Metacognitive Aspects of Gender Differences in Spatial Navigation

Chantal Lemieux

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfilling of the requirements
For the PhD degree in Experimental Psychology

School of Psychology
Faculty of Social Sciences
University of Ottawa

ACKNOWLEDGMENTS

My success through graduate school and the final outcome of this project required a lot of guidance and support from many people. First and foremost, I consider myself privileged to have had the chance to work under the supervision of Dr. Charles Collin. Thank you, Charles, for giving me the freedom to explore my curiosity and to develop my own work ethic, as well as the guidance and support I needed when I was lost. Through your incredible mentorship, you have ultimately shaped the researcher and person that I have become, and for this, I am eternally grateful. I could not have navigated my way through without you.

Besides my supervisor, I would also like to thank the rest of my thesis committee members: Dr. Andra Smith, Dr. Cary Kogan, and Dr. Ken Campbell, who have provided me with insightful and constructive feedback from various perspectives. Thank you to my external committee member, Guy Lacroix, for the thorough review he did on my thesis and for the insightful feedback during my defense. Thank you all for the care and hard work you put into my thesis.

I would also like to thank all my fellow labmates: Beth, Dhrasti, Heather, and Laura, for their help as well as friendship. I am especially grateful for the help and support from Dr. Nick Watier, who continued to mentor me after he graduated and helped to review my work.

Finally, I would not have been able to accomplish my dreams without the support and encouragement from my family and friends. Thank you for listening to and humouring my know-it-all ramblings. I am and will continue to be grateful for all the ways you have each enriched my life.

TABLE OF CONTENTS

ABSTRACT .... **V**

INTRODUCTION **1**

CHAPTER 1: SPATIAL NAVIGATION **5**

THE HISTORY OF RESEARCH IN SPATIAL NAVIGATION 5
    Cognitive Maps *5*
    Spatial Knowledge *6*
    Navigational Strategies and Frames of Reference *9*
    Convergence with Neuroscience and Animal Models *11*
REAL-WORLD VS VIRTUAL NAVIGATION 16
    First-Person Virtual Navigation Tasks *18*

CHAPTER 2: METACOGNITION **24**

THE HISTORY OF RESEARCH IN METACOGNITION 24
    Nelson and Narens (1990) Metamemory Framework *26*
MECHANISMS UNDERLYING METACOGNITIVE MONITORING JUDGMENTS 31
    Ease-Of-Learning Judgments *32*
    Judgments-of-Learning *33*
    Retrospective Confidence Judgments *36*
RATIONALE FOR APPLYING A METACOGNITIVE FRAMEWORK TO SPATIAL NAVIGATION 37
    Metacognition in Spatial Navigation *39*
CHAPTER SUMMARY 42

CHAPTER 3: GENDER DIFFERENCES IN SPATIAL NAVIGATION **43**

THE HISTORY OF RESEARCH ON GENDER DIFFERENCES IN SPATIAL COGNITION 43
GENDER DIFFERENCES IN SPATIAL NAVIGATION 47
    Factors Underlying Gender Differences in Spatial Navigation *52*
GENDER DIFFERENCES IN SPATIAL CONFIDENCE AND METACOGNITON 59
    Evidence from the Stereotype Threat Literature *64*
CHAPTER SUMMARY 67

CHAPTER 4: OVERVIEW OF EXPERIMENTS **69**

GENERAL PURPOSE 69
GENERAL PROCEDURE 70
GENERAL HYPOTHESES 72

CHAPTER 5: EXPERIMENT 1 .... **75**

GENDER DIFFERENCES IN METACOGNITIVE JUDGMENTS AND PERFORMANCE ON A FIRST-PERSON VIRTUAL MAZE LEARNING TASK 75
    Methods *83*
    Results *89*
    Discussion *96*
    References *102*

CHAPTER 6: EXPERIMENT 2 113

GENDER DIFFERENCES IN JUDGMENTS OF LEARNING AND RETROSPECTIVE CONFIDENCE IN A FIRST-PERSON VIRTUAL MAZE LEARNING TASK.. 113

Methods 123
Results.. 129
Discussion.. 134
References.. 144

CHAPTER 7: EXPERIMENT 3 157

EFFECTS OF STEREOTYPE ACTIVATION ON GENDER DIFFERENCES IN METACOGNITIVE ACCURACY AND NAVIGATION PERFORMANCE.. 157

Methods 168
Results.. 176
Discussion.. 187
References.. 194

CHAPTER 8: GENERAL DISCUSSION 208

CROSS STUDY COMPARISON TO METACOGNITION LITERATURE 208
GENDER DIFFERENCES IN NAVIGATION PERFORMANCE AND CONFIDENCE 213
GENDER DIFFERENCES IN LOCAL METACOGNITION... 217
THE EFFECTS OF STEREOTYPE ACTIVATION 222
STUDY LIMITATIONS.. 225
PRACTICAL IMPLICATIONS OF THIS RESEARCH 229
CONCLUSIONS.. 232

REFERENCES.. 235

APPENDIX A: SANTA BARBARA SENSE-OF-DIRECTION SCALE (SBSOD)........... 296

APPENDIX B: SPATIAL ANXIETY SCALE (SAS).. 298

ABSTRACT

Many studies have shown a gender difference in spatial navigation ability, including a related gender difference in global metacognitive self-assessment and spatial anxiety. However, it has yet to be determined whether trial-by-trial metacognitive accuracy differs between the genders and how this may be related to gender differences in navigation performance. The goal of this research was to determine, using the Nelson and Nerens (1990) metamemory framework, if there exist gender differences in trial-by-trial metacognitive monitoring on a first-person virtual maze navigation task, and how this may be related to gender differences in navigation performance. Considering that there is a relatively pervasive stereotype that women have poor navigation stills, an additional goal of this research was to determine if the effects of stereotype could, at least partially, explain the gender difference in navigation performance, confidence, and trial-by-trial metacognitive monitoring accuracy. Many studies have shown stereotype threat and lift to influence confidence and performance between the genders on a variety of spatial cognitive tasks, but mostly on mental rotation tasks. We investigated whether this effect applied to gender differences in a spatial navigation task. In order to accomplish this, we assessed trial-by-trial metacognitive accuracy during a first-person virtual maze navigation task under three stereotype facilitation conditions where participants were told that either: 1) men outperform women on this particular task, 2) women outperform men on this particular task, or 3) the genders perform equally. Over three experiments, the results showed that men generally have more accurate metacognitive monitoring than women, especially when assessing a previous performance. Contrary to our expectations, stereotype activation had no effect on trial-by-trial metacognition, though it did have an effect on navigation performance and confidence.

INTRODUCTION

Navigation research has been a central part of experimental psychology for much of its history. This is so much the case that the image of a rat in a maze is often held up as the quintessential image of the discipline. Despite an abundance of findings regarding navigation, generated over many decades, there remains much to be uncovered about this complex cognitive function. Notably, while research has been conducted on global confidence and anxiety effects in navigation, there has yet to be any work relating more fine-grained trial-by-trial metacognitive judgments to this task. The purpose of this dissertation is to examine this more granular aspect of metacognition as it pertains to the field of navigation. This will be done by adapting a metacognitive protocol, based on the Nelson and Narens (1990) metamemory framework, to a spatial navigation task.

While metacognition is also a very fundamental psychological function, the study of it is relatively new, at least compared to navigation research. Indeed, until relatively recently, only a few vague theoretical frameworks existed to explain it, and the existing research community acknowledges the need to extend the field into other areas of cognition (Diana & Reder, 2004; Nelson & Narens, 1994). It is for this reason we have decided to adapt Nelson and Narens' (1990) model to our navigation tasks, as it is the most empirically established and cogent theoretical framework regarding metacognition.

One factor in navigation performance that has long been widely studied is gender. While the gender gap in performance on many cognitive functions has been eliminated over time (for reviews, see Coluccia & Louse, 2004; Feingold, 1998; Hyde, 1981; Martens & Antonenko, 2012; Voyer, Voyer, & Bryden, 1995) or at least narrowed (Hyde, 2005), spatial navigation tasks still show strong differences between the genders (Feingold, 1998; Voyer et al., 1995; Voyer, Voyer,

& Saint-Aubin, 2017). There are studies that have shown gender differences in global metacognitive confidence favoring men (e.g., Hertzog, Dixon, & Hultsch, 1990); though very little research has investigated gender differences in trial-by-trial metacognitive accuracy. One aim of the present thesis work is to verify, using a virtual spatial navigation task, whether gender differences in performance continue to exist, and to discuss how this relates to the current literature on gender differences in spatial cognitive tasks. Another aim is to determine if gender differences exist in trial-by-trial metacognitive accuracy, and if so how this may be related to gender differences in performance.

Another important goal of the present thesis work is to examine whether trial-by-trial metacognitive accuracy, confidence and navigation performance are affected by stereotype threat/lift, also known as stereotype activation. Stereotype threat, a term originally coined by Steele and Aronson (1995), refers to a phenomenon whereby reminding participants of a stereotype causes the individuals who are negatively regarded in the stereotype to perform more poorly than they otherwise would have had the stereotype not been mentioned. For instance, telling or insinuating to participants that women are poor drivers can induce female participants to drive more poorly than females in control conditions (e.g., Derks, Scheepers, Van Laar, & Ellemers, 2011; Yeung & von Hippel, 2008). Steele and Aronson (1995) proposed that this works by putting pressure on the individual to represent their social group which induces anxiety, causing attentional deficits that lead these individuals to objectively perform worse than if they were not reminded of the stereotype. Several studies have shown that by alleviating the effects of stereotype threat, the gender gap in a variety of tasks is reduced (for a review, see Spencer, Logel, & Davies, 2016). One goal of the current study is to determine if the gender gap in metacognitive accuracy, confidence, and spatial navigation performance can be reduced by alleviating such effects. The findings from the current thesis also have the potential to add to the body of research supporting the effects of stereotype threat on spatial tasks, by expanding this effect to navigation-related tasks and stereotypes.

A final goal of this research is to initiate the development of methodologies to facilitate the study of trial-by-trial metacognition as it pertains to spatial navigation by applying the Nelson and Nerens (1990) metamemory framework. As is the case with work in stereotype threat, the contribution here is to expand previous work on metacognitive accuracy to a wider range of cognitive tasks. Previous work in this area has largely centered on word-pair association learning, and there have been calls to expand it to other types of stimuli and other tasks. Among other things, this will allow us to examine whether the Nelson and Narens framework applies to higher-order cognitive tasks, as opposed to simple paired-associate learning.

This dissertation begins with the history of research in spatial navigation, with specific attention paid to first-person virtual navigation tasks. This is done in order to establish a historical context for the research that follows. Next, the history of research in metacognition will be presented, with specific attention paid to Nelson and Narens' (1990) framework for metamemory. This is followed by a discussion as to the practical and theoretical significance of applying metacognitive research to the field of spatial navigation. Once the state of these two fields of research has been established, a more specific review of the scientific literature on gender differences in spatial navigation, metacognition, and confidence will be presented. The third chapter will end with a discussion as to how these gender differences may be related to stereotype threat.

The fourth chapter presents the rationale for the proposed research as well as the general protocol of three experiments designed to empirically answer the questions outlined above. The first two experiments are designed to examine if, when applying the Nelson and Nerens (1990) metamemory framework to our navigation task, we indeed see gender differences in navigation performance and/or trial-by-trial metacognitive accuracy. The third experiment involves using a stereotype facilitation procedure in order to determine if stereotype affects the gender difference in metacognitive accuracy and navigational performance. The final chapter of this dissertation includes discussions of the theoretical and practical implications of gender differences in navigation performance and trial-by-trial metacognition, and of the effects of stereotype activation in this novel context.

CHAPTER 1: SPATIAL NAVIGATION

THE HISTORY OF RESEARCH IN SPATIAL NAVIGATION

Navigation is a complex skill requiring the coordination of multiple cognitive processes. To understand it, we need to understand each cognitive sub-process and its role in the larger navigation system. A vast amount of research has been undertaken in an attempt to do this. However, the various models that have been developed to guide and organize this abundance of findings are relatively broad and general, and take a wide variety of perspectives. Moreover, there are varying degrees of overlap between one model and the next. In order to establish the historical context of my proposed work, I describe below the three most predominant navigation frameworks: Cognitive Mapping, Spatial Knowledge, and Navigational Strategies and Perspectives. The relevance of my own work to these models will be discussed in later sections. In order to provide a more complete foundation for the current research, neuroscientific findings as well as research pertaining specifically to first-person virtual navigation tasks will also be covered.

Cognitive Maps

One of the most widely referenced concepts in the literature on navigation is that of the *cognitive map*. Edward C. Tolman developed this concept in 1948, in an attempt to improve on the one-dimensional “stimulus-response” understanding of learning and navigation prevalent at the time. He developed the idea of cognitive maps in order to explain the results of his experiments involving rats, including some in which they navigated mazes using shortcuts. Tolman suggested that the rats could only do this by creating a cognitive map of the maze during the initial exploration phase of the experiment. Tolman argued that the ability to use novel shortcuts was evidence contrary to previous theories suggesting that non-human navigation was simply a result of stimulus-response associates, and suggested we view navigation as its own kind of learning (Tolman, 1948).

Since Tolman's time, cognitive maps have become almost synonymous with the simpler and broader concept of *spatial representations*. Generally, a “cognitive map” is defined as an internal mental representation of relative spatial information, or more explicitly the layout of a specific environment (Golledge & Spector, 1978; Waller & Nadel, 2013)[1]. It is important to note that most models of cognitive maps hold them to be incomplete and to routinely contain systematic distortions (Golledge & Spector, 1978; A. Stevens & Coupe, 1978; Tversky, 1992). Moreover, cognitive maps are thought to be sparse, in that they capture only certain classes of details in the environment depending on the individual, the strategy they are using, and the information available. For instance, individuals only given a map to learn a route perform better than those only given direct exposure to the environment to learn a route (Lloyd, 1989). This is thought to be because a map provides only minimal and relevant information about an environment, and thus may be a more efficient source of information for building an effective (i.e., sparse) cognitive map.

Spatial Knowledge

While the concept of cognitive maps has been influential, many researchers have preferred to work within the broader and more hierarchical theoretical context of "spatial

---

[1] Other researchers have defined cognitive maps in slightly different ways throughout the years (e.g., Gallistel, 1989; Thinus-Blanc, 1988).

knowledge". This has classically been categorized into three general types: landmark, route, and survey knowledge (Siegel & White, 1975). Landmarks are significant major features in the environment that can be the location of an associated action or can serve as a beacon to aim toward. The next level in spatial knowledge is “Route Knowledge”. It is usually defined by a series of place-action associations (i.e., a route), and it allows one to follow a path from one location to another (e.g., when giving directions). However, “Route Knowledge” does not require metrics or knowledge of the overall environment, only the information needed at specific decision points where it is necessary to change heading. The final and all-encompassing level is “Survey Knowledge” which includes information about the overall layout and how routes fit together with correct distances and angles independent of the navigator’s position. In other words, it is the kind of information contained in cognitive maps, perhaps like those described by Tolman in 1948 (McNamara, 2013).

A recent review of the literature on spatial knowledge has argued for the inclusion of “Graph Knowledge” in between Route and Survey Knowledge. Graph Knowledge represents decision points according to graph theory, without the need for metric distances like in Survey Knowledge (Chrastil, 2013). Thus, it allows the finding of new sequences of routes between landmarks. This is unlike Route Knowledge, which only represents one route. In addition, Graph Knowledge is highly influenced by the geometry of the environment, which many studies have found is information we are sensitive to (e.g., Hartley, Trinkler, & Burgess, 2004; Nardi, Newcombe, & Shipley, 2010; Schmidt & Lee, 2006; Shelton & McNamara, 2001). For instance, participants bias angles in their spatial learning towards right angles (Byrne, 1979), hinting that the person had encoded Graph Knowledge but not yet achieved more metrically-precise Survey

Knowledge. Indeed, most navigation tasks are likely assessing Graph Knowledge, as most tasks involve multiple possible routes with no need for metric distances.

Although the four forms of spatial knowledge seem to be part of a linear sequence (i.e., Landmark < Route < Graph < Survey), there is evidence that the order in which we learn them is different from one individual to another (Wolbers & Hegarty, 2010) and one task to another (e.g., Klatzky et al., 1990; Montello, Waller, Hegarty, & Richardson, 2004; Wiener, Buchner, & Holscher, 2009). Notably, Ishikawa and Montello (2006) found that a small proportion of participants had relatively accurate Survey Knowledge after only one tour of two connected routes, while only half of participants improved their Survey Knowledge at all over time. These results suggest that the sequence of learning Route and Survey information may not always manifest strictly the same for everyone, with some people prioritizing one over the other (Ishikawa & Montello, 2006). Moreover, Foo, Duchon, Warren, and Tarr (2007) found that despite having acquired survey knowledge their participants relied on erroneous landmark knowledge when taking a novel shortcut. This demonstrates that survey knowledge does not necessarily dominate lower forms of knowledge (Foo et al., 2007).

Many researchers have suggested the existence of “spatial profiles”, whereby some participants have a tendency to focus mainly on Route Knowledge while others use Survey Knowledge (Ishikawa & Montello, 2006; Nori & Giusberti, 2003). According to Nori and Giusberti (2003), survey-oriented participants do not show any *alignment effect*, unlike landmark- and route-oriented participants. The alignment effect is exhibited by better performance when the spatial perspective at test is the same as the spatial perspective at learning. Nori and Giusberti (2003) suggest that their results are likely due to survey-oriented participants representing spatial information in allocentric coordinates, which are independent of perspective.

These profiles may reflect differences in spatial strategies and perspectives, where a route learner uses a more egocentric point-of-view and survey learners use a more allocentric point-of-view. It is to this topic that we turn to next.

Navigational Strategies and Frames of Reference

It is widely held in the literature that there are two basic complementary processes for determining position and directional heading while navigating (Calton & Taube, 2009): *path-integration* and *landmark navigation*. Path integration, also known as dead-reckoning or self-orientation, is an egocentric process by which the subject's current position is estimated by performing a sequential integration of direction and speed of movement over time since the last known position (Gallistel, 1990; Mittelstaedt & Mittelstaedt, 1980). Mittelstaedt and Mittelstaedt (1980) provided some of the first documented evidence of the exclusive use of idiothetic cues (i.e., cues that are internal to the organism) in navigation in a mammal. They showed that a gerbil mother, upon finding her displaced pups, can take a direct route back to the nest in total darkness, and therefore in the absence of orienting landmarks. A second method of estimating one’s position is landmark navigation (also known as place recognition or piloting) which relies upon the presence of stable allothetic cues. These are cues external to the organism, like landmarks or geographical features (Gallistel, 1990). The Morris Water Maze is widely held to be the best example of a technique that emphasizes allocentric processing (Morris, 1984). The animal is placed in a pool of opaque water at random locations and must learn to locate the position of a hidden platform based on the established visual cues in the surrounding room. Landmark navigation can be more accurate than path integration, because it does not suffer from error accumulation; indeed positional errors can be continuously reassessed relative to new external cues (i.e., landmarks). However, its efficacy depends on the availability of salient and discriminable landmarks. These two forms of navigation—path integration and landmark navigation—are generally held to be complementary and are likely used in concert during every day navigation (Calton & Taube, 2009; Gallistel, 1990).

There is another theoretical perspective one encounters when examining the cognitive literature that is parallel to the landmarks versus dead-reckoning model. This involves a consideration of two complementary perspectives when approaching navigation called *egocentric* and *allocentric*. An egocentric perspective uses the self as the ultimate reference point and can involve an eye, head, or body-based coordinate system (McNamara, 2013). Conversely, an allocentric perspective uses external landmark cues as relative reference points, such as the Sun, the true North, or the features on the walls in a room (Wehner, Michel, & Antonsen, 1996). An allocentric perspective is often considered to involve a third-person point of view or frame of reference, such as one has while looking down at a map; conversely, an egocentric perspective is necessarily understood to involve a first-person point of view or frame of reference. Interestingly, men have been shown to prefer an allocentric strategy while women tend towards an egocentric strategy (for more on gendered strategy differences see Chapter 3). Several behavioral studies (e.g., Burgess, 2008; Mou, McNamara, Valiquette, & Rump, 2004; R. Wang & Spelke, 2002; Xiao, Mou, & McNamara, 2009) and neurophysiological studies (e.g., R. A. Andersen, Snyder, Bradley, & Xing, 1997; Matsumura et al., 1999; Snyder, Grieve, Brotchie, & Andersen, 1998) have concluded that egocentric and allocentric frames of reference are distinct but used in conjunction to represent the spatial structure of the environment.

Convergence with Neuroscience and Animal Models

As a complement to the cognitive literature, neuroscientific findings can be helpful in determining convergent and concurrent validity while also providing additional theoretical insights. Until the advent of non-invasive neuroimaging techniques, the majority of neuroscientific findings made use of animal models. In the context of animal spatial cognition research, the primate literature has mainly focused on tasks requiring the manipulation of spatial relationships within the personal space of the subject (e.g., Buneo & Andersen, 2006; Colby & Goldberg, 1999; Snyder et al., 1998). Rodent studies have tended to emphasize tasks on a much larger relative scale, addressing issues of whether an animal can accurately navigate from one location to another (e.g., Parron & Save, 2004; Save, Guazzelli, & Poucet, 2001; Save & Moghaddam, 1996). Although rodent studies in cognition have given us a vast amount of information about how creatures navigate, it is important to take their limitations into consideration, particularly with regards to translational comparability. It is safe to say that human navigation is special because they use navigational supports (e.g., maps, compasses) and navigational language, which likely involve far more spatial transformation processes. Animals also have different sensory strengths, which could change the information they use to navigate. Having said this, research does suggest that there are some similarities in how humans and other mammals navigate. For instance, there is evidence that animals and humans both navigate primarily by short-term egocentric representations that are limited by the information in the environment (R. Wang & Spelke, 2002). However, the authors agree that distinctly human navigation is characterized by building on these representations. Other research on similarities between humans and other animals' navigation systems has focused on identifying homologous brain regions between humans and other species (e.g., Amaral & Lavenex, 2007; Furtak, Wei, Agster, & Burwell, 2007; Reep & Corwin, 2009; Rushworth, Behrens, & Johansen-Berg, 2006; van Strien, Cappaert, & Witter, 2009). Among other things, this provides a basis of comparison between species in neurocognitive research.

