirmation – were used more by pairs in the Video condition than they were by pairs in both Audio or Text. Before moving forth with the discussion of our findings, it is important to reiterate the purpose of these types of utterances. Clark and Wilkes-Gibbs (1986) explain that when two people are speaking, over the course of the conversation one of them will utter a statement (specifically, a noun phrase) that their listening partner deems unacceptable or inadequate. The unacceptability could occur because the listener needs more of a description to understand what their partner is referring to (request for expansion), or because they want to clarify that what they heard is correct (request for confirmation) (Clark and Wilkes-Gibbs, 1986).

Although task performance overall and on complex tasks was significantly better in the audio condition, overall communicative behavior was essentially the same in the audio and text conditions. This suggests that the there were factors inherent to the textual condition that impacted task performance without effecting how pairs actually worked together. Previous research indicates that textual communication is simply more difficult for pairs to work with, which could be the case here. Cramton (2001) discusses how various traits of text-based communication, including the slower rate of information exchange and the demand to communicate typically non-verbal cues with words (i.e., saying 'yes' instead of nodding), impede performance in text communication mediums. Text-based systems do not provide significant feedback, like verbalizing 'yeah' or 'mmhm' to indicate understanding, and thereby imposes on pairs' ability to develop a shared knowledge of the situation (Cramton, 2001). Our text condition certainly presented these issues, which could have been the root of the resulting poorer performance. Another possibility is that the processing/resource demand of the text condition was higher than in audio, and while pairs could overcome the text condition's inherent difficulties in simpler tasks and they were unable to do so in more complex tasks. If this were the case it would follow predictions of theory on the impact of resources on multiple-task performance (Wickens, C. D., 1991). However, our study did not specifically examine any of these factors; therefore, a conclusion regarding the poorer performance in the text condition cannot be reached.

Performance in the video condition was also significantly poorer than that in the audio condition. The video condition *did* allow for some amount of Cramton's (2001) described non-verbal feedback so a lack of social and verbal cues cannot be entirely blamed for performance (though video communication is still deficient when compared to face-to-face; Cramton, 2001). It is possible that performance in the video and text conditions were both rooted in some cognitive/attentional load issue; however, the load induced by the video condition appears much higher than that of the text (as a large video feed of another individual was in close proximity to the task, compared to a text message box). If it were cognitive demand that decreased performance, one would expect the video condition to have experienced a more significant impact. And if the processing/resource demand had affected conversational behavior, one would expect text and video conditions to have communicated in the same manner. But, only the video condition incited pairs to use more requests for expansion and confirmation during their conversations. The difference in conversational behavior, and more specifically in the types that were differing (requests for confirmation /expansion, both statement types that indicate inadequacy of initial communication/a need to confirm what was said (Clark and Wilkes-Gibbs, 1986), suggesting that the video condition spurs a sub-optimal communicative behavior between partners. The video condition's poor performance levels indicate that this behavior was detrimental in some manner to task performance. The root of this behavior could lie in social effects imposed by the fact that video condition participants could see each other while communicating. For instance, perhaps individuals unfamiliar with each other restrict their display descriptions due to some discomfort felt by knowing that a stranger is watching them while they think. Additional research is needed to further examine the mechanism driving performance in different communication media, as we did not aim to examine such social influences. Follow-up work should include larger sample size and more trials, to better examine the effects of practice in varying conditions, and the aforementioned control of pairs' spatial abilities.

Conclusion

We found that audio communication allowed users to maintain a high level of performance on collaborative orientation tasks of varying complexities. Text and video communication mediums made the tasks more difficult to perform. In the case of video communication, this decrease in performance is likely tied to the style of conversational behavior incited by the communication medium. More research is needed regarding both the cause of decrease in performance observed in textual communication, and the reason behind the shift in conversational behavior observed in the video condition.

Acknowledgements

We thank Miriam Holtzman and Namrata Donthamsetti for help running the experiment and preparing data for analysis.

References

Clark, H. H., Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* (1-39).

Cramton, C. D. (2001). The Mutual Knowledge Problem and Its Consequences for Dispersed Collaboration. *Organizational Science, 12*(3) (346-371).

Diamant, E. I., Fussell, S. R., & Lo, F. (2008). Where did we turn wrong? Unpacking the effects of culture and technology on attributions of team performance. *Proceedings of CSCW 2008.*

Gunzelmann, G., Anderson, J. R., Douglass, S. (2004). Orientation Tasks with Multiple Views of Space: Strategies and Performance. *Spatial Cognition and Communication, 4*(3), 207-253.

Hegarty, M. & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence, 32* 175-191.

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.

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object-manipulation spatial ability and spatial orientation ability. *Memory and Cognition, 29,* 745–756.

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

Taylor, H.A. & Tversky, B. (1996). Perspective in Spatial Descriptions. *Journal of Memory and Language, 35*, 371-391.

Vandenberg S.G., Kuse, A.R. (1978). Mental rotations, a group test of three-dimensional mental rotation. *Perceptual and Motor Skills (47)*2, 599-604.

Wickens, C. D. (2001). Processing resources and attention. In Damos, D. L. *Multiple-task performance.* (pp. 3-34). Taylor & Francis, Inc., PA: Bristol.