Despite the caveats noted above, animal studies clearly have a great deal to offer, especially in terms of understanding the neurological substrates of navigation. Indeed, John O’Keefe, May-Britt Moser and Edvard Moser were awarded the 2014 Nobel Prize in Physiology or Medicine for their research with single-cell recordings showing that there are categories of neurons in the hippocampus and surrounding areas which help to code for an individual’s cognitive map. One of the central findings in this work is the discovery of “place cells”, which exhibit a particular firing pattern when the animal is at specific locations in an environment (O'Keefe & Dostrovsky, 1971). Another key finding regarding how brains navigate was the discovery of "grid cells" in the entorhinal cortex (Hafting, Fyhn, Molden, Moser, & Moser, 2005). These cells fire whenever an animal enters any one of a number of spatial locations, which are arranged in a triangular grid pattern. The discovery of this pattern of firing in grid cells led to the hypothesis that they encode a cognitive representation of Euclidean space (Hafting et al., 2005). Grid cells are similar to place cells in the sense that they fire according to the location of the animal. However, place cells fire according to specific relative locations like landmarks, and require visual input. In contrast, grid cells do not require visual input to fire and can adapt their configuration in a new environment (Quirk, Muller, Kubie, & Ranck, 1992). Grid cells fire according to an egocentric absolute perspective and are thought to track the relative movement of the individual (Hafting et al., 2005). Thus, grid cells have been proposed to form an important component of the neural substrate of navigation via path integration (Fuhs & Touretzky, 2006; McNaughton, Battaglia, Jensen, Moser, & Moser, 2006). Although, the same areas comprised of place cells and grid cells were recently found to also map non-spatial dimensions along a continuous sound frequency axis, with particular firing fields at particular sound frequencies (Aronov, Nevers, & Tank, 2017). The authors suggest the hippocampal-entorhinal systems may be used to help represent a variety of behavioral tasks beyond spatial navigation.

A third set of findings using single-cell recordings have found “head direction” cells, which are neurons that fire based on the direction the animal is facing and are commonly thought to form the basis of the “sense of direction” (Calton & Taube, 2009; Taube, 2007; Taube, Muller, & Ranck, 1990). They are found in a number of areas in and around the hippocampus (Calton & Taube, 2009; Mizumori & Williams, 1993; Ranck Jr, 1984; Taube, 1995; Taube et al., 1990; van Groen & Wyss, 1992, 1995; Vogt & Miller, 1983). They have been shown to connect to the entorhinal cortex, which contains grid cells (Wyss & van Groen, 1992). However, research on head direction cells is relatively new and it is unclear how exactly they are distributed and how they are connected to grid and place cells. More recently, subiculum hippocampal cells were found to keep track of an individual’s axis of travel (Olson, Tongprasearth, & Nitz, 2017).

Although these cells were originally discovered in rodents, studies using single-cell recordings on brain surgery patients have since confirmed similar cells in humans. Several studies found cells in the hippocampus that respond to specific spatial locations and cells in the parahippocampal region were found to respond to views of landmarks (Ekstrom et al., 2003). These give us evidence of human place cells. In addition, J. Jacobs, Kahana, Ekstrom, Mollison, and Fried (2010) found, in the entorhinal cortex, evidence for what they termed "path cells". That is, cells whose activity indicated whether a patient was taking a clockwise or counter clockwise path around a virtual square road. This same group of researchers also found that humans have grid cells that map onto the layout of a simulated town (J. Jacobs et al., 2013), although this line of evidence is preliminary.

While single-cell studies and other invasive techniques cannot routinely be used in humans, other neuroscientific methods using humans have corroborated much of the research using animal models. For instance, neuroscientific studies provide convincing evidence for a central role of the posterior parietal cortex (PPC) and the hippocampal formation during navigation. Based on lesion and fMRI studies, the PPC is likely involved in egocentric spatial processing (e.g., Janzen & Weststeijn, 2007; Leibowitz & Post, 1982; Mesulam, 1981; Posner, Walker, Friedrich, & Rafal, 1984; Schneider, 1967, 1969; Shelton & Gabrieli, 2002; Trevarthen, 1968; Van Asselen et al., 2006; Weniger, Ruhleder, Wolf, Lange, & Irle, 2009; Wolbers, Hegarty, Buchel, & Loomis, 2008; Wolbers, Weiller, & Buchel, 2004). Additionally, Iachini, Ruggiero, Conson, and Trojano (2009) showed that patients with right parietal damage were impaired on an egocentric task whereas those with left parietal damage were impaired on an allocentric task. This could be due to an egocentric frame of reference, which is attributed to the right parietal cortex, being required for action and motion, while the left parietal cortex is known for being specialized in categorical spatial processing (Iachini et al., 2009). Furthermore, fMRI and TMS (transcranial magnetic stimulation) studies have shown the PPC to be involved in spatial working memory tasks (e.g., D'Esposito et al., 1998; Kessels, d'Alfonso, Postma, & de Haan, 2000; Koch et al., 2005; Smith & Jonides, 1998). Several studies have also found sustained firing of cells in the PPC during the delay associated with spatial working memory tasks, suggesting these cells may have been “holding” the spatial information (Constantinidis & Procyk, 2004; Constantinidis & Steinmetz, 1996; Constantinidis & Wang, 2004; Curtis, 2006; Linden, 2007). More generally, it has been proposed that the PPC is involved in the integration of visual, spatial, and temporal information in working memory similar to the episodic buffer proposed by Baddeley (2000).

Conversely, implication of the hippocampal formation in navigation is supported by evidence from neural imaging, lesion, reversible inactivation, early gene expression, and electrophysiological studies (e.g., Frankland & Bontempi, 2005; Maguire, Woollett, & Spiers, 2006; Martin & Clark, 2007). This is in addition to the above-mentioned studies using single-cell recording techniques (Jayet Bray, Quoy, Harris, & Goodman, 2010; O'Keefe & Burgess, 2005; O'Keefe & Nadel, 1978; Taube et al., 1990). The hippocampal formation, along with related temporal lobe regions, are known to play a crucial role in episodic, contextual, and spatial memory functions (Burgess, Maguire, & O'Keefe, 2002; Eichenbaum, 2004; Moscovitch et al., 2005). Notably, H.M., known for having a bilateral medial temporal lobectomy, was deficient in storing and retrieving new spatial routes but did not show deficits in any other forms of spatial memory. This tells us the medial temporal lobe, including the hippocampal formation, is likely involved in the encoding and retrieval of new spatial memories. Similarly, patients with hippocampal damage display little if any navigational impairments in familiar environments despite being markedly impaired in novel ones (Calton & Taube, 2009).

Moreover, a well-known study on this topic showed, using PET scans, that London taxi drivers have larger mid-posterior hippocampi than non-taxi driving controls (Maguire et al., 2000) as well as compared to bus drivers (Maguire, Woollett, et al., 2006). This supports the idea that cognitive training can physically alter the brain and shows that this applies to spatial memory. In addition, several studies have found that increased hippocampal volume and activation were correlated with performance on an orientation-based wayfinding task or strategy; and, conversely, increased caudate nucleus volume and activation were related to performance on a route learning wayfinding task or strategy (Bohbot, Lerch, Thorndycraft, Iaria, & Zijdenbos, 2007; Head & Isom, 2010; Iaria, Petrides, Dagher, Pike, & Bohbot, 2003). Interestingly, those who had more grey matter in the caudate nucleus had less grey matter in the hippocampus and vice versa (Konishi & Bohbot, 2013). This suggests that navigation strategy may involve a "zero sum game" where neural resources devoted to one strategy causes fewer resources to be devoted to the other, corroborating the idea of spatial “profiles” described above.

In summary, the PPC is likely involved with egocentric spatial processing and the encoding of route knowledge, while the hippocampal formation is likely involved in allocentric processing and the encoding of graph/survey knowledge. It is important to note that both the PPC and hippocampal formation have been shown to be involved in spatial working memory, which makes it difficult to fully understand their individual roles and how spatial working memory in particular is manifested within the brain. Indeed, this would be an important discovery considering most navigation tasks require some form of spatial working memory, including first-person virtual navigation tasks.

REAL-WORLD VS VIRTUAL NAVIGATION

Research in navigation involves examining the combination of action, cognition, and perception as they interact in complex environments. It therefore imposes high demands on experimental design (van Veen, Distler, Braun, & Bülthoff, 1998). Researchers have attempted to meet these demands via either virtual or real-world experimental settings, which each have their own advantages and disadvantages. For instance, to engage in systematic research and to ensure reproducibility, precise control over the presented stimulus is required. In the case of navigation, this involves parametrically controlling the characteristics of the environment in which the subject is navigating, including landmarks, lighting, and traffic. Virtual environments better allow researchers to do this compared to real-world navigation research. Furthermore, it is important to engage the participant and to ensure proper interactivity with the environment using response feedback in order to close the action-perception loop. These latter requirements pose challenges for a virtual experimental setting, as the action console (e.g., joystick, keyboard) cannot be perfectly calibrated compared with real-world interactions. However, compared to other digital and in-lab tasks (e.g., static scene display tasks, passive navigational videos), three-dimensional virtual environments have a dynamic environment in which the participant is actively engaged allowing researchers to better measure the participant's interactions with the environment.

While virtual environments provide many advantages for navigation researchers, the artificial nature of the stimuli leads to concerns over the generalizability of their results to everyday wayfinding. For instance, there seem to be differences in spatial mental representations elicited by virtual versus real environments. Richardson, Montello, and Hegarty (1999) showed that, consistent with previous studies (e.g., Philbin, Ribarsky, Walker, & Hubbard, 1998; Witmer, Bailey, Knerr, & Parsons, 1996), spatial learning was generally worse for participants who learned from a virtual environment than for those who learned from a real environment. A reason for this could be that virtual environments generate greater alignment effects than real world environments (e.g., Chance, Gaunet, Beall, & Loomis, 1998; Klatzky, Loomis, Beall, Chance, & Golledge, 1998). In other words, virtual reality platforms may not be sufficient to elicit the same qualitative kind of performance as real world environments. However, some research provides data that goes against this idea. For instance, the development of spatial knowledge during navigation has been found to be no different between real and virtual environments (Ruddle, Payne, & Jones, 1997). In addition, learning from a virtual tour has been found to transfer to the corresponding real environment (Darken & Banker, 1998).

Another important limitation of virtual environments is that they often induce a form of motion sickness, called “simulator sickness” or “simulator adaptation syndrome”. Simulator sickness is likely due to a mismatch between visual cues of movement and a lack of vestibular inertial cues, which is similar to motion sickness where the opposite happens (Reed-Jones, Reed-Jones, Trick, & Vallis, 2007). Simulator sickness could have an effect on behavioral data, as participants are often dizzy and have difficulty concentrating when experiencing this phenomenon. Furthermore, simulator sickness is more often seen in women who, by consequence, also exhibit higher dropout rates, which could create a selection bias (Kennedy, Lanham, Massey, Drexler, & Lilienthal, 1995; Rizzo, Sheffield, Stierman, & Dawson, 2003). Research has shown that simulator sickness is associated with virtual set-ups with a high field-of-view and with higher-rated levels of immersion (R. C. Allen, 2000; Lin, Duh, Parker, Abi-Rached, & Furness, 2002). Despite researchers finding ways to mitigate simulator sickness, it remains an important consideration in virtual reality research settings.

First-Person Virtual Navigation Tasks

A first-person virtual navigation task is one where the participant can actively navigate through a virtual computer-generated environment in three dimensions and in which the view into the environment from the screen takes a first-person perspective. This is generally considered the most ecologically valid kind of task in a virtual setting as it mimics the dynamic visual input of real-life navigation, at least compared to two-dimensional and third person perspective tasks. First-person virtual navigation tasks can be designed to reflect day to day experiences and to be very similar to the real-world (e.g., Dalgarno & Lee, 2010; Dede, Ketelhut, & Ruess, 2002; Messinger et al., 2009). This type of task can vary depending on the type of virtual platform and technical set-up, from a simple desktop computer and keyboard to a 360 field-of-view simultation room with a dynamic floor. It is important to take these platform differences into consideration because individual differences, such as age and gender, have been shown to interact with variations in the virtual platform in determining spatial learning and performance outcomes (e.g., Calisir & Gurel, 2003; Chen & Ford, 1998; Hunt & Waller, 1999; Inal & Cagiltay, 2007; Jansen-Osmann, Schmid, & Heil, 2007; Waller, 2000). For instance, a study compared performance between men and women on a test of survey knowledge when using a large flat-screen display versus a full-dome simulator (Hedge, Weaver, & Schnall, 2017). It found that men improved their performance after learning in the simulator compared to the flat screen, but no benefit occurred for women. Furthermore, performance correlated with spatial visualization ability in men but not women. While these more elaborate platforms are becoming more prevalent, most research to date has used a simple desktop computer, flat screen, and keyboard buttons for controls.

Virtual navigation tasks and subsequent behaviors on these tasks can also vary based on the content and interactivity of the virtual environment (e.g., Yilmaz, Baydas, Karakus, & Goktas, 2015). For instance, researchers can control, based on their research question, whether to present landmarks or other cues, navigational aids like top-down maps, and/or interferring stimuli. Additionally, this type of control enables these tasks to study navigation in very different scenarios like in buildings, outdoor open spaces, towns or even mazes where participants can be given a variety of goals to perform against (e.g., Antonova et al., 2011; Astur, Ortiz, & Sutherland, 1998; Canovas, Espinola, Iribarne, & Cimadevilla, 2008; Maguire, Nannery, &

Spiers, 2006). This control allows researchers to create a task that can reliably be said to require a specific navigation strategy based on available cues or the type of goal used (for a review, see Wiener et al., 2009). For instance, landmark-based first-person virtual navigation tasks can be virtual versions of the Morris Water Maze (e.g., Astur et al., 1998), or they can present landmarks at specific decision points in order to assess spatial memory. This is in contrast to orientation-based first-person wayfinding tasks which are often devoid of landmarks, in order to avoid any confounding strategies, and often require the participant to find a shortcut or take a novel route based on an acquired allocentric cognitive map, otherwise known as either graph or survey knowledge (e.g., Nowak, Murali, & Driscoll, 2015). Another factor to consider is the type of navigational learning material (such as maps, sets of directions, etc.) given prior to performance in order to measure, for instance, spatial memory. Such materials have participants use either active or passive learning, which have been shown to favor different forms of navigation performance (Carassa, Geminiani, Morganti, & Varotto, 2002). For instance, active navigational learning has been shown to favor spatial mapping, otherwise known as graph or survey knowledge (Plancher, Tirard, Gyselinck, Nicolas, & Piolino, 2012). Conversely, landmark recognition has been shown to be improved after passive navigation versus active in unfamiliar environments (von Stulpnagel & Steffens, 2012).

The Hebb-Williams mazes (Hebb & Williams, 1946) are a set of mazes originally used for intelligence testing in mice (for a review, see R. E. Brown, 2016) whose layouts and protocol have since been standardized over several decades of research (for reviews, see Rabinovitch & Rosvold, 1951; Shore, Stanford, MacInnes, Klein, & Brown, 2001). While these mazes were originally conceived for mice, they have since been adapted to humans using virtual reality (e.g., Chebat, Maidenbaum, & Amedi, 2017; Shore et al., 2001; Therrien & Collin, 2010). These studies have confirmed the robust gender difference in maze performance favoring males in mice (e.g., Caballero-Reinaldo, Navarro-Francés, & Arenas, 2017) and in both mice and humans (e.g., Shore et al., 2001). For instance, Shore et al. (2001) found that women compared to men took longer paths and went down more blind alleys, in addition to taking longer pauses, and that the same was true of female mice compared to male mice. Shore et al. (2001) were not only able to replicate a large body of research showing such gender differences on a wide range or spatial tasks (e.g., Galea & Kimura, 1993; Gron, Wunderlich, Spitzer, Tomczak, & Riepe, 2000), but their results also showed that Hebb-Williams mazes can be adapted to humans using virtual reality. In addition, Virtual Hebb-Williams mazes have also been described as a rich enough environment to provide a sense of depth without being so immersive as to induce simulator sickness (S. Nichols, 1999).

The current study used a series of first-person virtual Hebb-Williams mazes, adapted from Shore et al. (2001), in order to assess orientation-based navigation performance. The mazes were devoid of landmarks or any other spatial or non-spatial cues other than the first-person perspective's visual information and movement information (e.g., time it took to walk to the next wall). Prior to performance, participants were shown passive video tours of the same virtual maze environments. As described above, active learning would have been better suited to building graph knowledge and supporting an orientation-based navigation strategy. However, the Hebb-Williams mazes are quite simple and performance in these mazes using active learning would have potentially been at ceiling. It is for these reasons that we chose to use a passive learning procedure in the work reported in this thesis.

Considering a large majority of recently developped video games are also virtual environments in three dimensions with a first-person perspective (e.g., Call of Duty, Halo, and

Portal), it is not unreasonable to presume that experience playing these video games may also result in better performance on research-based first-person navigation tasks (e.g., Murias et al., 2016). In addition, men have been shown to have more experience with computers (Cooper, 2006) and video games in general (Lucas & Sherry, 2004), which may not only help to make them more comfortable using computers peripherals (e.g., keyboards and joysticks) but could also result in better-developed related cognitive skills (e.g., Achtman, Green, & Bavelier, 2008; Blumberg & Fisch, 2013; Dye, Green, & Bavelier, 2009). In fact, a recent study found that individuals with experience playing navigational video games more commonly employed optimal navigation strategies compared to those without experience (e.g., Murias et al., 2016). Moreover, not only did they find that males performed better than females and had more experience playing video games, but their results also showed that those with experience playing video games performed better irrespective of gender. However, their analyses also indicated that men's superior performance could not be entirely explained by video game experience alone (Murias et al., 2016). In contrast, there is also research that did not find video game experience to affect performance during a simulated driving assessment (e.g., Stinchcombe, Kadulina, Lemieux, Aljied, & Gagnon, 2017). Additionally, in two previous studies, each using a very similar first-person virtual spatial navigation task to the current study, video game experience was shown to have no effect on performance (Shore et al., 2001; Therrien & Collin, 2010).

Thanks to developing technology in the last decade, research using virtual reality is becoming less expensive and more ecologically valid, while also providing greater experimental control. The use of virtual reality environments has made it possible to implement specific tasks that enable researchers to explore the effects of environment layout and spatial information on subjects' strategy and performance. Nevertheless, virtual reality fidelity issues, content validity concerns related to a lack of vestibular (and other) cues, and a lack of external validity due to virtual environments not being naturalistic, are all limitations when compared to real world navigation studies. However, these are generally considered small drawbacks compared to the ability to address the complementary limitations of real-world environments, which include lack of experimental control over the environment and its parameters/conditions. It is for all these reasons we have decided to use a first-person virtual navigation task in our endeavors to understand the metacognitive aspects of spatial navigation.

CHAPTER SUMMARY

Cognitive research on spatial navigation has focused on determining the nature of spatial mental representations and their characteristics. Our review of the literature has shown that there are two broad perspectives to consider: a first-person egocentric point-of-view and a third-person allocentric perspective. Related to this are different strategies that can be used, which make use of varying forms of spatial knowledge. As we acquire higher forms of spatial knowledge with more exposure, we begin to build a representation of the navigation environment using a variety of cell types in the hippocampal formation and posterior parietal cortex. The choice of navigation strategy depends on the characteristics of the task being done, the nature of the information present in the environment, and the individual tendencies of the navigator.

We have determined that the most widely used task in human navigation research, first-person virtual navigation, has many advantages and disadvantages. In the current project, the advantages outweigh the disadvantages, though important considerations will be kept in mind, including ways to avoid simulator sickness, as well as its effects on our sample.

CHAPTER 2: METACOGNITION

The capacity to reflect on our cognitive processes is a remarkable human characteristic. Indeed, many scholars suggest that it is this characteristic that separates us from other animals (Descartes, 1637; Metcalfe & Kober, 2005). The purpose of this chapter is to review this complex higher-order cognitive process. We will begin with a review of research in metacognition. This will be followed by a description of a widely used theoretical framework of the process. The chapter concludes with a rationale for applying the framework to spatial navigation research, an endeavor that has not been previously attempted.

THE HISTORY OF RESEARCH IN METACOGNITION

Interestingly, despite William James declaring introspection to be the cornerstone for measuring psychological events in 1890 (James, 1890) there was little research done in the realm of metacognition until the 1960's. Once cognitive psychology came to the fore in the latter half of the 20th century, scientists began developing empirical studies on the subjective awareness of cognitive processes and metacognition moved from the realm of philosophical discussion to empirical investigation (for a review, see Hacker, 1998; Nelson, 1992). Flavell (1979), who was studying children's theory of mind, was particularly influential, coining the term "metacognition" and defining it as "knowledge and cognitions about cognitive phenomena" (p.906, Flavell, 1979). The general concept of metacognition can be applied to a variety of specific cognitive processes, yielding such phenomena as metacomprehension (Maki & Berry, 1984), metamemory (Nelson & Narens, 1990) and visual metacognition (Levin, 2004). It is no surprise then that metacognition has developed into a broad field of study and that its constructs have been operationally defined in many ways (e.g., Kluwe, 1982; Schooler, 2002). Despite varying conceptions, the universal notion of metacognition remains true to its etymology: “cognition about cognition” or “thinking about one’s own thoughts”.

Notably, Flavell (1979) refers to the awareness of ongoing cognitions as metacognitive experience while awareness of how a cognitive process operates is considered metacognitive knowledge. For example, knowing that one's navigation ability is poor would be considered metacognitive knowledge, whereas realizing one went the wrong way during an ongoing navigational task would be considered metacognitive experience.

It is important to note that the contemporary scientific literature makes a distinction between *global metacognition* and *local metacognition*. (e.g., Kelemen, Frost, & Weaver, 2000; Lieberman, 2004). Global metacognition is synonymous with metacognitive knowledge, as it represents a general self-impression of how a specific cognitive process personally operates. Thus, global metacognition can be measured using self-reported questionnaires on general ability, comfort, or confidence during a specific type of cognitive task. For instance, the Santa Barbara Sense of Direction Scale (Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002) measures a form of global metacognition related to an individual’s self-perceived ability in various daily navigation scenarios. Conversely, local metacognition is synonymous with metacognitive experience, as it represents online metacognitive processing specific to an ongoing task. It is also generally accepted that the source of global and local metacognitive thoughts is an internal representation rather than an external source (Nelson & Narens, 1990). A more recent framework by Nelson and Narens, in development since 1990, has focused on a belief that the role of a local metacognitive system is to monitor and control that internal representation (Nelson & Narens, 1990). That is, it not only involves knowing if one is currently lost but also determining what one should do about it (e.g., checking the map again) in order to better reach a specific goal.

Nelson and Narens (1990) Metamemory Framework

According to Nelson and Narens' (1990) framework for metamemory, cognitive processes are divided into two hierarchical frames of reference, and communication between these two frames of reference is accomplished through a cycle of monitoring and control processes. The lower order frame of reference is at the *object-level*, which contains information on the cognitive process in question, for instance a representation of your current cognitive map. This information then flows up from the object-level to the higher order *meta-level*, with the purpose of updating the meta-level with regard to the state of the object-level. This process is called "monitoring". The meta-level contains a dynamic representation of the object-level, and it evaluates the accuracy of that representation based on a specific goal. An example of the meta-level would be a representation of how incomplete or potentially false your cognitive map may be (e.g., knowing you do not know the layout of a particular corridor in order to get to where you need to go). Information from the meta-level then flows back down to the object-level using a process called "control" in order to modify the state of the object-level by informing our actions (e.g., exploring the missing corridor, looking at a road map, or asking for directions). In other words, the role of metacognition is to regulate our cognitive processes in relation to the inputs and outputs from the external world.

While Nelson and Narens' theoretical framework for metamemory was an important advance in metacognition research, their most critical contribution to the field is their development of an empirical approach to studying the monitoring and control processes that exist at each stage of learning. In their approach, control processes are operationally defined by measuring study time, strategy selection, and the duration of memory search, whereas monitoring processes are operationally defined via prospective and retrospective judgments of retrieval performance. The ability to operationally define metacognition in this way has allowed the scientific community to develop methods of examining prospective and retrospective judgments of retrieval information that are now well validated and that are applied in a variety of research settings. These include three prospective judgments and one retrospective judgment, namely: the Ease-of-Learning (EOL), the Judgment-of-Learning (JOL), the Feeling-of-Knowing (FOK) and the Retrospective Confidence Judgment (RCJ). EOLs reflect the *a priori* estimate of item difficulty and JOLs are estimates of future retrieval performance (prospective confidence). FOKs are estimates of the likelihood of *recognizing* an item either prior to retrieval, after a failed retrieval, or long after retrieval as a measure of long-term metamemory. RCJs are estimates of the accuracy of the response *after* retrieval (retrospective confidence). There is evidence to confirm that these judgments are indeed distinct from each other and reflect separate monitoring processes (Nelson & Narens, 1990). For instance, correlations between the four judgments are weak (Leonesio & Nelson, 1990). In addition, JOLs and RCJs seem to rely on different sources of information (e.g., Busey, Tunnicliff, Loftus, & Loftus, 2000; Dougherty, Scheck, Nelson, & Narens, 2005) and are differentially affected by encoding manipulation (e.g., study time manipulations affects JOLs but not RCJs; Dougherty et al., 2005), and neurological impairments (e.g., Modirrousta & Fellows, 2008a; Pannu & Kaszniak, 2005). Finally, FOK and RCJ’s have been shown to have common and distinct neural mechanisms (Chua, Schacter, & Sperling, 2009).

An index of participants' metacognition abilities can be obtained by comparing the accuracy of their monitoring judgments to objective performance measures. The accuracy of metacognitive judgments can be measured in two different ways (e.g., de Bruin & van Gog, 2012; Dunlosky & Rawson, 2012; Koriat, 2012). First, using the correlation between a metacognitive judgment and performance on a criterion test (Nelson, 1984), one can determine what Nelson and Narens call *relative accuracy*; this measure allows one to examine how participants monitor items relative to each other. That is, relative accuracy reflects the degree to which participants can discriminate between items that will and will not be remembered (i.e., monitoring) and affects which items participants select for further study (i.e., control). It is usually determined using the Goodman-Kruskal gamma correlation between a numerical rating of prospective confidence and a dichotomous variable indicating whether the participant was correct or not in his/her answer.

The second method looks at the difference between the mean predicted memory performance and the mean observed memory performance (Keren, 1991). This measure is called *absolute accuracy* or *calibration*, and it is useful for determining overall under-confidence or over-confidence. Absolute accuracy tells us how accurate someone is at predicting or assessing their overall performance (i.e., monitoring) and affects the amount of overall time students invest for further learning (i.e., control). Absolute accuracy is usually assessed by calculating a calibration curve that demonstrates the relationship between the proportions correct and subjective probability per condition. The position of this line relative to the perfect-judgment line (i.e., a 45° line linking performance and predicted performance) indicates whether overall over-confidence or under-confidence is being exhibited. For instance, the example calibration curves in Figure 1, using data from Experiment 3 of this project, show both men and women to be slightly overconfident for RCJs as their calibration curves are below and to the right of the perfect-judgment line. Absolute accuracy thus refers to the height of the line linking performance to prediction (or, equivalently, the y-intercept), while relative accuracy refers to its slope.

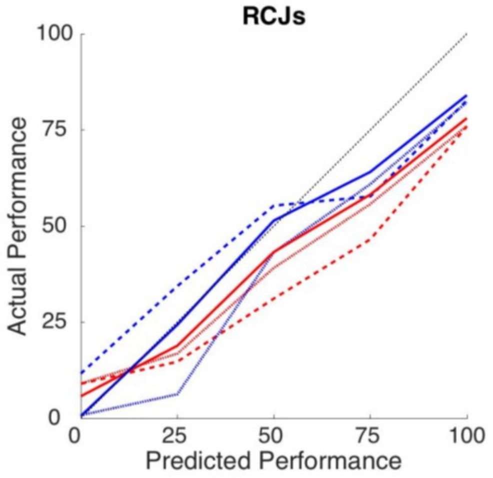


Figure 1. Example calibration curve using data from Experiment 3 of the current project. Blue lines represent male groups, and red lines female. Line styles represent different study conditions.

Research using these kinds of judgments has shown us, for instance, that EOLs are correlated with study time (Nelson & Leonesio, 1988), as well as being a significant predictor of an individual's rate of learning (Leonesio & Nelson, 1990; Underwood, 1966). JOLs provide insight into the current level of mastery as well as whether additional study time is warranted, and are a significant predictor of retrieval performance, especially after a delay from study (Thiede & Dunlosky, 1994). In a more recent study, overconfidence (i.e., monitoring) was associated to earlier termination of practice (i.e., control) as well as underperformance (Dunlosky & Rawson, 2012). Generally speaking, students are more effective at regulating their study time as monitoring accuracy improves, and their performance increases (e.g., Hacker, Bol, Horgan, &

Rakow, 2000; Mihalca, Mengelkamp, & Schnotz, 2017), but see a Dunlosky and Ariel (2011) for a review. For example, participants spend less time studying self-judged easy items and focused more on self-judged difficult items (for a review, see Son & Metcalfe, 2000). However, recent findings have suggested that the relationship between monitoring accuracy and performance is not so simple and may depend on task demands. For instance, Son and Metcalfe (2000) found that, contrary to previous studies, participants allocated *more* study time (i.e., control) to easy items, but only if study time was restricted and insufficient.

Despite a large body of research showing a positive relationship between monitoring accuracy, control, and performance (e.g., Dunlosky & Ariel, 2011; Dunlosky & Rawson, 2012; Hacker et al., 2000; Nietfeld, Cao, & Osborne, 2005; Thiede, Anderson, & Therriault, 2003), there is little research on the causal link between these variables (Thiede et al., 2003). That is, there is little evidence that increased monitoring accuracy influences control behaviors such that they cause an improvement in performance. Furthermore, recent evidence related to a causal link points to a more complex explanation than provided by previous theories (Schwartz & Efklides, 2012). For instance, several studies have posited a bidirectional link between monitoring and control whereby a control process (i.e., spending extra time studying an item) can be used as a cue and feed into monitoring ("that item must have been hard if it took so long to memorize, therefore it probably won't be remembered") (e.g., Koriat, 2006, 2012; Koriat, Ackerman, Adiv, Lockl, & Schneider, 2014; Koriat, Ma'ayan, & Nussinson, 2006). Recent work in local metacognition has focused on the various types of internal and external cues individuals use to help with accurate monitoring but also how they could cause metacognitive illusions (e.g., Castel, McCabe, & Roediger, 2007; Rhodes & Castel, 2009). Metacognitive illusions happen when a particular piece of information (cue) affects metacognitive monitoring but not memory performance (for a review, see Bjork, Dunlosky, & Kornell, 2013). A further discussion on the effects of various cues on each monitoring judgment will follow in the next subheading.

MECHANISMS UNDERLYING METACOGNITIVE MONITORING JUDGMENTS

The purpose of this section is to provide a more detailed understanding of three monitoring judgments relevant to this dissertation (EOLs, JOLs, and RCJs)[2]. A large number of studies have shown these judgments to be statistically distinct (e,g, Busey et al., 2000; Dougherty et al., 2005; Modirrousta & Fellows, 2008b; Pannu, Kaszniak, & Rapcsak, 2005) including a study that found the judgments correlated poorly with each other (e.g., Leonesio & Nelson, 1990). In view of that, each metacognitive judgment will be reviewed in its own section, including the procedure for measuring it, followed by the relevant factors that can influence its accuracy and the possible theoretical mechanisms underlying it. Despite several other types of judgments having been introduced since the original Nelson and Narens (1990) framework, such as the Judgment of Retention (Tauber & Rhodes, 2012) or the Judgment of Improvement (Townsend & Heit, 2011), the literature thus far has focused far more on the original judgments (for a review, see Dunlosky & Metcalfe, 2009) and so will this project. It is important to note that most studies examining these monitoring judgments have used word-pair associations as their cognitive task. There is no reason why the Nelson and Narens framework cannot not be applied outside this context (e.g., Watier & Collin, 2010, 2011, 2012), but one must keep in mind that much of the research is limited to the realm of verbal stimuli.

---

[2] FOKs were not included in this section because they were not used in the experiments described below. See Nelson and Narens (1990) for more about FOKs.

Ease-Of-Learning Judgments

The degree of subjective ease of acquiring an item to an ideal level of mastery is reflected in the EOL. Accordingly, an EOL can be thought of as a monitoring device used to determine how much study time is to be allocated to a particular item (Nelson & Narens, 1990). For instance, it has been found that, as the predicted level of difficulty increases, the time spent studying also increases (Nelson & Leonesio, 1988). In addition, during self-paced study conditions participants spend fewer trials memorizing easy-to-learn word pairs (Leonesio & Nelson, 1990). EOLs have also been found to be significant predictors of retrieval accuracy (Nelson & Leonesio, 1988; Son & Metcalfe, 2000), and rate of acquisition (Leonesio & Nelson, 1990). They enable study time to be efficiently allocated through effective monitoring.

EOLs can be collected in a variety of ways. For instance, participants can be asked to draw a line whose length corresponds to the perceived level of difficulty (Underwood, 1966), or to rank items based on ease of learning (e.g., Leonesio & Nelson, 1990; Son & Metcalfe, 2000). However, they are typically collected by having participants rate the perceived difficulty of each item separately on a numerical scale (e.g., de Carvalho & Yuzawa, 2001). The predictive relative accuracy of EOLs is determined by comparing EOL ratings with performance across each trial, using the Goodman-Kruskal gamma correlation. Absolute accuracy cannot be computed for EOLs, as the EOL does not refer to the probability of a successful performance, but instead the subjective ease or difficulty of an item. It cannot, therefore, be compared to observed performance as there is no predicted performance.

While little research has examined the underlying mechanisms of EOLs, they are presumably based on inferential processes. Indeed, this must be the case, as the participants have yet to memorize the items. Koriat has suggested the cue-utilization hypothesis to explain EOLs.

This is where a priori knowledge of memory ability (metacognitive knowledge) and extrinsic properties of the learning conditions are used to determine an item’s difficulty (Koriat, 1997, 2000). In other words, if participants believe that their memory for particular items will be poor, then their EOLs will reflect those beliefs. In other words, participants base their EOL judgments on a combination of their assessment of their own global memory abilities and characteristics of the stimuli that suggest that the items will be easier or harder to learn. For instance, individuals might believe that they generally have a poor memory for words, but believe that a given stimulus word is very easy to memorize because it is short and distinctive; In this case they might give a moderately high EOL judgment value when predicting how hard it would be to memorize the word. Other cues can also influence EOLs. For instance, if participants are told that certain word pairs are considered easy to memorize, they bias their EOLs to correspond with the information (de Carvalho & Yuzawa, 2001). Further support for the cue utilization hypothesis comes from the fact that EOL ratings form a bi-modal distribution if word pairs are grouped according to easy versus hard difficulty levels (de Carvalho & Yuzawa, 2001) or into semantically related versus unrelated groups (Dunlosky & Matvey, 2001). It is important to note that EOLs are known for being the least accurate compared to the other judgments (Dunlosky & Matvey, 2001; Leonesio & Nelson, 1990).

Judgments-of-Learning

A JOL reflects a monitoring process by which one estimates the likelihood of future recall for a recently or currently learned item. For example, when navigators ask themselves whether they can successfully recall the route they have just travelled, they are essentially conducting a JOL. Like EOLs, JOLs determine whether additional study time is required (Nelson

& Leonesio, 1988). However, in addition, they can be used as a control device to determine if a new learning strategy is warranted, since learning has already begun (Nelson & Narens, 1990). In order to account for how the outputs of JOLs control study time, Dunlosky and Hertzog (1998) proposed the discrepancy-reduction hypothesis. The hypothesis assumes that participants can monitor the strength of an item's memory and compare it to a desired level of mastery, allocating additional study time if the strength of the memory is insufficient. However, under a time constraint, participants will allocate additional study time to the easiest items in order to maximize the number of items retained (Thiede & Dunlosky, 1999) showing that a discrepancy-reduction mechanism is not necessarily the optimal strategy for all learning situations.

Typical JOL experiments consist of a word-pair association task with a cued-recall test, where the cues presented at the time of the JOL are also presented at test. Just as with EOLs, the predictive relative accuracy of JOLs is determined by comparing JOL ratings with performance across each trial, using the Goodman-Kruskal gamma correlation. Unlike EOLs, absolute accuracy can be computed for JOL as it does refer to the probability of a successful performance. It is, therefore, possible to compare observed performance to predicted performance via a calibration curve, as described above.

A few reviews have established JOLs to be predictive of future recall performance with an average gamma correlation of 0.56 (e.g., Koriat, 1997, 2000; Koriat, Sheffer, & Ma'ayan, 2002), however, these conclusions are based on studies with verbal stimuli (e.g., Koriat, 1997, 2000; Mazzoni & Nelson, 1995; Nelson, 1984; Nelson & Dunlosky, 1991) and it remains to be seen whether these results will generalize to other domains, such as navigation.

Research on the underlying mechanisms of JOLs has demonstrated that they are inferential in nature (Koriat, 1997, 2000). According to the cue-utilization hypothesis and supported by research findings, JOLs are based upon a variety of cues where the relative importance of each cue depends on one's a priori beliefs and on the demands of the task (e.g., Dunlosky & Matvey, 2001). Intrinsic cues refer to inherent characteristics of the stimuli, for instance, concrete words produce increases in JOL and memory performance compared to abstract words (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989). This is likely because concrete words are easier to process and yield superior recognition memory compared to abstract words (Begg et al., 1989). Interestingly, JOLs increase for high frequency words but recognition accuracy does not, likely because high frequency words are easier to process but do not necessarily yield superior recognition memory, producing a metacognitive illusion. More examples of intrinsic cues causing metacognitive illusions include an increase in JOL rating for words spoken more loudly at study (Rhodes & Castel, 2009) or words presented in larger font sizes during study (McDonough & Gallo, 2012; Rhodes & Castel, 2009). This effect is even seen for words that are associated with a greater physical weight during study (Alban & Kelley, 2013). In addition, Matvey, Dunlosky, and Guttentag (2001) found a negative correlation between response latency and JOL rating, which suggests that processing fluency can influence JOL ratings. Thus, participants will be less confident in their ability to recall an item the longer it takes to come to mind (i.e., low processing fluency).

In contrast, extrinsic cues are defined by learning conditions and strategies used by the learner. For instance, relative JOL accuracy is also mediated by over-learning (Leonesio & Nelson, 1990) as well as by practice (e.g., Koriat et al., 2002). That is, relative JOL accuracy was greatest for items that continued to be presented after four consecutive correct trials compared to items that were presented after only one correct trial (Leonesio & Nelson, 1990). Interestingly, while some studies have found relative JOL accuracy to be improved when degree of learning is controlled (Shaughnessy & Zechmeister, 1992), many others have found a counter-intuitive effect whereby increased practice results in increased under confidence both among relative and absolute accuracy analyses (Koriat et al., 2002). This effect has been named the "under-confidence with practice effect" and is considered to result from mnemonic debiasing; that is, by counteracting the pre-existing biases we have towards overconfidence (Koriat, Ma'ayan, Sheffer, & Bjork, 2006).

Retrospective Confidence Judgments

According to the Nelson and Narens (1990) framework, an RCJ is an assessment of the probability that an item retrieved from memory at test is the correct target (e.g., Koriat, Pearlman-Avnion, & Ben-Zur, 1998). It is the only metacognitive judgment in their framework that is solicited after completing the task in order to assess past performance. The accuracy of RCJs is usually assessed through calibration curves (Keren, 1991). This allows one to determine the degree of over- or under-confidence being exhibited, although relative accuracy (i.e., Goodman-Kruskal gamma correlation) is sometimes reported as well.

Findings regarding the absolute accuracy of RCJs on general knowledge questions usually demonstrate an overall trend of overconfidence (Lichtenstein, Fischhoff, & Phillips, 1982). This could possibly occur due to a bias to focus only on the evidence supporting a retrieved answer while ignoring contradictory evidence (Koriat, Lichtenstein, & Fischhoff, 1980). Contrary to what one might expect, overconfidence is usually observed for difficult items, whereas underconfidence is typically observed for easy items (Juslin, Winman, & Olsson, 2000). This could be due to a greater ignorance of contradictory evidence during difficult tasks as opposed to easy ones. RCJs also vary according to exposure duration, where relative accuracy decreases as exposure duration decreases (e.g., Brewer, Weber, & Wells, 2004; Busey et al., 2000; Kelley & Lindsay, 1993; Memon, Hope, & Bull, 2003). Kelley and Lindsay (1993) found that prior exposure to either correct or incorrect answers significantly decreased response latency and increased confidence. That is, if an answer comes to mind quickly, regardless of its objective accuracy, participants will be more confident in that answer. This again shows the impact of retrieval fluency on metacognitive judgments.

The mechanisms that underlie RCJs appear to involve two general processes: an effortful and deliberate consideration of selected answers against possible alternatives (Koriat & Goldsmith, 1996; Koriat et al., 1980; Nelson & Narens, 1990) and an automatic and unconscious application of an ease-of-processing heuristic (Kelley & Lindsay, 1993; Koriat & Levy-Sadot, 1999). Thus, retrospective confidence can emerge from either a thoughtful deliberation regarding a previous performance, or a rapid and automatic assessment of one’s subjective feelings regarding a previous performance.

RATIONALE FOR APPLYING A METACOGNITIVE FRAMEWORK TO SPATIAL NAVIGATION

In order to operationally define the constructs underlying metacognition, Nelson and Narens and others have proposed the use of prospective (i.e., EOLs, JOLs, and FOKs) and retrospective (i.e., RCJs) judgments of local metacognition. These judgments are now commonly used in a variety of research settings, including those with non-verbal stimuli such as face identification (e.g., MacLin, 1999; Sommer, Heinz, Leuthold, Matt, & Schweinberger, 1995; Watier & Collin, 2010, 2011, 2012), picture identification (Butterfield, Nelson, & Peck, 1988), and even olfactory recognition (Jonsson & Olsson, 2003). Because only a few papers have investigated specific indicators of metacognition in relation to navigation or spatial cognition (e.g., Schwartz, 2006; Vecchi, Albertin, & Cornoldi, 1999), it is hard to come to any firm conclusions on the subject (but see next subsection for a brief review). That is, the literature in regards to trial-by-trial self-rated performance during spatial navigation is sparse, which is one motivation for conducting the present research.

While it is of interest to study the metacognitive aspects of spatial navigation in their own right, there are also several practical and theoretical advantages to developing this line of research. The role of conscious experience during navigational processes has been largely neglected in the literature. For instance, ascertaining that we navigate more accurately in the presence of more distinct and salient landmarks does not provide any understanding of the subjective experience that participants have, or the conscious strategies that participants use to make navigational decisions. By examining and assessing the subjective characteristics of navigational behaviors, a more complete account of real-world navigation can be established.

Within Nelson and Narens' (1990) framework, in order for a storage and retrieval system to be effective it must be able to 1) monitor whether an item has been sufficiently processed, 2) control whether further learning is necessary, and 3) determine whether a retrieved item is the best answer to be offered. Subsequently, examining how monitoring and control mechanisms operate during spatial navigation tasks will inform us regarding the conditions under which navigational errors take place. For instance, prospective monitoring judgments (EOL and JOL) can be used to determine if certain task variables (e.g., navigational aids like maps, number of learning trials) or stimulus variables (e.g., characteristics of landmarks or routes) lead participants to erroneously believe that a route is accurately and sufficiently committed to memory. Similarly, examining RCJs can inform us regarding what kinds of task and stimulus characteristics lead people to falsely believe that they have successfully completed a navigation task when they have not.

Applying a metacognitive paradigm to a spatial navigation task should elucidate several underlying processes of route learning; for instance, how participants allocate their study time, the encoding strategies they use and whether these are effective, what recall strategies they use in order to make the best navigational decisions, and whether prospective confidence is an accurate reflection of actual navigation ability. It also opens the door to hypotheses related to other findings found in the spatial navigation literature, such as awareness of cognitive maps or spatial knowledge, differences in implicit and explicit spatial knowledge, as well as whether spatial perspectives can be controlled or optimized depending on environmental conditions. Finally, it could provide an explanation for the persistent gender gap in navigation performance; That is, it might be that men outperform women due to differences in domain-specific metacognitive ability.

Metacognition in Spatial Navigation

As previously noted, research on metacognition, and more specifically the Nelson and Narens' (1990) framework, has been largely based on verbal tasks such as word-pair association learning. It is therefore unclear to what degree this work is applicable to spatial navigation tasks. That is, the findings from word-pair association typical of the metacognitive literature may not generalizable to a higher-order cognitive process like spatial navigation, which typically use complex dynamic visual stimuli and a variety of lower-order cognitive processes. Hence, it is possible that distinct metacognitive systems for words and places might exist, and that these might be qualitatively different in how they function. Evidence for this comes from Perfect and

Hollins (1996, 1999) who showed that metacognitive judgments were not predictive of performance for episodic information, though they were for semantic information. However, it is also possible that a more general metacognitive system exists which may subserve various cognitive functions. The current study seeks to launch research in this area by helping to determine if the Nelson and Narens (1990) metamemory framework can be applied to a spatial navigation task.

As noted earlier, there is very little research that combines spatial navigation (or even spatial cognition in general) and metacognition. The only previous study to use the Nelson and Narens framework within a spatial navigation task was by Schwartz (2006), who collected JOLs in two experiments, respectively examining the use of maps and verbal directions. The results show very strong gamma correlations, 0.81 and 0.85 respectively, between prospective confidence and performance. These are very high compared to the findings of other studies examining immediate JOLs, which typically find correlation values between 0.3 and 0.4 (Nelson & Dunlosky, 1991). In addition, results showed good calibration (i.e., absolute metacognitive accuracy) with both map and verbal direction tasks, with little evidence of over- or under-confidence. Schwartz suggests that the high gamma correlations for his tasks may be due to their ecological validity, since people typically have much more experience receiving directions and using maps than, for example, memorizing word-pair associates.

While there has been only one wayfinding study examining the kind of trial-by-trial self-assessments that are used in association with the Nelson and Narens (1990) framework (Schwartz, 2006), there have been many examining the relationship of more global self-ratings with spatial cognition performance. For instance, Vecchi et al. (1999) found no correlation between a self-assessment questionnaire on everyday spatial memory and six spatial memory tasks. However, they did find that young and old participants indicated similar self-assessments despite having significantly different performance levels (favoring the young participants). This may indicate that confidence regarding spatial tasks does not adjust with age-related memory decline, a finding also seen with other cognitive tasks (Vecchi et al., 1999). There are other forms of relevant subjective knowledge that could be helpful in understand the metacognition of spatial navigation. For instance, research has shown that confidence is directly predicted by metacognitive knowledge (e.g., Jiang & Kleitman, 2015; Kleitman & Gibson, 2011; Kleitman & Stankov, 2007; Morony, Kleitman, Lee, & Stankov, 2013), by academic and problem solving self-concept (Efklides & Tsiora, 2002; Kröner & Biermann, 2007), by self-efficacy (Stankov, Lee, Luo, & Hogan, 2012), and spatial cognition (e.g., Bryant, 1991; Cooke-Simpson & Voyer, 2007; Moe & Pazzaglia, 2006; Wraga, Duncan, Jacobs, Helt, & Church, 2006). In addition, spatial anxiety, defined as anxiety about performing spatial tasks (Lawton, 1994), is negatively associated with spatial ability (e.g., Ramirez, Gunderson, Levine, & Beilock, 2012). It follows then that variations in spatial abilities could partially be due to variations in one's metacognitive abilities. The role of metacognition is to regulate our cognitive processes in relation to the inputs and outputs from and to the external world. For instance, navigational metacognition processes help us know when to stop exploring an environment or where to search next; they may also help to determine our level of confidence or self-doubt in taking a turn or in giving directions to others.

Research on gender differences generally shows that men are more confident than women on a variety of tasks (e.g., Lawton, Charleston, & Zieles, 1996; O'Laughlin & Brubaker, 1998; Picucci, Caffo, & Bosco, 2011). Men also exhibit less spatial anxiety (e.g., Lawton, 1994; Lawton & Kallai, 2002). It is difficult to know how these findings regarding global confidence might be related to possible gender differences in the more granular trial-by-trial metacognitive processes addressed by Nelson and Narens' model. However, they do suggest the possibility of gender differences with regards to these kinds of metacognitive judgments. A more thorough look at gender differences in metacognition will be covered in the following chapter.

CHAPTER SUMMARY

The capacity to reflect on our cognitive processes is a vital human ability. While empirical research in the field of metacognition is relatively new compared to that in spatial navigation, there nonetheless exists a cogent framework within which one can attempt to apply the concepts of metacognition to various fields, including navigation. Briefly, this framework suggests that the role of the *meta-level* is to evaluate what is being monitored, and based on this evaluation, control *object-level* processing via feedback. The ability to calibrate confidence is an important aspect of metacognition because accurately knowing what one knows and what one does not know has important implications for learning behaviors. I have covered findings that relate EOLs, JOLs, and RCJs to performance and task characteristics – now we want to see if these concepts generalize to research in navigation. Only one past study has examined similar issues. It found strong absolute and relative accuracy measures on spatial tasks. Aside from the general issue of examining how current models of metacognition generalize to navigation, there are a number of more specific novel research questions to be explored. These include questions regarding whether differences in metacognition can explain performance differences between the genders and whether manipulating confidence via stereotype activation affects metacognitive performance.

CHAPTER 3: GENDER DIFFERENCES IN SPATIAL NAVIGATION

THE HISTORY OF RESEARCH ON GENDER DIFFERENCES IN SPATIAL COGNITION

Research on differences between the genders has been conducted since the birth of psychology itself and is found within every subfield of the discipline. There are many books on gender[3] and gender differences in psychology (e.g., Chrisler & McCreary, 2010), and more specifically, cognition (e.g., Caplan, Crawford, Hyde, & Richardson, 1997). Most of these have a specific chapter covering the strong, clear, and persistent gender difference in spatial cognition. This is contrasted with other established gender differences in cognition, which mostly demonstrate a very weak effect (Hyde, 2005). Despite recent meta-analyses demonstrating a disappearance or reduction of gender difference in many areas (e.g., Feingold, 1998; Hyde, 1981, 2005; Martens & Antonenko, 2012; Voyer et al., 1995), those in spatial cognition have been reducing at a much slower rate (e.g., Feingold, 1998; Voyer et al., 1995). The decrease in the magnitude of gender differences in recent years argues that attitudes concerning gender-related cognitive differences have changed, which has likely affected the way children are raised and the way the genders approach different tasks.

The first extensive review to explicitly look at gender differences in psychology was in 1974 by Maccoby and Jacklin. It clearly established the existence of gender differences in spatial abilities favoring males. Based on the magnitude of gender differences found in a meta-analysis done by Hyde (1981), where effects sizes among the studies reported in Maccoby and Jacklin (1974) were calculated, it was concluded that gender only accounted for 5% of the variance in

---

[3] As cultural and biological factors cannot be separated when comparing gender/sex differences (A. Kaiser, 2012), the authors chose to use the word "gender" throughout as it more accurately refers the differences in the state of being male and female by acknowledging social/cultural differences in addition to biological differences.

spatial tasks. A subsequent meta-analysis by Linn and Petersen (1985) used a psychometric as well as a cognitive rationale to classify spatial tests based on homogenous effect sizes between categories. They found three distinct categories: spatial perception, mental rotation, and spatial visualization. They defined spatial perception as the ability to determine spatial relations in spite of distracting information, which produced a significant mean effect size of 0.44. They defined mental rotation as the ability to quickly and accurately mentally rotate figures in two or three dimensions, which produced a significant mean effect size of 0.73. Finally, they defined spatial visualization as the ability to manipulate complex spatial information, which produced a mean effect size of 0.13 that was not significant. Thus, Linn and Petersen's (1985) analysis indicates a significant and quite robust difference in certain categories of spatial tasks (spatial perception and mental rotation) but not for others (spatial visualization). This result suggests that Hyde (1981) may have underestimated the percentage of variance accounted for by gender in spatial performance by averaging across all tasks and not considering the possibility that different types of spatial tasks may produce greater gender differences than others. However, many researchers have taken issue with these findings (e.g., Caplan, Macpherson, & Tobin, 1985). Specifically, many have noted the difficulty with defining and categorizing spatial tasks (e.g., Burnett, 1986; Chrastil, 2013; Halpern, 1986; Hiscock, 1986; Sanders, Cohen, & Soares, 1986).

Another meta-analysis by Voyer et al. (1995) categorised each spatial task into one specific component of spatial abilities and looked at gender differences in this context. They found similar results to those found in Linn and Petersen’s (1985) meta-analysis (i.e., effect sizes of 0.56 for mental rotation, 0.44 for spatial perception, and 0.19 for spatial visualization). Voyer et al. (1995) also split studies by age of participants and year of publication, and were able to determine that there was a small decrease in gender differences over time. However, this was only significant for a minority of the tasks. They also found that gender differences were not found in prepubescent participants for most tasks. They noted that one must take these last two findings lightly, as their sample size for young participants was very small, and the tasks are likely not comparable when administered to young children.

Several researchers have posited that the gender difference seen in most spatial cognition tasks may be simply due to differences in visual-spatial working memory. A recent meta-analysis comparing gender on visual-spatial working memory (VSWM) ability found a small male advantage in this ability across a variety of tasks, with the exception of object-location memory (Voyer et al., 2017). It is important to note that the effect size of this gender difference (excluding object-location memory tasks) is small ($d = .21$) and could not explain the gender difference in spatial tasks, which is much larger ($d = .94$) (Linn & Petersen, 1985). That is, gender differences in VSWM might partially account for gender differences in spatial abilities, but there also is much variance left to be explained. Interestingly, the gender difference in VSWM was not significant before puberty indicating a potential influence from pubescent hormones (Voyer et al., 2017). This last finding corroborates a similar finding described above in a meta-analysis among spatial tasks, where gender differences were also only seen during and after puberty (Voyer et al., 1995).

Another hypothesis regarding gender difference in spatial cognition is related to evidence showing that the genders process visual information differently, with women focusing on object-imagery processing and men on spatial-imagery processing (e.g., Blajenkova, Kozhevnikov, & Motes, 2006). Indeed, these differences in visual-cognitive style between the genders have been related to differences in spatial cognition (e.g., Blazhenkova & Kozhevnikov, 2009; Yoon, Choi, & Oh, 2015). For instance, mental rotation and maze navigation tasks were better performed by spatial visualizers, whereas picture recognition tasks were better performed by object visualizers (e.g., Farah, Hammond, Levine, & Calvanio, 1988). Thus, the gender difference in spatial tasks could be due to the fact that women tend to be object visualizers and not spatial visualizers. The evidence supporting this hypothesis is quite robust: a series of studies have consistently demonstrated, in both children and adults, that these visual cognitive styles have high internal reliability as well as predictive, discriminative, and ecological validity (e.g., Chabris & Glickman, 2006).

Based on these findings, it is likely not valid to make general statements concerning gender differences in spatial ability overall. Rather, based on the cognitive literature, it appears gender differences in spatial ability depend upon the nature of the task, VSWM ability, and visual-cognitive style. There is only one type of task that has shown a relatively reliable yet weak female advantage and that is in object-location memory (e.g., Andreano & Cahill, 2009; De Goede & Postma, 2008; Lejbak, Vrbancic, & Crossley, 2009; Levy, Astur, & Frick, 2005; Silverman, Choi, & Peters, 2007; Spiers, Sakamoto, Elliott, & Baumann, 2008). Object-location memory tasks typically involve the pairing of, usually random, objects and specific locations within an experimental environment. The only relevant meta-analysis on this subject confirmed that women outperform men in tasks requiring them to memorize the location of objects, with a small average effect size (Voyer, Postma, Brake, & Imperato-McGinley, 2007). However, this effect was not seen in studies with pre-pubescent participants, nor was it seen for certain object types, including uncommon and gendered objects (Voyer et al., 2007). As mentioned above, the female advantage in object-location memory tasks may be due to the fact that they tend to be object visualizers, allowing them to better code an object's characteristics to its location in memory.

GENDER DIFFERENCES IN SPATIAL NAVIGATION

The male advantage in spatial navigation can be expressed in a variety of ways: they make fewer errors, take less time and use their time more efficiently both in real-world (e.g., Malinowski & Gillespie, 2001; Silverman et al., 2000) and virtual environments (e.g., Astur et al., 1998; Astur, Tropp, Sava, Constable, & Markus, 2004; Tascon, Leon, & Cimadevilla, 2016). However, the gender difference in spatial navigation is perhaps best illustrated in how each gender prefers to give and receive directions: it has been well documented that men are more likely to use distances and cardinal directions, whereas women are more likely to use landmark information (e.g., L. N. Brown, Lahar, & Mosley, 1998; Dabbs, Chang, Strong, & Milun, 1998; Galea & Kimura, 1993; Lawton, 2001; MacFadden, Elias, & Saucier, 2003). In an eye-tracking study, women fixated more on landmarks than did men and performance differences were only seen when landmarks were missing (N. E. Andersen, Dahmani, Konishi, & Bohbot, 2012). The likely reason for this is that men and women pay attention to and use different forms of spatial information, and therefore use different strategies. It follows that they tend to produce different kinds of spatial representations.

With regards to the types of spatial knowledge (i.e., landmark, route, graph, and survey knowledge), the findings discussed to this point suggest that women tend to focus more on landmark and route knowledge. This is because research shows that they encode landmark information from an egocentric first-person perspective (for a review, see Coluccia & Louse, 2004). Conversely, men focus more on graph and survey knowledge, as they are better able to transform the more configural global information into an allocentric representation, like a cognitive map (for a review, see Coluccia & Louse, 2004). This ability to transform knowledge into an allocentric representation could be what is giving men their performance edge (Tascon et al., 2016). In contradiction to this, it has been suggested that men tend toward a more egocentric reference frame when undertaking mental rotation tasks, which gives them an advantage (Bosco, Longoni, & Vecchi, 2004; Linn & Petersen, 1985; Voyer et al., 1995). Similarly, the preferential use of a landmark-based frame of reference in women may explain the female advantage in object-location memory tasks found in numerous studies (e.g., Andreano & Cahill, 2009; De Goede & Postma, 2008; Lejbak et al., 2009; Levy et al., 2005; Silverman et al., 2007; Spiers et al., 2008).

Most of the literature on gender differences in spatial navigation revolves around the differences in the strategies men and women use and in their spatial representations. For instance, in the presence of both proximal and distal information, men perform better. However, when distal information and cardinal directions are made more salient, the gender difference effectively disappears, likely because women make use of this information as well (Ward, Newcombe, & Overton, 1986). A more recent study found a gender difference in environments with only distal cues (similar to the Morris Water Maze), whereas differences were not found in environments that also contained proximal cues (Rahman, Abrahams, & Jussab, 2005). This demonstrates that women are not impaired in the presence of proximal cues but only non-salient distal cues. This could be because distal cues, such as those seen in the Morris Water Maze, require an allocentric perspective, which men have a greater tendency to use, while proximal cues require an egocentric perspective, which women have a greater tendency to use. Moreover, women and men are equally successful when provided with landmark cues alone, but men perform better in the presence of geometric cues alone, as well as when only one distant landmark is available (Kelly & Bischof, 2005). This could be because, compared to men, women take more time to choose and have more difficulty when using the appropriate strategy (Astur, Purton, Zaniewski, Cimadevilla, & Markus, 2016). Indeed, when women do choose an allocentric strategy, they are not as successful as men are at using it (van Gerven, Schneider, Wuitchik, & Skelton, 2012). The idea that men are better than women at taking a third-person perspective (or top-down perspective) was further corroborated by Kaiser et al. (2008) and Tascon et al. (2016), in addition to males outperforming females in the speed of perspective transformation (Gardner, Sorhus, Edmonds, & Potts, 2012). The general consensus is that landmarks play a more critical role in navigation for women than for men because proximal cues are necessary for the default navigation strategy typically used by women (L. F. Jacobs & Schenk, 2003).

Strategy differences could be explained by the general spatial cognition differences discussed above. In fact, it has been proposed that the ratio of visual-spatial to verbal working memory capacity may determine the spatial strategy difference between the genders (L. Wang & Carr, 2014). That is, better VSWM compared to verbal working memory would help select more effective holistic spatial strategies, whereas superior verbal working memory relative to VSWM would help select less effective analytic strategies. Therefore, since women have better developed verbal working memory whereas men have better developed visual-spatial working memory, these differences in cognitive processing alone could dictate the type of navigation strategy favored by each gender. Moreover, spatial abilities but not verbal memory contributed to route learning for boys, while girls used both spatial abilities and verbal memory (Merrill, Yang, Roskos, & Steele, 2016). This difference may reflect the precursor of the strategy difference developed into adulthood. Further support for the spatial-verbal divide between men and women comes from EEG studies. For instance, during a virtual wayfinding task, while men showed increased activation in spatial working memory regions in the right hemisphere, women had increased activation in the verbal-analytical regions of the left hemisphere (Ramos-Loyo & Sanchez-Loyo, 2011).

Additional evidence for gender differences in navigation comes from studies showing that men may have better cognitive maps than do women. For instance, men are more accurate when aligning sketch maps with cardinal directions (Harrell, Bowlby, & Hall-Hoffarth, 2000). They are also more accurate at pointing in the direction of hidden targets (e.g., Fields & Shelton, 2006; Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Holding & Holding, 1989; Lawton, 1996; Lawton & Morrin, 1999; Prestopnik & Roskos-Ewoldsen, 2000; Waller, Knapp, & Hunt, 2001), which may indicate the use of an allocentric cognitive map that allows one to understand the configuration of unseen targets. Men also perform better than women on wayfinding tasks in which maps are used (e.g., Devlin & Bernstein, 1997; Malinowski & Gillespie, 2001; Tlauka, Keage, & Clark, 2005) and are more accurate than women on tests of map reading skills (e.g., G. L. Allen, 2000; Chang & Antes, 1987; Coluccia, Bosco, & Brandimonte, 2007; Gilmartin, 1986; Henrie, Aron, Nelson, & Poole, 1997). It is important to note that there is no evidence to support that one gender is better than another at cognitive mapping per se, only that males and females differ in the way they use various types of spatial information in order to build spatial representations (e.g., De Goede & Postma, 2008; Lawton, 1994, 2001; Lawton & Kallai, 2002; Postma, Izendoorn, & De Haan, 1998). However, it seems the information that men generally use to build their cognitive maps is likely more efficient and effective than that used by women, which explains why they are less susceptible to alignment effects (Roskos-Ewoldsen, McNamara, Shelton, & Carr, 1998).

Looking at neurological differences, many cortical areas show similar patterns of neural activation between women and men during navigation (e.g., R. J. Blanch, Brennan, Condon, Santosh, & Hadley, 2004). However, the few differences that are found show men to have more activity in the left hippocampus as well as in the right parietal cortex (e.g., Gron et al., 2000; Gur et al., 2000). The left hippocampus is known to have a role in representing external space (Nadel & MacDonald, 1980), whereas the right parietal cortex, as mentioned above, is thought to be involved in egocentric perspective taking. Women have been shown to have greater activation in the right prefrontal and parietal cortices, which are involved in working memory when doing spatial tasks (e.g., Gron et al., 2000; Jonides et al., 1993). In addition, women have smaller parietal lobe surface area compared to men and this difference was related to mental rotation task performance (Koscik, O'Leary, Moser, Andreasen, & Nopoulos, 2009). These findings are likely related to how men tend to outperform women on perspective-taking tasks and women on object-location memory tasks. Generally speaking, activation patterns in women demonstrate more effortful deliberate verbal-analytic processing, whereas activation patterns for men demonstrate more automatic processing, reflecting more visuomotor strategies (e.g., Butler et al., 2006; Heil & Jansen-Osmann, 2008; Hugdahl, Thomsen, & Ersland, 2006). However, it is not possible from these findings to determine the direction of the causal relationship; that is, whether neural activation differences are a result of or a cause of differences in strategy use.

Despite some inconsistent findings on gender differences in cortical activation, cognitive studies have demonstrated clear differences between men and women when it comes to navigation behaviors and performance. Generally, women have a greater dependency on landmark information as part of their default route-based navigation strategy. In the absence of landmark information, women have greater difficulty using allocentric cues as part of an orientation-based strategy. In contrast, men are better able to use more stable allocentric cues which help to keep their orientation-based navigation strategy more flexible and holistic, whether landmark information is present or not. By acknowledging these differences, we can now investigate the underlying factors that may be contributing to these differences and discuss what can be done to improve women’s navigation performance.

Factors Underlying Gender Differences in Spatial Navigation

As with most complex human abilities, spatial cognition arises from an interaction of inherited and environmental factors. The reduction in gender differences in spatial cognition performance over the course of the last few decades almost certainly reflects changes in environmental factors as society becomes more overtly egalitarian. Conversely, the fact that these differences remain quite robust and vary in magnitude across various spatial tasks suggests there may also be a less changeable biological cause. For instance, the hunter-gatherer theory suggests that men were evolutionarily selected to have greater spatial cognition skills in order to successfully navigate and explore the larger natural environment and hunt. Correspondingly, women were evolutionarily selected to remember where things like fruits and nuts were located near camp, thus explaining why object-location memory is the only spatial task on which women outperform men (Silverman & Eals, 1995). In addition, research has found that girls and boys who favored activities such as climbing, building with blocks, or playing with toy vehicles demonstrated greater spatial ability scores than did girls and boys who favored playing with dolls or housekeeping toys (Serbin & Connor, 1979). There is a broad consensus that both effects of nurture and nature are at play (Halpern, 2013). Indeed, numerous biological and sociocognitive factors have been posited to explain the existence of gender differences in spatial abilities (for reviews, see S. C. Levine, Foley, Lourenco, Ehrlich, & Ratliff, 2016; Schug, 2016). The most important factors are discussed below.

Biological Factors

One of the strongest arguments supporting the effects of biological factors is that the gender difference favoring men on most spatial tasks is highly consistent across cultures world-wide (e.g., Jahoda, 1980; Lynn, 1992; Mann, Sasanuma, Sakuma, & Masaki, 1990; Porteus, 1965; Silverman, Phillips, & Silverman, 1996). Moreover, it is seen across species, though research is mostly with rodents (e.g., Barrett & Ray, 1970; Binnie-Dawson & Cheung, 1982; Gaulin & Fitzgerald, 1986; Joseph, Hess, & Birecree, 1978; Shore et al., 2001; Williams & Meck, 1991).

Several behavioral genetic theories were proposed in the 1970's and 1980's to try to explain the gender difference in spatial abilities (for reviews, see Henderson, 1982; Pezaris & Casey, 1991). However, although spatial test score variation appears to be moderately heritable (e.g., Bouchard, Segal, & Lykken, 1990; DeFries et al., 1976; Plomin et al., 1994; Tambs, Sundet, & Magnus, 1984; Vanderberg, 1969) there have been very few studies to support any of these theories (e.g., Boles, 1980; Corley, Defries, Kuse, & Vandenberg, 1980; DeFries, 1980; DeFries et al., 1976; Fralley, Eliot, & Dayton, 1978; Guttman & Shoham, 1979; McGee, 1979, 1982). In the meantime, recent research has shown the importance of taking into consideration differences in phenotypes caused by sex chromosome complements (XX vs XY) and their interactions with hormones (e.g., Arnold & Chen, 2009; McCarthy, Arnold, Ball, Blaustein, & De Vries, 2012).

There are several hypotheses regarding biological causes of gender differences in spatial navigation, and they mainly focus on the effects of prenatal and postnatal sexual hormones (e.g., Buchsbaum & Henkin, 1980; Imperato-McGinley, Pichardo, Gautier, Voyer, & Bryden, 1991; McGee, 1979; Resnick, Berenbaum, Gottesman, & Bouchard, 1986; Williams, Barnett, & Meck, 1990; Williams & Meck, 1991). Studies using animals models have suggested that prenatal androgens might promote an acceleration in the development of the magnocellular layer of neurons in the brain (which provides input to the dorsal, object-location and motion stream) relative to the development of the parvocellular layer (which provides input to the ventral, object recognition stream) (Alexander, 2003; Bachevalier & Hagger, 1991; Lauer, Udelson, Jeon, & Lourenco, 2015; Salyer, Lund, Fleming, Lephart, & Horvath, 2001). According to this hypothesis, as male infants are exposed to more androgens compared to female infants, the development of their dorsal streams relative to their ventral stream would be favored, which would lead to early structural neurological differences (Lauer et al., 2015). These structural differences, which exist from a young age, may explain early toy preference differences, with young boys preferring objects with moving parts such as trucks and may eventually lead to enhanced mental rotation ability (Lauer et al., 2015).

In addition to prenatal hormonal differences, there is strong evidence for the role of postnatal hormones in gender differences in spatial ability. For instance, women with high testosterone levels have more accurate graph knowledge (e.g., Allen, Kirasic, Dobson, Long, & Beck, 1996; Burkitt, Widman, & Saucier, 2007). Correspondingly, performance using Euclidean information while navigating a virtual maze has been found to be negatively correlated with estrogen levels across women’s menstrual cycles (Chabanne, Peruch, & Thinus-Blanc, 2004). Interestingly, while males with androgen deficiency early in development exhibit a variety of visuospatial impairments (Buchsbaum & Henkin, 1980), males who acquire androgen deficiency after puberty do not (Slabbekoorn, van Goozen, Megens, Gooren, & Cohen-Kettenis, 1999). Altogether, these findings suggest differences between organizational and activational effects between prenatal and postnatal hormones. In brief, variations in prenatal and postnatal hormone exposure could explain the variation in spatial task performance via differences in cortical activation and neuroanatomy between the dorsal and ventral streams across gender.

Sociocognitive Factors

While there is no single sociocognitive factor underlying the effects of environment on gender differences in spatial cognition, there is a long list of potential factors all revolving around the ideas of differences in socialization (e.g., Baenninger & Newcombe, 1989), and gender role identification (e.g, Nash, 1975; Signorella & Jamison, 1986).

There are several studies showing differences in socialization at different developmental stages favoring boys' development with spatial tasks (for a review, see S. C. Levine et al., 2016). Namely, studies have shown that the type of language parents use with boys and girls differs and that this is related to spatial performance. For instance, parents provided boys with significantly more spatial language than girls for both easy and difficult puzzles (Pruden & Levine, 2017). In addition to teaching them to attend to spatial information on a regular basis, the greater emphasis on spatial language for boys may help to better develop their spatial thinking by leading them to more spatial information, thus reducing cognitive load on the task at hand and allowing for more sophisticated processing to occur (Pruden & Levine, 2017). Indeed, this was supported by a study showing that young children who heard a greater proportion of spatial language from their parents used a greater proportion themselves and performed better on a mental rotation task (Pruden, Levine, & Huttenlocher, 2011). Along this same line of research, a study found that compared to young girls, young boys received more high quality puzzle play, which was defined by a composite of puzzle difficulty, parent engagement, and parent spatial language (S. C. Levine, Ratliff, Huttenlocher, & Cannon, 2012).

Interestingly, the mere presence of boys could also change the way girls perform on spatial tasks. For instance, girls from single gender schools outperform girls from co-ed schools and show performance similar to boys (Titze, Jansen, & Heil, 2011). The authors speculate this could be because girls from single-gender schools are not as exposed to gender stereotypes (Titze et al., 2011). Indeed, children of both genders have been found to believe the gender stereotype favoring males in mental rotation, showing how stereotypes can develop early (Neuburger, Ruthsatz, Jansen, & Quaiser-Pohl, 2015).

Generally speaking, engaging in typical "boy" activities, such as playing with blocks or other construction toys, as well as playing video games, is associated with higher performance on spatial tasks, including spatial visualization tasks (e.g., Caldera, Huston, & O'Brien, 1989; Nazareth, Herrera, & Pruden, 2013; Ruthsatz, Neuburger, Jansen, & Quaiser-Pohl, 2015; Terlecki & Newcombe, 2005). Indeed, boys are more likely to engage in these activities, which involve spatial manipulations and transformations and part-whole thinking, compared to girls at home and at school (e.g., Saracho, 1995; Tracy, 1987). However, it is unclear how much this difference in play activities is driven by early intrinsic interests compared to socialization such as parent/teacher encouragement reflected by societal stereotypes.

In addition, there are social explanations that increase the spatial experience of boys over girls. For instance, boys are generally allowed more freedom than girls are to travel far from home (e.g., Herman, Heins, & Cohen, 1987; Newson & Newson, 1987). Consequently, at about the age of 8, when boys' home ranges become larger, their sketch maps of areas surrounding their homes become more integrated, accurate, and richer in detail than those of girls the same age (Matthews, 1986). Indeed, less childhood wayfinding experience was associated with higher levels of spatial anxiety and lower spatial aptitude in adults (Hoang, Mora, Vieites, Reeb-Sutherland, & Pruden, 2017).

It is important to note that the argument for differences in socialization include any differences in experience/nurture across the lifetime - not just developmentally. That is, there are several studies showing differences in socialization among adult women and men such that women generally have less exposure to and experience with spatial tasks in their everyday lives compared to men. Even in adulthood, men are more likely to have a larger range of travel than are women (i.e., range of travel for work, leisure, and errands), which was particularly associated with men's small-scale spatial abilities such as mental rotation and object-location memory (Ecuyer-Dab & Robert, 2004). Additionally, several studies have shown that men have more expertise in motor skills, such as athletics (Eccles & Harold, 1991) and skilled trades (Denissen, 2010). Indeed, a recent meta-analysis determined that individuals with an expertise in motor skills, such as athletes, have significantly greater spatial abilities than those without motor expertise (Voyer & Jansen, 2017). In addition, there are generally fewer female drivers on the road than there are male drivers (Sivak, 2013), and males are more likely to drive when women are present (Ozkan & Lajunen, 2006). Several studies have also found confidence to be a better predictor than gender for spatial performance, and women may be less confident due to having less spatial experience (Estes & Felker, 2012). These differences in socialization in adulthood only further exacerbate any spatial cognition differences between men and women that may have already developed throughout childhood. Indeed, it is possible that these spatial behavior differences are due to differences in biological skill or interest. However, it is undeniable that there are strong social influences steering girls' and women's interest away from spatially oriented tasks, and therefore their cognitive development in these tasks.

Many studies have found certain cognitive tasks to be considered gendered, meaning that certain tasks are socially “reserved” for one gender – relating to gender roles. For instance, spatial skills are rated as being more masculine and participants believe men partake in these types of tasks more often than do women (Newcombe, Bandura, & Taylor, 1983). Similar results were found in the case of beliefs about memory for spatial information (Crawford, Herrmann, Holdsworth, Randall, & Robbins, 1989). Interestingly, expectations based on the gendered nature of a task seem to affect performance. This was found to be true using a mental rotation task, which, as mentioned above, is known for having the most robust gender difference of all the spatial tasks. Sharps, Welton, and Price (1993) replaced the standard instructions for the task with a paragraph that either accentuated or de-emphasized the spatial nature of the task. Interestingly, the gender difference was eliminated when the spatial nature of the task was de-emphasized. In addition, Sharps, Price, and Williams (1994) used the same instructional manipulation for a mental rotation task with items differing in difficulty. They only found a gender difference for the most difficult type of item (3D block figures compared to 2D) and only under spatial instructions. In their second experiment, the same 3D task was presented as being predictive of a person’s aptitude in either interior design (“feminine” instructions) or combat aircraft flying (“masculine” instructions). Similarly to previous experiments, men given “masculine” instructions performed better than men or women in any other conditions (Sharps et al., 1994). All together, these studies suggest that gender differences in mental rotation tasks are potentially the product of the participants’ gender-based expectations about their own performance, which leads to a self-fulfilling prophecy when faced with a gendered task. This may also be evidence of the effects of stereotype threat (for more on stereotype threat see corresponding below subsection).

Although very little can be done to improve girls' and women's spatial cognition by changing biological factors, there are several sociocognitive factors that can be changed. While early findings were used as a justification for the underrepresentation of women in STEM (Science, Technology, Engineering, and Mathematics) fields (e.g., Fennema, 1975; Maccoby & Jacklin, 1974; Sherman, 1967), researchers more recently have argued that the discussion would be better served by focusing on the sociocognitive factors preventing women from entering these fields (e.g., Hines, 2007; Newcombe, 2007). There are several hypotheses about how this can be done including developing training programs to improve girls' and women's spatial skills, understanding and thwarting the effects of stereotype threat, and understanding and alleviating women's low self-confidence regarding these skills in order to prevent girls' loss of interest in science (S. C. Levine et al., 2016; Miller, Eagly, & Linn, 2015). Indeed, two meta-analyses on the efficacy of training programs for improving spatial ability found that women's scores could in fact be improved to match those of men (Baenninger & Newcombe, 1989; Uttal et al., 2013). The following subsections will cover gender differences in confidence and anxiety during spatial tasks as well as how gender stereotypes may be related to differences in spatial task performance.

GENDER DIFFERENCES IN SPATIAL CONFIDENCE AND METACOGNITON

Generally speaking, both men and women tend to display overconfidence across many domains (for a review, see Moore & Healy, 2008). Despite this effect of overconfidence, a large body of scientific literature demonstrates that women, on average, are less overconfident and more risk averse than men (e.g., Hardies, Breesch, & Branson, 2013). The gender difference in overconfidence has been found to be highly task dependent (Lundeberg, Fox, & Punccohar, 1994) with differences at their greatest for stereotypically masculine tasks (Beyer & Bowden, 1997). In a recent study, boys had higher math self-confidence than equally able girls (Parker, Van Zanden, & Parker, in press). That is, while the performance gap in math has closed, girls' self-confidence has stayed the same. Other studies have shown that women underestimate their academic performance in a variety of fields from economics to medicine (e.g., D. C. Blanch, Hall, Roter, & Frankel, 2008; Jakobsson, 2012; Jones & Jones, 1989; Langan et al., 2008). Many researchers have also made a distinction between overconfidence, which is when individuals overestimate their own performance, to overplacement (Moore & Healy, 2008) which is when individuals overestimate their performance relative to others. A recent study found that while both genders were equally overconfident, men overplaced themselves relative to women and other men significantly far more than women did (Ring, Neyse, David-Barett, & Schmidt, 2016).

Additionally, women underestimate their performance on spatial tasks. For instance, there is a persistent gender difference in confidence on the mental rotation task (e.g., Cooke-Simpson & Voyer, 2007; Cross, Brown, Morgan, & Laland, 2017; Estes & Felker, 2012). Women have also been found to exhibit low spatial confidence in a cognitive mapping task (O'Laughlin & Brubaker, 1998), as well as in a wayfinding task (Lawton et al., 1996). In addition, Picucci et al. (2011) demonstrated that men were significantly more confident in their performance on a spatial reorientation task compared to women with a moderate effect size, despite only finding a small effect for gender difference in task accuracy. This study supports research that has found that gender differences in spatial performance are affected by self-confidence (e.g., Cooke-Simpson & Voyer, 2007; Lawton & Kallai, 2002; Moe & Pazzaglia, 2006; Wraga et al., 2006). The authors argue that this could be because women have a lower tolerance for uncertainty about spatial information. This idea is supported by a study that found that men exhibit more guessing behaviors than women (Baldiga, 2014). This has also been shown to be true at a young age, where boys are less likely than girls to select "I don't know" when they are uncertain (Linn & Petersen, 1985).

Many studies have suggested that this confidence difference favoring men could help to explain strategy differences between the genders. For instance, confidence positively predicted allocentric strategy preference in older adults, perhaps demonstrating a reluctance to use memory-based strategies (Ariel & Moffat, 2017). This is in addition to research that has also suggested differences in spatial anxiety could explain performance and strategy differences between the genders (e.g., Lawton, 1994; Lawton & Kallai, 2002; Nowak et al., 2015). For instance, spatial anxiety was positively related to the use of a route strategy when wayfinding (Hoang et al., 2017). One possible explanation for this could be that a route-based navigation strategy requires less VSWM which anxiety has been shown to negatively impact (e.g., Ganley & Vasilyeva, 2014; Moran, 2016; Shackman et al., 2006; Vytal, Cornwell, Letkiewicz, Arkin, & Grillon, 2013). Thus, with anxiety-induced impaired VSWM, women may not be able to perform an orientation-based navigation strategy.

A recently proposed evolutionary hypothesis suggests that men and women navigate differently due to fitness costs associated with long-distance travel. This is thought to explain the gender difference in mobility, specifically, why women have greater levels of spatial anxiety and hesitant wayfinding behavior (e.g., Cashdan & Gaulin, 2016). A recent study supports this idea, finding that women generally take a more cautious approach to spatial exploration, and suggested that harm avoidance may partially explain their hesitant spatial behavior as well as their tendency towards a different strategy compared to men (Gagnon, Cashdan, Stefanucci, & Creem-Regehr, 2016).

The topic of gender differences in local metacognition is a relatively unexplored one. One study that investigated metamemory and age across a variety of tasks demonstrated that, irrespective of the task, women generally underestimated their performance more than men (Hertzog et al., 1990). Lundeberg et al. (1994), examining performance on a psychology exam, found that women showed more accurate metacognition when evaluating their incorrect answers than did men, who tended to show more overconfidence when wrong.

There is only one study that specifically used the Nelson and Narens (1990) framework with a map and verbal direction-learning task; though a gender difference analysis was not the main goal of the paper (Schwartz, 2006). The preliminary analyses showed neither age nor gender predicted performance outcomes, which justified excluding them from all analyses. Regardless, the results showed very strong gamma correlations, much higher than most other immediate JOL paradigms; therefore, gender differences might have been masked by a ceiling effect (refer to Chapter 2 for more details on this study). Schwartz explains that these high values may be due to his tasks having high ecological validity. Considering that people typically have much more experience receiving directions and using maps than learning word-pair associates, as is typically used in metacognition research, it is possible our metacognitive skills are better developed for spatial tasks. Consequently, our expertise with navigation might result in superior monitoring and control mechanisms compared to what has been found with verbal stimuli. Along similar lines, it has been found that experts demonstrate relatively superior monitoring and control abilities compared with novices across a variety of tasks (e.g., Gobet, 1998; Hallam, 2001; Martini & Shore, 2008). To the extent that we have a greater expertise with navigation tasks than word-pair associates, we would expect high gamma correlations using the former compared to the latter. This expertise effect also leads to a prediction regarding gender differences in spatial metacognition. That is, if men have greater experience with navigation tasks than women (e.g., Lucas & Sherry, 2004) then we would expect them to have more accurate metacognitive performance as well.

Some studies have asked general metacognitive questions regarding spatial abilities or have asked participants to explicitly state the strategy used. Women self-report having a strategy that emphasizes attention to landmarks and associated route turns, while men report that they orient to global cues and develop a top-down perspective of the larger environment (e.g., Charleston, 2008; Lawton, 1994, 1996; Lawton & Kallai, 2002). These results corroborate the behavioral literature on gender differences in spatial strategies covered above. Women self-report having lower spatial skills (e.g., Hegarty et al., 2006), as well as greater levels of spatial anxiety (Lawton, 1994; Lawton & Kallai, 2002). In this context, it is important to note that spatial confidence can alter navigational behavior, which, in turn, could lead to inflated gender differences in performance (Lavenex & Lavenex, 2010; Lawton, 1996). For instance, a lack of confidence might increase exploratory behaviors, which would cause longer latencies and perseveration errors (Lavenex & Lavenex, 2010). Though it is possible women's lack of confidence stems from a lack of ability, the many studies cited above have found that women exhibit lower levels of confidence even when they perform as well as men (e.g., Beyer & Bowden, 1997; O'Laughlin & Brubaker, 1998; Parsons, Meece, Adler, & Kaczala, 1982). Moreover, self-reported confidence has been shown to be influenced by factors other than ability or performance, including gender stereotypes (Oswald, 2008).

Evidence from the Stereotype Threat Literature

Stereotype threat is a phenomenon whereby an individual's performance is hindered simply due to an awareness of a negative stereotype concerning the (ostensibly) poor aptitude of a group they belong to (Steele & Aronson, 1995). There is evidence to suggest that gender differences in spatial performance are subject to stereotype threat. Indeed, there is a stereotype that men possess superior spatial skills, which is reflected in how men report higher expectations for success on spatial tasks (Meehan & Overton, 1986). Moreover, a recent study, comparing a variety of gender stereotype beliefs found that the magnitude in current stereotyped beliefs have not changed since 1980 (Haines, Deaux, & Lofaro, 2016). Additionally, both genders report believing that men are better at finding a previously visited place as well as remembering directions (Crawford et al., 1989). In a classic series of studies, Sharps and colleagues demonstrated that gender differences in performance on spatial tasks is influenced by manipulations that emphasize masculine or feminine stereotypes (Sharps et al., 1994; Sharps et al., 1993). For instance, they found that women performed worse on an object-location memory test when the title of the stimuli was presented as a "map" (stereotypically masculine) compared to a "model" (Sharps et al., 1993). Interestingly, simply telling participants that their gender scores higher on a spatial task can increase their performance (Wraga et al., 2006). Moreover, women with feminine gender role beliefs performed better when a spatial visualization task was said to be a test of empathetic ability (stereotypically feminine), whereas women with masculine gender role beliefs performed better when the same task was said to be a test of spatial ability (Massa, Mayer, & Bohon, 2005).

There are a few studies that have successfully manipulated performance on a spatial task (mostly mental rotation) via various stereotype activation manipulations (e.g., Hausmann, Schoofs, Rosenthal, & Jordan, 2009; Heil, Jansen, Quaiser-Pohl, & Neuburger, 2012; Moe, 2009, 2012; Wraga, Helt, Jacobs, & Sullivan, 2007). For example, a study by Moe and Pazzaglia (2006) tested participants' mental rotation skills before and after a stereotype facilitation instruction that was either "males do better" (male facilitation condition), "females do better" (female facilitation condition) or "they perform equally" (control condition). Women in the female facilitation group showed the greatest pre to post-test improvement in performance, and next were the men in the male facilitation group. Correspondingly, women in the male facilitation group demonstrated the largest pre to post-test decrease in performance, and men in the female facilitation group demonstrated the second largest pre to post-test decrease. In another related study, Moe (2009) found that performance in mental rotation did not differ between the genders in the female facilitation group. In a recent study, this effect was also seen among children where implicit gender stereotype activation had the effect of reducing the gender difference in mental rotation (Neuburger et al., 2015). However, the male advantage persisted for the difficult version of the mental rotation in both the stereotype threatening and the non-threatening conditions (Neuburger et al., 2015). Overall, this suggests, consistent with the effects of stereotype threat that women under the effects of stereotype threat tend to do worse than control groups on mental rotation tasks. That is, they do worse than when there is no stereotype threat manipulation at all (e.g., Hausmann et al., 2009; Moe, 2009; Wraga et al., 2007). This may simply be because being made aware that they are in a gender study induces stereotype threat. Interestingly, stereotype facilitation and self-affirmation have been shown to improve performance (e.g., Hausmann et al., 2009; Moe, 2009; Wraga et al., 2007), a phenomenon known as stereotype boost or lift (Walton & Cohen, 2003). However, it is important to note that studies have shown that men and women react differently to stereotype activation, with men, for instance, showing improved verbal fluency scores during stereotype threat compared to control (Hirnstein, Freund, & Hausmann, 2012). An additional study, from the same research group, suggested that it is difficult to induce stereotype threat and lift simultaneously when tests favoring both men and women are used (Hirnstein, Andrews, & Hausmann, 2014).

More recent work has put into question the idea of stereotype threat in general. For instance, a recent meta-analysis investigated the effects of stereotype activation in the context of gender differences in math and spatial tasks (Doyle & Voyer, 2016). It concluded that, among spatial tasks, there was no consistent effect of stereotype threat in either men or women, and only a small effect of stereotype lift on women only. However, this should be interpreted with caution, as the authors were only able to analyze 18 manuscripts, most of which regarded mental rotation and none of which used a navigation/wayfinding task. Thus, it is difficult to know how these results relate to navigation, which may very well be a spatial task with a more pervasive stereotype. Another meta-analysis looked at the effects of stereotype threat on girls and grouped studies using math, science, and spatial tasks. They found an average effect size of .22 (Flore & Wicherts, 2015). It is important to note that both meta-analyses found evidence of a publication bias. Overall, it is unclear how stereotype threat or boost may influence the gender difference in navigation, as there is only one stereotype study to specifically use a navigation task. Rosenthal, Norman, Smith, and McGregor (2012) had students learn to navigate to a hidden goal using geometric and/or landmark-based environmental cues under different stereotype activation conditions. Under stereotype activation, men’s performance was boosted compared to control for tasks with both types of cues but women’s performance was generally unaffected.

The way stereotype threat was originally thought to work is as follows. First, a reminder of one's group affiliation and its ostensible performance deficit creates a heightened sense of anxiety in a person as they attempt to suppress stereotype related thoughts. This in turn narrows the individual's attention and hinders their ability to perform (Steele & Aronson, 1995). Accordingly, stereotype boost/lift is thought to stem from a decrease in self-doubt (Walton & Cohen, 2003). More recently, Schmader, Johns, and Forbes (2008) established a process model of stereotype threat that identifies three interconnected factors. Their model links stereotype threat to a physiological stress response as well as in increase in performance monitoring processes; these work in a vicious cycle with suppression processes to suppress negative thoughts and emotions based on one's concept of self. That is, in an effort to avoid failure, and under the influence of stereotype threat, one's mind switches from an automated state of functioning into a more conscious and controlled state of monitoring the self within the situation. This in turn creates an additional cognitive load, hindering working memory and performance. While Schmader et al. (2008) do not explicitly mention metacognition in their model; one can see their theory has implications for how stereotype threat may be linked to metacognition by influencing performance monitoring. In addition, an increase in cognitive load and a shift from automated to deliberate performance monitoring, implied by their model, would be expected to hinder metacognitive processes as well as cognitive ones.

CHAPTER SUMMARY

A very large body of research shows gender differences in a variety of spatial skills, with the most robust difference seen in mental rotation tasks. Beyond this, a variety of other gender differences that may partially explain the gender difference in spatial cognition have also been found, including differences in socialization, confidence, spatial anxiety, and the effects of stereotype threat. Although females' and males' spatial skills can be effectively manipulated through stereotype instruction protocols, further research is needed to determine whether this type of manipulation can eliminate gender differences. We have found that metacognition may be related to stereotype threat based on the Schmader et al. (2008) model. While little research has looked into possible gender differences in metacognition, there is some evidence of a difference specifically related to performance expectations, which may in turn be informed and influenced by gender stereotypes. This demonstrates a relatively important gap in the literature; gender differences in spatial navigation might, at least in part, be explained by metacognitive differences brought on by stereotype threat.

CHAPTER 4: OVERVIEW OF EXPERIMENTS

GENERAL PURPOSE

The goal of the present thesis work is to determine if gender differences exist in local metacognitive accuracy and confidence in the context of a spatial navigation task, and if this is related to the well-established gender differences in spatial navigation performance. That is, the present research aims to determine whether men possess more accurate metacognition and whether this is related to them outperforming women in navigation. A related goal is to examine whether local metacognitive judgments and navigation performance are affected by stereotype threat/lift. The present findings could not only help to inform the theoretical basis for this gender difference but could also have practical implications. For instance, the results from this research could help inform spatial cognition training programs that include metacognitive aspects for girls and women or education reform targeting improved spatial metacognition in math and science classes. An ancillary goal of this research is to investigate the feasibility of applying the Nelson and Narens (1990) metamemory framework to a navigation task setting, in order to expose and help structure this new research direction.

Three experiments were conducted to assess gender differences in local metacognition during a first-person virtual maze navigation task. The first two experiments were designed to examine if, when applying the Nelson and Narens (1990) metamemory framework to our navigation task, we indeed see gender differences in navigation performance and/or local metacognitive accuracy. The third experiment involved manipulating confidence via a stereotype activation procedure in order to compare how stereotyped instructional facilitation affects local metacognition and navigational performance, and whether these effects differ between men and women.

GENERAL PROCEDURE

All three experiments follow the Nelson and Narens (1990) framework, in which metacognition is operationally defined by the accuracy of trial-by-trial self-assessment judgments in relation to future and past performance. All three experiments involved participants learning maze layouts from a first person perspective and then attempting to navigate from the entrance to the exit of those mazes via the shortest possible route; similarly, in all three experiments, participants were asked to provide monitoring judgments between trials. Monitoring judgment instructions were adapted to the navigation task with the intention of preserving as much comparability as possible with traditional metacognition methods. According to the framework, local metacognition is typically evaluated using relative and absolute accuracy between monitoring judgment scores and performance. Both of these methods were adapted to the current task, and both are described and discussed in each experiment.

Two prospective metacognitive monitoring judgments (Ease-Of-Learning and Judgment Of Learning) and one retrospective judgment (Retrospective Confidence Judgment) have been adapted to suit a navigation task (see “Materials” section under each study for adapted metamemory instructions). While most studies have usually solicited only one type of metacognitive judgment per trial, others have solicited multiple judgments for each trial (e.g., Busey et al., 2000; de Carvalho & Yuzawa, 2001; Dougherty et al., 2005; Leonesio & Nelson, 1990; Modirrousta & Fellows, 2008b; Watier & Collin, 2010, 2011, 2012). In an attempt to achieve a more complete understanding of metacognitive monitoring during spatial navigation, two judgments were solicited for the first two experiments: one prospective (EOL in Experiment 1, JOL in Experiment 2) and one retrospective (RCJ). All three were solicited for the third experiment. Although the literature suggests that each judgment measures something different and that they are independent from one another (e.g., Busey et al., 2000; Dougherty et al., 2005; Modirrousta & Fellows, 2008b; Pannu & Kaszniak, 2005) it remains possible that eliciting one kind of judgment will have an effect on another. To test if this is the case, we calculated the correlations between monitoring accuracy scores in each experiment. We hypothesized that the correlations would be low, reflecting the independence of the judgments.

In all experiments, participants navigated through thirty virtual mazes, and an error was recorded whenever they crossed invisible "trip wires" that were strategically placed in dead ends (Shore et al., 2001). Participants navigated using the arrow keys on a typical computer keyboard. Video tours of the same 30 mazes were created to be used as learning stimuli. These were either shown in a series or paired with maze performances, or a combination thereof, dependant on the needs of the monitoring judgment. A prospective monitoring judgment was taken after each video tour, either an EOL or JOL, and a retrospective monitoring judgment was taken after each maze performance.

In the first and second experiments, the maze task was programmed in C++ using OpenGL and video tours were created by hand via video screen capture. For these experiments, the experimenter was in the room and recorded metacognitive judgment scores, coordinated the video tours to display on Windows Media Player, as well as launched the maze program when appropriate. However, for the third experiment, instructions, maze performances, video tours, and metacognitive judgments were all integrated into the same program in C++ using OpenGL. The third experiment also included, at the end in paper and pen format, the Santa Barbara Sense of Direction Scale (SBSOD; see Appendix A; Hegarty et al., 2002), and the Spatial Anxiety Scale (SAS; see Appendix B; Lawton, 1994). More details on each scale are provided in the description of Experiment 3.

What most distinguishes Experiment 3 from Experiments 1 and 2 is the stereotype threat/lift instructional facilitation protocol. While there was no mention of gender differences in Experiments 1 and 2, there were three Stereotype Facilitation conditions in Experiment 3, where different deceptive study introductions were given: 1) Pro Male facilitation condition: “men outperform women on this particular task”, 2) Pro Female facilitation condition: “women outperform men on this particular task”, and 3) Neutral facilitation condition: “the genders perform equally on this task” (see each experiment for more specific task instructions). Data collection for Experiment 3 was done at the INSPIRE lab, collecting data from up to four individuals at a time, whereas data collection for Experiments 1 and 2 was done in a single testing room, one participant at a time. All three experiments were run by the same female experimenter.

General Hypotheses

Because this research examined a quite novel intersection between local metacognition, gender differences in spatial navigation, and stereotype activation it was necessarily partly exploratory in nature. By adapting the methodologies from the metacognition literature and stereotype threat/lift literature to our navigation task, we were better able to operationalize our hypotheses in a way that is highly structured and comparable to the existing literature in the various relevant fields.

The research was driven by a number of hypotheses derived from past literature examining gender differences in navigation. We predicted that men would outperform women on all navigation tasks, showing fewer errors, shorter maze completion times, and fewer failed mazes. In addition, we hypothesized that men would demonstrate greater overall confidence than would women, with larger mean EOL, JOL, and RCJ ratings.

In relation to metacognitive accuracy, we hypothesized that men would have more accurate relative and absolute metacognitive values compared to women. These hypotheses are derived from multiple sources, including a study that previously examined a form of local metacognition during a mental rotation task and found that men had more accurate confidence (Cooke-Simpson & Voyer, 2007). Although the only other relevant study which found gender to not predict performance and metacognitive outcomes on a map and verbal direction memory task (Schwartz, 2006). Regardless, women are expected to have greater difficulty in distinguishing between mazes that will be (or were) correctly or incorrectly performed because they have less experience with tasks that hone spatial cognition skills compared with men (e.g., Lawton & Morrin, 1999) as well as a poorer VSWM (for a review, see Voyer et al., 2017). This is corroborated by studies showing experts demonstrate relatively superior monitoring and control abilities compared with novices across a variety of tasks (e.g., Gobet, 1998; Hallam, 2001; Martini & Shore, 2008). Thus, we expected women's metacognitive confidence ratings would be less well correlated with their performance on a trial-by-trial basis (relative metacognitive accuracy). In addition, we expected any difference between observed and predicted performance probability means to be greater for women than for men (absolute metacognitive accuracy).

In addition to the above hypotheses, which apply to all three experiments, there were a number specific to Experiment 3, which involve stereotype facilitation effects. The hypotheses related to stereotype activation are based on a combination of the Nelson and Narens (1990) metamemory framework and the Schmader et al. (2008) model of stereotype activation effects. This latter model suggests that, during stereotype threat activation, increased performance monitoring and stress arousal, due to an increase in self-consciousness and anxiety, deplete working memory resources and thus limit performance. They consider stereotype lift to alleviate stress, which allows working memory resources to fully support performance. Related work has examined the role of other variables surrounding the concept of self-efficacy. For instance, a lack of self-confidence, self-efficacy and/or increased anxiety have been shown to influence global metacognition and metacognitive monitoring (e.g., Coutinho & Neuman, 2008; Kleitman & Stankov, 2007; McDonald-Miszczak, Gould, & Tychynski, 1999; Stankov; Tobias & Everson, 1997; Veenman, Kerseboom, & Imthorn, 2000; West, Boatwright, & Schleser, 1984). As the framework by Nelson and Narens (1990) implies improved performance with improved metacognitive monitoring and control, it is possible to combine this idea with Schmader et al.'s stereotype threat model and hypothesize that stereotype threat reduces performance by reducing metacognitive monitoring accuracy via changes in self-confidence and/or anxiety. That is, we posit that stereotype threat may increase anxiety and reduce self-confidence which influence metacognitive monitoring and control processes, which in turn deplete working memory resources, ultimately limiting performance. In short, metacognitive factors may act as an intermediary between stereotype activation and performance modulations. The present study sought to investigate this hypothesis by determining if confidence, metacognitive monitoring and performance are affected by stereotype activation. We also explored whether stereotype threat would influence spatial anxiety using the SAS and global metacognition using the SBSOD.

More specific hypotheses are described in each individual experiment. As stated above, the novel nature of this study required that some of the methodology be exploratory in nature. For instance, we examined the relationship between different types of metacognitive judgments (EOL, JOL, and RCJ) and performance on our navigation task in order to compare the results on each judgment to the metacognitive literature. This was done without a guiding hypothesis, with the goal of initiating research on the relationship between local metacognition and navigation.

CHAPTER 5: EXPERIMENT 1

GENDER DIFFERENCES IN METACOGNITIVE JUDGMENTS AND PERFORMANCE ON A FIRST-PERSON VIRTUAL MAZE LEARNING TASK

Chantal L. Lemieux, Charles A. Collin
University of Ottawa

Nicholas N. Watier
Brandon University

Author Note

Chantal L. Lemieux, School of Psychology, University of Ottawa; Charles A. Collin, School of Psychology, University of Ottawa; Nicholas Watier, Brandon University.

This work was supported by a grant from the *Canadian Natural Sciences and Engineering Research Council* to CC [Grant # 2015-05067]. We thank Joe MacInnis for creating the computer program used to present the virtual mazes.

Abstract

Many studies have shown a gender difference in spatial navigation ability including a related gender difference in global metacognitive self-assessment (e.g., Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006) and spatial anxiety (e.g., Lawton & Kallai, 2002; Nowak, Murali, & Driscoll, 2015). However, very few studies have examined whether there are gender differences in trial-by-trial self-assessment, or what we term *local* metacognition. We assessed trial-by-trial metacognitive performance in a sample of men and women engaging in a first-person virtual maze navigation task. Methods for assessing trial-by-trial metacognitive performance were adapted from Nelson and Narens' (1990) metamemory framework. Results showed that men were more accurate at assessing their trial-by-trial performance than women when the assessment was made after performance. This suggests that women are more likely to err in assessing their past navigational performance, and thus may be less likely to undertake corrective control actions in the future (Nelson & Narens, 1990).

Keywords: metacognition, gender differences, spatial navigation, virtual reality, confidence, metamemory, wayfinding

Gender Differences in Metacognitive Judgments and Performance on a First-Person Virtual Maze Learning Task

While the gender gap in many cognitive abilities has been closing over recent decades, there remains a strong and persistent difference favoring males in spatial cognition (for reviews, see Coluccia & Louse, 2004; Linn & Petersen, 1985; Martens & Antonenko, 2012; Voyer, Voyer, & Bryden, 1995). More specifically, males generally have faster maze running times as well as fewer errors than females. Men and women also rely on different environmental cues for navigation, and have correspondingly different navigational strategies (for a review, see Coluccia & Louse, 2004). For instance, males tend to employ a more orientation-based strategy using allocentric cardinal points when giving directions, whereas women tend to employ a more landmark-based strategy using landmarks at specific decision points to give directions (e.g., Saucier et al., 2002). Several studies have evidence to suggest this gender difference arises due to biological differences such as genetics and hormones (e.g., Collaer & Hines, 1995; Halari et al., 2005; Hines et al., 2003; Kimura, 1999; Silverman, Choi, & Peters, 2007), while others have revealed sociocognitive factors to have an influence (Casey, 1996; Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005).

Many studies have suggested this performance and strategy difference may be partially due to confidence differences favoring men (e.g., Lundeberg, Fox, & Puncochar, 1994) or differences in spatial anxiety disfavoring women (e.g., Lawton, 1994; Lawton & Kallai, 2002; Nowak et al., 2015). These studies have mainly used self-reported scales designed to determine overall confidence during specific navigational contexts, like the Santa Barbara Sense of Direction Scale (Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002), the Spatial Anxiety Scale (Lawton, 1994), and the International Wayfinding Strategy Scale (Lawton &

Kallai, 2002). In other words, they assess a form of *global metacognition* (Kelemen, Frost, & Weaver, 2000) by asking participants to rate their abilities, confidence, or comfort during specific types of spatial tasks. Several studies using global measures have shown gender differences in metacognition demonstrating that women generally underestimated their performance more than men do (e.g., Hertzog, Dixon, & Hultsch, 1990). In the present study, we are interested in exploring a more fine-grained form of metacognition that has to do with how well men and women assess and predict their performance on individual trials of an ongoing spatial navigation task. This is referred to as *local metacognition* (Lieberman, 2004) to distinguish it from more global measures of overall self-assessment.

Nelson and Narens (1990) developed a framework that accounts for the acquisition of local metacognitive knowledge. According to their framework, cognitive processes are divided into two hierarchical frames of reference: the object-level, which is the cognitive process itself, and the meta-level, which contains a dynamic representation of the object-level. Communication between these two frames of reference is accomplished through a cycle of monitoring and control processes. In monitoring processes, information from the object-level (e.g., stored knowledge of study material) is sent to the meta-level, so that the state of the object-level can be evaluated according to some goal (e.g., to have enough knowledge stored to pass an exam). In control processes, information from the meta-level is sent back to the object-level, so as to adjust the state of the object-level through behavior (e.g., spending more time studying). Several studies have supported this model and found an association between metacognitive accuracy and performance in the context of classroom-based tasks (e.g., Dunlosky & Ariel, 2011; Dunlosky & Rawson, 2012; Hacker, Bol, Horgan, & Rakow, 2000; Nietfeld, Cao, & Osborne, 2005; Thiede, Anderson, & Therriault, 2003). As monitoring accuracy increases, students are more effective at regulating their study time, which is related to increased performance (e.g., Hacker et al., 2000). Many studies have shown that participants spend more time studying items they previously judged as difficult compared to items they had previously judged as easy (for a review, see Son and Metcalfe (2000). However, recent research has shown that study time allocation (i.e., control) is more strategic depending on task demands (Son & Metcalfe, 2000). For instance, Son and Metcalfe (2000) found that participants allocated more study time to items they had previously judged as easy when study time was restricted and insufficient. More recently, researchers have argued for a bidirectional link between monitoring and control during learning, whereby a control process (spending extra time on an item) can feed into monitoring ("that item was hard if it took so long, therefore it probably won't be remembered") (Koriat, 2006; Koriat, Ackerman, Adiv, Lockl, & Schneider, 2014).

If men have greater confidence and women have greater anxiety related to spatial tasks, they likely perceive their own spatial competency differently, which could influence the way they approach these tasks (i.e., monitor and control their memory). In other words, differences in global metacognition (i.e., self-efficacy, Bandura, 1997) could influence, among other things, local metacognition for an ongoing task. In a recent review by Dunlosky, Mueller, and Thiede (2016), the authors argue that much of the metacognition research to date suggests metacognitive judgments (measures of local metacognition) are influenced by both epistemological beliefs, such as task heuristics or subjective feelings, and perceptions of how easy an item was processed (e.g., Frank & Kuhlmann, 2016; Mueller & Dunlosky, 2017; Mueller, Dunlosky, Tauber, & Rhodes, 2014).

In the context of spatial tasks, only a few studies have measured local metacognition, and these used a methodology outside of an existing metacognitive framework (e.g., Pfuhl, Barrera, Living, & Biegler, 2013; Stevens & Carlson, 2016). For instance, Stevens and Carlson (2016) employed a directional pointing task where participants gave pointing confidence intervals as a measure of confidence (larger intervals representing lower confidence). They found that participants used multiple sources of information to evaluate confidence, and were sensitive to, yet still underestimated, processing fluency difficulties related to egocentric and allocentric reference framing when monitoring performance (Stevens & Carlson, 2016). While this study took into consideration the effects of gender by considering it a covariate in the analyses, the authors did not analyze or interpret any data related to gender. However, in one of their linear mixed effects regression models predicting pointing confidence interval, the fixed factor of gender was significant ($p<.001$) showing women to have larger pointing confidence intervals (Stevens & Carlson, 2016). Relatedly, after collecting confidence scores after each trial of a well-known mental rotation task, Cooke-Simpson and Voyer (2007) determined that men had more accurate confidence ratings than women.

The Nelson and Narens framework proposes the use of several self-reported monitoring judgments about past and future trial-by-trial performance to assess local metacognitive knowledge. These include Ease of Learning Judgments (EOLs), Judgments of Learning (JOLs), and Retrospective Confidence Judgments (RCJs), among others. EOLs are collected after showing the participant the items to learn, but before they have had a chance to learn them. They reflect the degree of subjective ease a participant believes they would have in acquiring an item to an ideal level of mastery. They are typically collected by having participants rate the perceived difficulty of each item separately on a numerical scale (e.g., de Carvalho & Yuzawa, 2001). JOLs are similar to EOLs, but are collected after a participant has had a chance to learn the items. They reflect a participant's subjective sense of how well they have learned the items and how well they will perform when recalling the items at some future time. RCJs are collected after participants perform a trial of a recall task. They are defined as an assessment of the probability that an item retrieved from memory at test was the correct target (e.g., Koriat, Pearlman-Avnion, & Ben-Zur, 1998). Thus, RCJs are retrospective estimates of performance whereas EOLs and JOLs are prospective. All of these judgments are collected on a trial-by-trial basis and represent the participant's metacognitive confidence for that particular trial. Metacognitive accuracy can then be measured by comparing each trial's actual performance with the metacognitive judgment.

While much of the research on spatial navigation has focused on global self-assessed metacognition (global metacognition), only one study has examined metacognition in spatial navigation on a trial-by-trial basis using a metacognition framework. Schwartz (2006) assessed participants' prospective estimates of memory performance for maps and verbal directions. On each trial, participants were asked to study a map or a verbal direction for a fixed time interval. Afterwards, a JOL was solicited, whereby participants were asked to rate their confidence in their ability to recall the studied item. Immediately after completing the JOL, participants performed a recall task by either writing or drawing the directions from memory. The results indicated that participants were twice as accurate in their ability to self-assess future navigation performance in comparison to typical JOL paradigms that rely on word-pair associates as the to-be-remembered items. However, there was no explicit comparison of male and female participants as preliminary analyses determined that age and gender did not predict any of the outcome variables. Thus, while Schwartz (2006) revealed that memory monitoring for navigational information is well above chance levels, it remains to be confirmed whether this pattern differs between males and females.

In the present study, our goal was to better understand how gender differences in local metacognition might be related to gender differences in spatial navigation performance. To do so we applied Nelson and Narens' (1990) metacognitive framework to a navigation task. We collected data on trial-by-trial metacognitive performance by having participants make prospective Ease of Learning (EOL) judgments as well as Retrospective Confidence Judgment (RCJ) during a first-person virtual maze-learning task. As with previous paradigms examining metacognition (e.g., in word-pair association tasks), task performance was assessed in a dichotomous fashion. Participants were judged to "pass" a maze upon successfully completing it without making any wrong turns. Otherwise, the trial was scored as a failure. This was done for a series of thirty first-person virtual mazes.

In keeping with previous studies in metacognition, a participant's relative metacognitive accuracy (i.e., how well a participant is able to monitor items relative to each other) was assessed by correlating their EOL or RCJ ratings to dichotomous navigation performance using the Goodman-Kruskal gamma correlation (Nelson & Narens, 1990). Furthermore, absolute metacognitive accuracy was assessed by examining calibration curves, which plot mean predicted performance as a function of mean observed performance, in addition to a comparison of means between observed and predicted performance. Absolute accuracy is often used to determine if a participant is generally under or over confident (Nelson & Narens, 1990). For a recent review of metacognitive methodology, see Dunlosky et al. (2016).

Questions & Hypotheses

The goal of this study is to determine whether the genders differ in their navigational and metacognitive performance during a virtual spatial navigation task. If we do find a gender difference in metacognitive accuracy, this may explain, at least in part, gender differences in navigation performance warranting further research (Hacker et al., 2000; Nietfeld et al., 2005; Thiede et al., 2003). We hypothesized that men would have greater navigation performance than women. That is, we predicted that they would exhibit faster maze completion times with fewer errors. This prediction is based on a large number of studies that have shown a gender difference in navigation performance favoring men (for a review, see Coluccia & Louse, 2004). Similarly, we hypothesized that men would have higher metacognitive confidence ratings than women. This again is based on a large number of previous studies showing men to have more confidence (e.g., Lawton, 1994; Lawton & Kallai, 2002; Lundeberg et al., 1994). Finally, and most critically, we predicted that women's trial-by-trial self-assessments would be less accurate than men's would. Considering that women have a poorer visuo-spatial working memory span and less experience with tasks that hone spatial cognition skills compared with men (e.g., Lawton & Morrin, 1999), women should have greater difficulty in distinguishing between mazes that will be (or were) correctly or incorrectly navigated. Consequently, we expected women's metacognitive confidence ratings would be less well correlated with their performance on a trial-by-trial basis.

Methods

Participants

Seventy participants were recruited to participate via either the departmental subject pool or recruitment posters. Ten participants could not be included in the analyses: Seven women and one man had to terminate their participation due to simulator sickness, and technical difficulties prevented one man and one woman from completing the experiment. Participants were to be excluded if they did not identify on the traditional gender binary (i.e., either male or female) and instead indicated “other” when filling out demographic information; however there were no cases of this. The final sample was 30 men and 30 women, for a total N of 60. Average age was 21.1±4.12 years. Sample size was initially determined by conducting an a-priori power analysis using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) for a one-tailed independent-measures t-test with $\alpha = .05$, and a medium-to-large effect size ($d = .65$; Cohen, 1988). Effect size was chosen as a compromise between the large effect size (e.g., $d \geq .80$) seen in previous studies that used virtual Hebb-Williams mazes to examine gender differences in spatial navigation (e.g., Shore, Stanford, MacInnes, Klein, & Brown, 2001)[4] and the moderate effect size ($d = .58$) seen in a study comparing confidence accuracy between genders after a mental rotation task (Cooke-Simpson & Voyer, 2007). The outcome of the analysis revealed that a minimum sample size of 60 participants would be required to obtain a power of .80.

Materials

Virtual Navigation Task

Thirty first-person perspective virtual mazes were created from 10 Hebb-Williams mazes (Hebb & Williams, 1946). These 10 were shown to be the most difficult in past research (Shore et al., 2001; Therrien & Collin, 2010). In addition to the 10 original maze layouts, 20 more were created by either transposing the maze left-to-right relative to the axis linking the entrance to the exit, or reversing the maze by making the entrance the exit and vice versa. See Figure 1 for a top town view of the mazes. The virtual maze program recorded an error whenever the participant crossed invisible “trip wires” that were strategically placed in dead ends (Shore et al., 2001).

---

[4] While a measure of effect size was not directly reported in this study, it could nonetheless be calculated from the reported main effect of gender collapsing across species by exploiting the fact that $F = t^2$, and $d_s = t\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ (Lakens, 2013).

Video tours of the mazes were also generated. The first person perspective view in each video tour moves through the entire maze environment, following either a right-hand or left-hand wall-follower algorithm. Thus, there were two videos per maze, one left-handed and one right-handed. Participants were shown only one randomly selected video per maze. Each video tour was approximately 45 seconds long, and was shown with a screen resolution of 800 x 600 pixels. Videos were played using Windows Media Player. A 45-degree field-of-view was used in all video tours and maze performances, as a low field-of-view has been shown to be ideal for avoiding simulator sickness (e.g., Allen, 2000; DiZio & Lackner, 1998; Lin, Duh, Parker, Abi-Rached, & Furness, 2002). The experiment was run on a standard desktop PC computer running Windows XP with a 19-inch LCD screen. The maze task was programmed in C++ using OpenGL. Every maze was designed with the same texture on all wall, floor and ceiling surfaces, which was a fractal noise pattern (see Figure 2). This pattern mimics the spatial frequency content of an average natural scene, but with all structure randomized. The maze program automatically collected the time taken to complete the maze, error counts per maze, and raw navigation data (x/y position and orientation) every 100 ms. Participants navigated using the arrow keys.

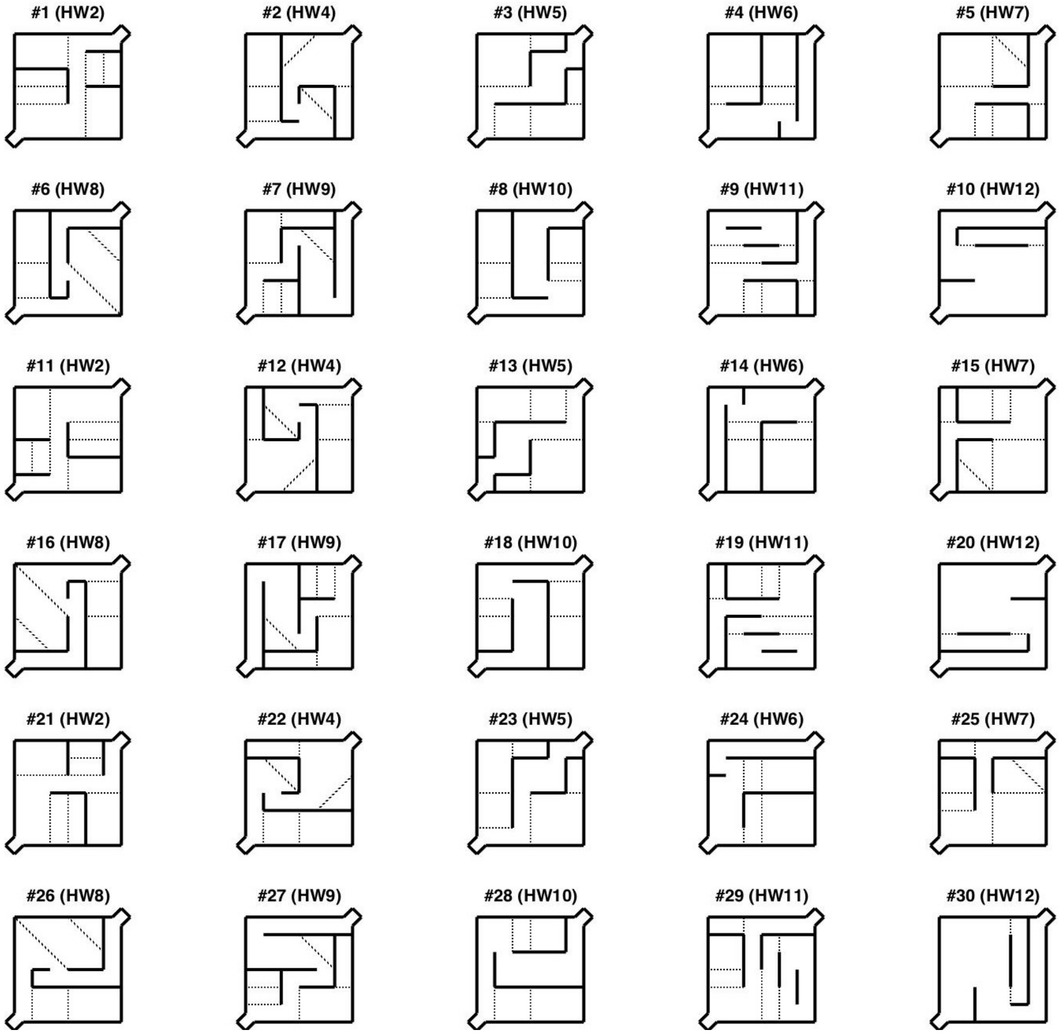


Figure 1. Layouts of the 30 mazes used in this experiment. Mazes 1 to 10 are reproductions of the original 12 Hebb-Williams mazes (Hebb & Williams, 1946), excluding #1 and #3, which were deemed too easy to perform. Mazes 11 to 20 are reversed versions of mazes 1 to 10, where the entrance has been made into the exit, and vice versa. Mazes 21 to 30 are transposed versions of mazes 1-10, where the x and y coordinates of the walls have been exchanged. Dotted lines show virtual “trip wires” replicated from Shore et al. (2001).

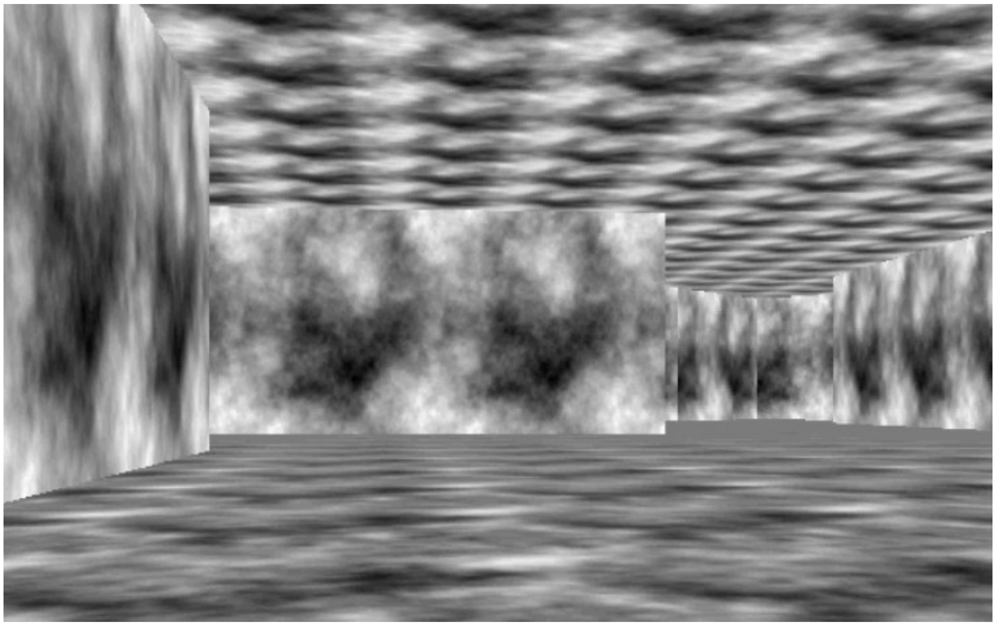


Figure 2. Screen shot demonstrating what participants saw for both the video tours and maze performances.

Metacognitive Materials

For the present study, we modified Nelson and Narens' (1990) prospective Ease of Learning (EOL) judgment instructions as follows:

> *"Clearly indicate on a scale of 0-100, going from "difficult" to "easy" with five options "0-25-50-75-100", how easy you believe it would be to learn the maze's layout such that you could navigate from the entrance to the exit without making any errors."*

Similarly, we retained the probabilistic scale of the Retrospective Confidence Judgment (RCJ) but modified the instructions as follows:

> *"Clearly indicate on a scale of 0-100, going from "not confident" to "confident" with five options "0-25-50-75-100", how confident you are that you successfully navigated the maze without making any errors."*

Additional relevant instructions for both were:

> *"Note that an error is defined as not taking the shortest route between start and finish (e.g., going down a dead end, or taking a wrong turn). Please answer as honestly as possible and use the entire scale."*

Procedure

There were two phases to this experiment. The first was the video phase, where participants were shown video tours of the virtual mazes and gave EOLs following each tour. This was followed by the performance phase, where participants attempted to navigate the mazes and gave RCJs following each navigation attempt. The video phase (30 maze videos) was completed in its entirety before the performance phase (30 maze performances). Instructions and practice trials were given before each phase. The metacognitive materials were also presented to them and explained. Participants were informed that each maze had a square layout and that the exit was always diagonally opposite to the entrance.

During the video phase, the 30 tours were presented in random order and each was immediately followed by the collection of an EOL from the participant. The experimenter verbally repeated the EOL instruction and recorded the participant's answer for each video tour. A 5-minute break was given halfway through (approximately 20 minutes into testing) and additional short breaks when needed were encouraged. Once the video phase was over, a 5-minute break was given between phases where the participant could sit with their eyes closed or walk in the hallway to get water or go to the washroom.

During the performance phase, participants navigated through each maze using the arrow keys on the keyboard. When the participant reached the exit of each maze, they entered an empty virtual room. While there, the experimenter immediately gave the RCJ instructions and recorded the participant's answer. Time to complete a maze was limited to 5 minutes. If a participant exceeded this time, the experimenter gave them the directions to the exit and moved on. This occurred rarely (in <.005% of trails) and the data for these trials were included for the outlier analysis. The mazes were navigated in random order (i.e., a different random order than the one in which the video tours were shown). Another 5-minute break was given halfway through the performance phase and short breaks were encouraged whenever needed. Notably, the RCJ was always immediately collected after the last performance before any break. Once RCJs for all 30 maze performances were collected, the experiment was complete. In total, testing took approximately 80 minutes. The experimenter was in the room for both phases in order to give instructions throughout, assign consistent breaks, play the video tours, start the maze program, and record the metacognitive judgments. Metacognitive judgments were recorded on a paper form by the experimenter as the participant provided their EOL and RCJ scores aloud.

To reiterate, the video phase (30 maze videos) was entirely done before the performance phase (30 maze performances). This arrangement was used to minimize learning transfer. That is, it was assumed that most participants would not be able to recall the layout of 30 sequentially presented mazes, especially given that the order of initial presentation was randomized relative to the order in which they navigated the mazes. Because EOLs are a measure of the perceived difficulty of learning the maze in the future, we wanted to ensure that actual learning was minimized during these ratings (otherwise the rating becomes a JOL instead). Our task was therefore organized such that it did not elicit a conscious effort to learn the maze layout while viewing the video tours. Performance, therefore, was done on a largely novel environment with the knowledge that the layout of each maze was square and the exit was at the opposite diagonal corner to the entrance.

Results

Univariate outlier analysis was conducted on all dependent variables. An outlier was defined as having a z score above 2.43 in accordance with our sample size of 30 individuals per condition (Van Selst & Jolicoeur, 1994). In order to retain an adequate sample size, outliers were winsorized (Hogg, 1979; Huber, 2011). In other words, their scores were changed ("tucked") to the closest non-outlying score. Overall, 1.78% of scores were considered univariate outliers and were winsorized. All DVs were considered normally distributed as skewness and kurtosis scores were all near 0 within a ±2 criterion range (Field, 2009; Gravetter & Wallnau, 2016; Trochim & Donnelly, 2006).

In addition to the performance variables recorded by the task program, another dependent variable for performance named *Mazes Passed* was created. This was done by recoding error count to either a pass or fail (1 or more errors) for each of the 30 mazes. Local metacognition is typically evaluated using relative and absolute accuracy between confidence scores and performance. Relative accuracy for EOLs and RCJs is typically assessed by the Goodman-Kruskal gamma correlation (Nelson, 1984) which provides a statistical index of concordance between two ordinal variables. Performance is traditionally a dichotomous variable (Mazes Passed: pass or fail) and an ordinal variable (confidence score 0-25-50-75-100) for each participant. One limitation of using gamma as a measure of metacognitive accuracy is that it is bounded between -1 to +1, as with most coefficients of association. This means that it inherently violates the assumptions required by parametric hypothesis tests. To address this issue, gamma can be converted into a standardized measure called G* (Benjamin & Diaz, 2008)[5]. G* is normally distributed and unbounded (Benjamin & Diaz, 2008). For G* as with gamma, the higher the score the more accurate the participant's metacognitive accuracy. Mean EOL and RCJ ratings were also computed as general indicators of metacognitive confidence.

In accordance with the structure of our hypotheses, the analyses consisted of a series of one-tailed independent sample t-tests between the genders on all performance DVs (Maze

---

[5] The equation used to transform gamma correlations into G* was : $G^* = \log\left(\frac{\gamma+1}{1-\gamma}\right)$ (Benjamin & Diaz, 2008)

Completion Time, Error Count, Mazes Passed), metacognitive confidence DVs (EOL and RCJ means), and average relative metacognitive accuracy (EOL and RCJ G*). Cohen's d was calculated as a measure of effect size. Variables demonstrating heterogeneity of variances, via Levene's Test of Equality of Variances, had their df adjusted as equal variances were not assumed. Please refer to Table 1 for descriptive statistics, t-test results, and effect sizes.

In addition to analyzing relative accuracy, we assessed absolute accuracy for RCJs by examining the slope of the calibration curve that is formed by plotting the observed proportion correct for each estimated proportion correct (Keren, 1991). See Figure 3. As EOL judgments are not on a probabilistic scale, absolute accuracy was not calculated (Keren, 1991). As a more objective analysis of absolute accuracy, a 2x2 mixed factorial ANOVA was also conducted comparing observed performance (Percent Mazes Passed) with predicted performance (RCJ mean) across both genders.

Gender Differences in Performance

A series of one-tailed independent sample t-tests were run in order to test our hypotheses regarding our three performance variables: Maze Completion Time, Error Count, and Mazes Passed (see Table 1). Results show women to have significantly longer maze completion times than men ($t = 5.003$, $df = 58$, $p < .001$) with a large effect size of 1.292. Women also exhibited a significantly greater number of errors ($t = 1.807$, $df = 58$, $p = .038$) compared to men, with a moderate effect size of .467. They also averaged fewer successfully completed mazes ($t = -1.677$, $df = 58$, $p < .05$) compared to men, with a moderate effect size of .433. These data confirmed our hypothesis that women would have poorer maze navigation performance compared to men.

Table 1.

One-tailed independent t-test results and descriptive statistics comparing gender

				Women		Men	
DV	t(df)	Sig.	d	$\bar{x}$	s	$\bar{x}$	s
Maze Completion Time (s)	5.003(58)	<.001	1.292	32.10	7.13	24.42	4.47
Error Count	1.807(58)	.038	.467	39.10	10.95	33.47	13.10
Mazes Passed	-1.677(58)	<.050	.433	14.80	3.69	16.33	3.39
EOL mean	-4.098(44.95)^{t}	<.001	1.051	52.25	9.36	66.83	17.10
RCJ mean	-1.851(58)	.035	.478	68.44	10.16	73.14	9.47
EOL gamma				0.04	0.33	0.16	0.30
EOL G*	1.301(58)	.100	.344	0.04	0.31	0.14	0.27
RCJ gamma				0.65	0.38	0.83	0.19
RCJ G*	2.299(58)	.013	.594	0.94	0.70	1.32	0.61
RCJ and Error Count^{p}	2.786(47.76)^{t}	.004	.746	-0.53	0.29	-0.71	0.18
RCJ and Maze Completion Time^{p}	2.079(58)	.021	.508	-0.54	0.26	-0.66	0.21

Note. ^{t} = equal variances not assumed. ^{p} = average Spearman's Roh correlations between RCJ and performance variables (either Error Count of Maze Complation Time) per participant across all 30 mazes. N=60 (30 females and 30 males)

Gender Differences in Metacognitive Confidence and Relative Metacognitive Accuracy

A series of one-tailed independent sample t-tests were run in order to test our hypotheses concerning metacognifive confidence variables (EOL mean and RCJ mean) as well as relative metacognitive accuracy for both EOL and RCJ judgments (G*; see Table 1).

Women showed significantly lower confidence scores for their EOLs (t = -4.098, df = 44.95, p < .001) as well as their RCJs (t = -1.851, df = 58, p = .035) in comparison to men with strong to medium effect sizes (d = 1.051 and d = .478, respectively). These findings confirm our hypothesis that women would have significantly lower metacognitive confidence compared to men.

Men showed greater retrospective metacognitive accuracy compared to women, with greater RCJ G* scores (t = 2.299, df =58, p = .013). However, there was no significant difference between men and women on prospective accuracy using EOL G* scores (t = 1.301, df = 58, p = .100). As EOL G* scores were quite low for both women ($\bar{x}$ = -.04, s = .31) and men ($\bar{x}$ = -.14, s = .27), we decided to compare their scores to chance (a G* score of 0; meaning no relationship between EOL and performance) using a one sample one-tailed t-test for each gender. Interestingly, when comparing EOL G* to chance, we found that women's average EOL G* was not significantly different from chance (t = -.727, df = 29, p = .237) but men's average EOL G* was (t = -2.800, df = 29, p = .005) with a moderate effect size of .519. That is, women's prospective metacognitive accuracy was no different than random chance whereas men's was.

In order to better understand the gender difference in retrospective metacognitive accuracy, non-parametric Spearman's Rho correlations were computed between RCJ and the two other performance variables (Error Count and Maze Completion Time) for each participant across all 30 mazes. Similar one-tailed independent sample t-tests were run comparing the average correlation between each pair of variables (RCJ and Error Count, RCJ and Maze Completion Time) for each gender. The results show (see Table 1) that men indeed have higher correlations than women between RCJ and Error Count ($t = 2.786$, $df = 47.76$, $p = .004$) with a large effect size of .746, as well as RCJ and Maze Completion Time ($t = 2.079$, $df = 58$, $p = .021$) with a moderate effect size of .508. In other words, men demonstrate better retrospective metacognitive accuracy across all three performance variables (Mazes Passed via G*, Error Count, and Maze Completion Time) with the greatest effect size seen in the correlation means between RCJ and Maze Completion Time ($d = .746$).

In summary, women were less accurate in their metacognitive judgments when compared to men if the judgment was made after performance (retrospective metacognitive accuracy), but no gender differences were found if the metacognitive judgment was made before performance (prospective metacognitive accuracy). However, women's prospective metacognitive confidence was not related to performance at all while men's prospective metacognitive confidence was related to performance, albeit weakly.

Gender Differences in Absolute Metacognitive Accuracy

Absolute accuracy is typically analyzed by examining the calibration curve depicting the mean observed performance on a memory test as a function of the mean estimated performance on a memory test, collapsing across all participants in a given condition (e.g., Keren, 1991; Watier & Collin, 2011). This was created in the present study by plotting the observed proportion of correctly completed mazes for each estimated proportion of correctly completed mazes (see Figure 3).

Perfect absolute accuracy occurs when the calibration curve shows a perfectly linear relationship between predicted and observed performance. Absolute accuracy allows one to assess whether participants are generally under confident or overconfident by examining whether their slope is above or below the perfect absolute accuracy line. A visual assessment of the calibration curves for RCJ judgments shows both genders to generally be overconfident when assessing their past performance. That is, both their calibration curves are below and to the right of the perfect absolute accuracy line. However, compared to men, women tend to be more accurate, and even slightly under-confident, when performance is low. Conversely, men show a stronger sense of overconfidence when performance is low (below 25%). However, the genders appear to be equal in their accuracy and confidence when predicted performance is above 50%.

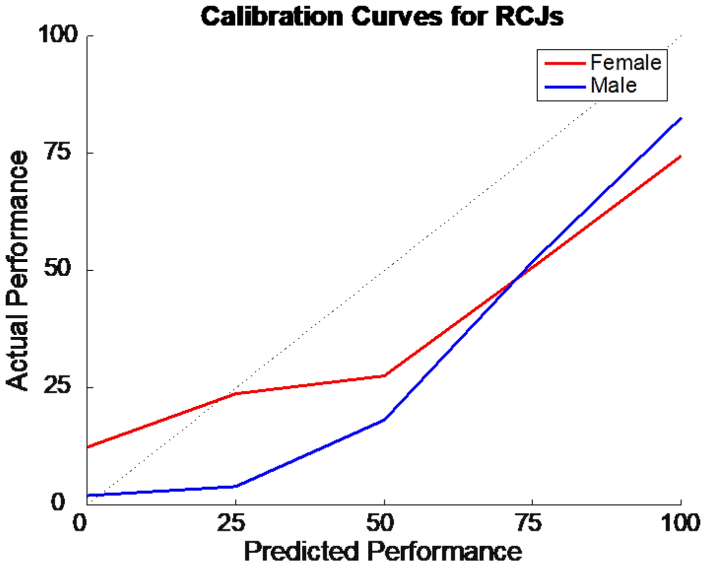


Figure 3. Calibration curves for male and female participants, showing absolute accuracy. N=60 (30 females and 30 males)

As a more objective analysis of absolute accuracy, a 2x2 mixed factorial ANOVA was conducted comparing observed performance (Percent Mazes Passed) with predicted performance

(RCJ mean) across both genders. The main effect for observed vs predicted performance shows both genders to be overconfident, as their predicted performance (RCJ mean) was statistically greater than their actual performance (Percent Mazes Passed; $F = 88.716$, $df = 1$, $p < .001$). The main effect for gender was also significant, with men showing greater observed and predicted performance ($F = 6.271$, $df = 1$, $p = .015$). However, the interaction between observed vs predicted performance and gender was not significant ($F = .011$, $df = 1$, $p = .918$). In other words, it can be concluded that the genders did not differ in their absolute accuracy because the relationship between observed and predicted performance was not statistically different across genders.

Discussion

In accordance with our hypotheses, the results show that women demonstrated poorer virtual navigation performance, lower confidence, as well as poorer local retrospective metacognitive accuracy, as compared to men. This latter finding suggests that women are more likely to err in assessing their past navigational performance, and thus may be less likely to undertake corrective control actions in the future (Nelson & Narens, 1990). While a causal connection between metacognitive monitoring and control in spatial navigation tasks still needs to be determined, our results point to the possibility that the disadvantage women exhibit in navigational tasks may at least in part be explained by metacognitive factors.

Our navigation performance data show, in accordance with previous literature (for a review, see Coluccia & Louse, 2004), that women took longer to complete the mazes, failed a greater number of mazes, and overall made a greater number of navigational errors in comparison to the men in our sample. The greatest effect size ($d = 1.292$) was seen with maze completion times, with women taking nearly 8 seconds more on average to complete a maze, against an average completion time of about 30 seconds. Several studies have shown that if participants are allowed to take as much time as necessary on spatial tasks, then the performance difference between genders disapears (e.g., Goldstein, Haldane, & Mitchell, 1990; Voyer & Sullivan, 2003) or is at least reduced (Voyer, 2011). Some researchers believe it may take longer for women to change their strategy from route-based to orientation-based, and that this explains their longer latencies compared with men (e.g., Barkley & Gabriel, 2007; Lawton, 1994; Sandstrom, Kaufman, & Huettel, 1998). Our first-person virtual navigation task was designed to be devoid of overt landmarks, thus forcing participants to adopt an orientation-based strategy. Other studies have shown that during open exploration tasks, women generally travel slower and exhibit more hesitant behaviors (e.g., Lavenex & Lavenex, 2010; Lawton, Charleston, & Zieles, 1996). Both of these are possible explanations for why the women in the present study showed longer maze completion times. That is, perhaps the women needed time to change their strategy from route to orientation based and/or they may have generally travelled slower with more hesitance. It is important to note that movement speed through the maze was fixed in the current study, so any difference in travel time across a given distance would necessarily be based on behavioral differences such as hesitance. Such hesistance may be related to metacognitive confidence differences between the genders, as it implies a lack of confidence in a navigational decision.

Our mean metacognitive confidence ratings data showed that women had lower trial-by-trial confidence both prospectively (i.e., when making EOL judgments) and retrospectively (i.e., when making RCJs). There is much research showing confidence differences between the genders in a variety of fields (e.g., Lawton et al., 1996; O'Laughlin & Brubaker, 1998; Picucci, Caffo, & Bosco, 2011; Stevens & Carlson, 2016) and there was no reason to believe metacognitive confidence related to spatial navigation would be any different. It is for this reason we hypothesized men would have greater confidence than women. Our results support this hypothesis with both EOL and RCJ judgments. The confidence difference between men and women was greater for EOLs than for RCJs, which is reflected in their corresponding effect sizes. In other words, the gender difference was greater for judgments made prior to performance compared to judgments made after performance. As there is a relatively robust stereotype that women have poor navigation skills, perhaps this greater gender difference in confidence before performance can be explained by the effects of stereotype threat (Steele & Aronson, 1995). Indeed, stereotype threat has been shown to reduce performance in women in a variety of tasks related to spatial cognition (e.g., Massa, Mayer, & Bohon, 2005; Moe & Pazzaglia, 2006; Wraga, Duncan, Jacobs, Helt, & Church, 2006). In addition, confidence has been shown to mediate the gender difference in mental rotation (Estes & Felker, 2012). Future research is needed to further investigate whether stereotype threat is at play in women performing navigation tasks and whether this is related to metacognition via a reduction in confidence.

Our metacognitive accuracy data—that is, the correlation between metacognitive judgment and performance on a trial—showed gender differences as well. Specifically, retrospective metacognitive relative accuracy was significantly different between the genders, with women showing poorer metacognitive accuracy compared to men. In other words, females were poorer at distinguishing between which mazes were and were not accurately completed compared with males. In addition, women's retrospective metacognitive judgments were also significantly less well correlated with the number of errors they made as well as the time it took to complete each maze. A possible explanation for this gender difference is that men may have more experience with spatial tasks and/or virtual environments. This additional experience may have allowed them to better discriminate between low and high performances by making better use of the number of errors and the time it took to complete the maze as cues (i.e., better cue-utlization, Koriat, 1997).

In contrast to the RCJ accuracy data, prospective (EOL) metacognitive relative accuracy was not significantly different between the genders, though there was still a moderate effect size (*d* = .344). However, men's average EOL G* values were approximately three times those of women, though both were quite low nearing chance. A further investigation found that women's prospective metacognitive judgments were not different from chance. In other words, their prospective judgments were random and not correlated with performance at all. Conversely, with men, prospective judgments showed a relationship with performance that was significantly above chance. Nonetheless, men still showed poor prospective accuracy. This finding is consistent with other studies that have examined performance using word-pair associates (e.g., Leonesio & Nelson, 1990) as well as face-name associates (e.g., Watier & Collin, 2011) which have shown EOLs to have the poorest correlation with performance compared to all other metacognitive judgments. The Nelson and Narens' metacognitive monitoring and control framework predicts higher accuracy when the information used to make the judgment is more similar to the information needed at time of test. In our task, EOLs were collected after video tours of the same environment later used during performance; moreover, both learning and testing phases used the same first-person perspective parameters. Thus, the information presented during both phases was quite similar. Despite this, EOL accuracy was quite poor. A possible explanation for this finding is that the video tours were viewed passively and involved travelling through the entire maze, including incorrect passages (dead-ends). This could have made it more difficult to accurately assess the difficulty of the maze layout as active and passive spatial learning have been shown to affect performance (Carassa, Geminiani, Morganti, & Varotto, 2002).

While there were gender differences on overall confidence levels for *both* prospective and retrospective judgments, a strong gender difference was only found for retrospective metacognitive accuracy and not prospective. This pattern of findings points to the need to assess not only global metacognitive factors, such as overall confidence in one's navigation abilities, but also local metacognitive performance, as the two are not necessarily directly related.

Compared to word-pair association tasks used in most local metacognitive studies, the relative accuracy gamma correlations for RCJs were quite high in the present study[6]. This was similar to results found by Schwartz (2006), who argued that navigation tasks might have greater ecological validity compared to word-pair association tasks and therefore might yield more developed and more accurate metacognitive judgments.

Our results regarding absolute accuracy (see Figure 3) indicate a general sense of overconfidence in both genders when making retrospective judgments. This general overconfidence is entitled the "overconfidence effect" and has been found before in a variety of tasks (e.g., Lichtenstein, Fischhoff, & Phillips, 1982; Moore & Healy, 2008; Svenson, 1981).

One potential limitation to this study was a possible sampling bias. While simulator sickness attrition is an issue in and of itself, it is important to note that there was a greater number of women who experienced simulator sickness than men (7 versus 1). This may create a sampling bias, in that a potentially larger proportion of women who can tolerate virtual 3D virtual environments were represented in our sample compared to the actual female population. A gender difference in simulator sickness also brings up another important limitation: if men

---

[6] Although G* was used in the analyses, RCJ gamma correlation means were .65 for women and .83 for men. See Table 1 for gamma correlation descriptive statistics by gender.

exhibit less simulator sickness perhaps they are generally better performers in virtual environments. Interestingly, though gender differences in navigation have been found among real-world navigation studies (e.g., Galea & Kimura, 1993; Halpern, 2013; Rossano & Moak, 1998), they are not as strong and robust as seen in virtual environments (for a review, see Coluccia & Louse, 2004). Both these limitations should be taken into consideration when interpreting the results for any study using virtual reality comparing gender differences.

In summary, women showed poorer navigational performance, lower overall confidence, and (critically) lower trial-by-trial metacognitive accuracy compared to men. The gender difference for confidence was larger when the judgment was made prior (EOLs) to the navigational task compared to after (RCJs). Judgments made prior to performance were not at all correlated with performance for women yet were modestly correlated with performance for men. Though both genders showed high retrospective metacognitive accuracy, there was still a gender difference where men were significantly more accurate in assessing their past performance. These results provide support for the hypothesis that metacognitive factors may contribute to gender differences in navigation ability. However, additional research is necessary in order to establish a causal link between gender differences in monitoring and gender differences in control when performing navigation tasks.

References

Allen, R. C. (2000). *The effect of restricted field of view on locomotion tasks, head movements, and motion sickness.* (Doctoral Dissertation), University of Central Florida, Retrospective Theses and Dissertations: 1758. Retrieved from http://stars.library.ucf.edu/rtd/1758

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W H Freeman/Times Books/Henry Holt & Co.

Barkley, C. L., & Gabriel, K. I. (2007). Sex differences in cue perception in a visual scene: Investigation of cue type. *Behavioral Neuroscience, 121*(2), 291-300. doi:10.1037/0735-7044.121.2.291 17469918

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 73-94). New York, NY: Psychology Press; US.

Carassa, A., Geminiani, G., Morganti, F., & Varotto, D. (2002). Active and passive spatial learning in a complex virtual environment: The effect of efficient exploration. *Cognitive Processing, 3*(4), 65-81.

Casey, M. B. (1996). Gender, sex, and cognition: Considering the interrelationship between biological and environmental factors. *Learning and Individual Differences, 8*(1), 39-53. doi:10.1016/S1041-6080%2896%2990006-0

Cohen, J. (1988). *Statistical analysis for the behavioral sciences* (2nd ed.). Hillsdale: NJ: Lawrance Erlbaum Associates.

Collaer, M. L., & Hines, M. (1995). Human behavioral sex differences: a role for gonadal hormones during early development? *Psychological Bulletin, 118*(1), 55-107.

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology, 24*(3), 329-340. doi:10.1016/j.jenvp.2004.08.006

Cooke-Simpson, A., & Voyer, D. (2007). Confidence and gender differences on the mental rotations test. *Learning and Individual Differences, 17*(2), 181-186. doi:10.1016/j.lindif.2007.03.009

de Carvalho, M. K., & Yuzawa, M. (2001). The effect of social influences and general metacognitive knowledge on metamemory judgments. *Contemporary Educational Psychology, 26*(4), 571-587. doi:10.1006/ceps.2000.1077

DiZio, P., & Lackner, J. (1998). *Alleviation of motion sickness and postural instability during and after virtual environment exposure*. Retrieved from Naval Air Warfare Center Training Systems Division: Orlando, FL

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 103-140). San Diego, CA: Elsevier Academic Press.

Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2016). Methodology for investigating human metamemory: Problems and pitfalls *The Oxford handbook of metamemory* (pp. 23-37). New York, NY: Oxford University Press.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271-280. doi:10.1016/j.learninstruc.2011.08.003

Estes, Z., & Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. *Arch Sex Behav, 41*(3), 550-570. doi:10.1007/s10508-011-9875-5 22130691

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods, 41*(4), 1149-1160.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, England: Sage Publications.

Frank, D. J., & Kuhlmann, B. G. (2016). More Than Just Beliefs: Experience and Beliefs Jointly Contribute to Volume Effects on Metacognitive Judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*(680-693). doi:10.1037/xlm0000332 27709983

Galea, L. A., & Kimura, D. (1993). Sex differences in route-learning. *Personality and Individual Differences, 14*(1), 53-65. doi:10.1016/0191-8869%2893%2990174-2

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition, 18*(5), 546-550. doi:10.3758/BF03198487 2233267

Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the behavioral sciences* (10th ed.). Belmont, California: Wadsworth Publishing.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*(1), 160-170. doi:10.1037/0022-0663.92.1.160

Halari, R., Hines, M., Kumari, V., Mehrotra, R., Wheeler, M., Ng, V., & Sharma, T. (2005). Sex Differences and Individual Differences in Cognitive Performance and Their Relationship to Endogenous Gonadal Hormones and Gonadotropins. *Behavioral Neuroscience, 119*(1), 104-117. doi:10.1037/0735-7044.119.1.104 15727517

Halpern, D. F. (2013). *Sex differences in cognitive abilities* (4th ed.). New York, NY: Psychology Press.

Hebb, D. O., & Williams, K. (1946). A method of rating animal intelligence. *Journal of General Psychology, 34*, 59-65. doi:10.1080/00221309.1946.10544520 21015350

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence, 34*(2), 151-176. doi:10.1016/j.intell.2005.09.005

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30*(5), 425-447.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Metamemory in adulthood: differentiating knowledge, belief, and behavior. In T. M. Hess (Ed.), *Aging and cognition: knowledge organization and utilization* (pp. 161-212). Oxford, England: Elsevier.

Hines, M., Fane, B. A., Pasterski, V. L., Mathews, G. A., Conway, G. S., & Brook, C. (2003). Spatial abilities following prenatal androgen abnormality: targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. *Psychoneuroendocrinology, 28*(8), 1010-1026.

Hogg, R. V. (1979). Statistical robustness: One view of its use in applications today. *The American Statistician, 33*(3), 108-115.

Huber, P. J. (2011). *Robust statistics*. Hoboken, NJ: John Wiley & Sons.

Kelemen, W. L., Frost, P. J., & Weaver, C. A., III. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition, 28*(1), 92-107. doi:10.3758/BF03211579 10714142

Keren, G. (1991). Calibration and Probability Judgments - Conceptual and Methodological Issues. *Acta Psychologica, 77*(3), 217-273. doi:10.1016/0001-6918(91)90036-Y

Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: The MIT Press.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology-General, 126*(4), 349-370. doi: 10.1037//0096-3445.126.4.349

Koriat, A. (2006). Are we frightened because we run away? Some evidence from metacognitive feelings *Memory and emotion: Interdisciplinary perspectives* (pp. 83-103). Malden: Blackwell Publishing.

Koriat, A., Ackerman, R., Adiv, S., Lockl, K., & Schneider, W. (2014). The effects of goal-driven and data-driven regulation on metacognitive monitoring during learning: A developmental perspective. *Journal of Experimental Psychology: General, 143*(1), 386-403. doi:10.1037/a0031768 23421442

Koriat, A., Pearlman-Avnion, S., & Ben-Zur, H. (1998). The subjective organization of input and output events in memory. *Psychological Research-Psychologische Forschung, 61*(4), 295-307. doi:10.1007/s004260050034

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol, 4*, 1-12. doi:10.3389/fpsyg.2013.00863

Lavenex, P. B., & Lavenex, P. (2010). Spatial relational learning and memory abilities do not differ between men and women in a real-world, open-field environment. *Behavioral Brain Research, 207*(1), 125-137. doi:10.1016/j.bbr.2009.09.046

Lawton, C. A. (1994). Gender Differences in Way-Finding Strategies - Relationship to Spatial Ability and Spatial Anxiety. *Sex Roles, 30*(11-12), 765-779. doi:10.1007/Bf01544230

Lawton, C. A., Charleston, S. I., & Zieles, A. S. (1996). Individual- and gender-related differences in indoor wayfinding. *Environment and Behavior, 28*(2), 204-219. doi:10.1177/0013916596282003

Lawton, C. A., & Kallai, J. (2002). Gender differences in wayfinding strategies and anxiety about wayfinding: A cross-cultural comparison. *Sex Roles, 47*(9-10), 389-401. doi:10.1023/A:1021668724970

Lawton, C. A., & Morrin, K. A. (1999). Gender differences in pointing accuracy in computer-simulated 3D mazes. *Sex Roles, 40*(1-2), 73-92. doi:10.1023/A:1018830401088

Leonesio, R. J., & Nelson, T. O. (1990). Do Different Metamemory Judgments Tap the Same Underlying Aspects of Memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 16*(3), 464-470. doi:10.1037//0278-7393.16.3.464

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic Status Modifies the Sex Difference in Spatial Skill. *Psychological Science, 16*(11), 841-845. doi:10.1111/j.1467-9280.2005.01623.x 16262766

Lichtenstein, S., Fischhoff, B., & Phillips, L. (1982). Calibration of probabilities: The state of the art to 1980. D. Kahneman, P. Slovic, and A. Tverski (Eds.) Judgment under uncertainty: Heuristics and biases: Cambridge, UK: Cambridge University Press.

Lieberman, D. A. (2004). *Learning and Memory: An Integrative Approach*. Belmont, United States: Thomson/Wadsworth.

Lin, J. J. W., Duh, H. B. L., Parker, D. E., Abi-Rached, H., & Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings IEEE Virtual Reality*, 164-171. doi:10.1109/Vr.2002.996519

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*(6), 1479-1498. doi:10.2307/1130467 4075870

Lundeberg, M. A., Fox, P. W., & Puncochar, J. (1994). Highly Confident but Wrong - Gender Differences and Similarities in Confidence Judgments. *Journal of Educational Psychology, 86*(1), 114-121. doi:10.1037/0022-0663.86.1.114

Martens, J., & Antonenko, P. D. (2012). Narrowing gender-based performance gaps in virtual environment navigation. *Computers in Human Behavior, 28*(3), 809-819. doi:10.1016/j.chb.2012.01.008

Massa, L. J., Mayer, R. E., & Bohon, L. M. (2005). Individual differences in gender role beliefs influence spatial ability test performance. *Learning and Individual Differences, 15*(2), 99-111. doi:10.1016/j.lindif.2004.11.002

Moe, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences, 16*(4), 369-377. doi:10.1016/j.lindif.2007.01.002

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502-517. doi:10.1037/0033-295X.115.2.502 18426301

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language, 93*, 245-258. doi:10.1016/j.jml.2016.10.008

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language, 70*, 1-12. doi:10.1016/j.jml.2013.09.007

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109-133. doi:10.1037/0033-2909.95.1.109 6544431

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The psychology of learning and motivation, 26*, 125-141.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education, 74*(1), 7-28.

Nowak, N. T., Murali, A., & Driscoll, I. (2015). Factors related to sex differences in navigating a computerized maze. *Journal of Environmental Psychology, 43*, 136-144. doi:10.1016/j.jenvp.2015.06.007

O'Laughlin, E. M., & Brubaker, B. S. (1998). Use of landmarks in cognitive mapping: Gender differences in self report versus performance. *Personality and Individual Differences, 24*(5), 595-601. doi:10.1016/S0191-8869%2897%2900237-7

Pfuhl, G., Barrera, L. B. G., Living, M., & Biegler, R. (2013). Do adjustments in search behavior depend on the precision of spatial memory? *Learning & Behavior, 41*(1), 77-93. doi:10.3758/s13420-012-0087-1 22886715

Picucci, L., Caffo, A. O., & Bosco, A. (2011). Besides navigation accuracy: Gender differences in strategy selection and level of spatial confidence. *Journal of Environmental Psychology, 31*(4), 430-438. doi:10.1016/j.jenvp.2011.01.005

Rossano, M. J., & Moak, J. (1998). Spatial representations acquired from computer models: Cognitive load, orientation specificity and the acquisition of survey knowledge. *British Journal of Psychology, 89*(3), 481-497. doi:10.1111/j.2044-8295.1998.tb02698.x

Sandstrom, N. J., Kaufman, J., & Huettel, S. A. (1998). Males and females use different distal cues in a virtual environment navigation task. *Brain Res Cogn Brain Res, 6*(4), 351-360.

Saucier, D. M., Green, S. M., Leason, J., MacFadden, A., Bell, S., & Elias, L. J. (2002). Are sex differences in navigation caused by sexually dimorphic strategies or by differences in the ability to use the strategies? *Behavioral Neuroscience, 116*(3), 403-410. doi:10.1037/0735-7044.116.3.403 12049321

Schwartz, B. L. (2006). The feeling of going: Judgments of learning (JOLs) for maps and directions. *The International Journal of Creativity & Problem Solving, 16*(2), 5-16.

Shore, D. I., Stanford, L., MacInnes, W. J., Klein, R. M., & Brown, R. E. (2001). Of mice and men: Virtual Hebb-Williams mazes permit comparison of spatial learning across species. *Cognitive Affective & Behavioral Neuroscience, 1*(1), 83-89. doi:10.3758/Cabn.1.1.83

Silverman, I., Choi, J., & Peters, M. (2007). The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Arch Sex Behav, 36*(2), 261-268. doi:10.1007/s10508-006-9168-6 17351740

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 204-221. doi:10.1037/0278-7393.26.1.204 10682298

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol, 69*(5), 797-811. doi:10.1037/0022-3514.69.5.797 7473032

Stevens, C. A., & Carlson, R. A. (2016). Are you sure the library is that way? Metacognitive monitoring of spatial judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(7), 1034-1049. doi:10.1037/xlm0000229 26866660

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47*(2), 143-148. doi:10.1016/0001-6918%2881%2990005-6

Therrien, M. E., & Collin, C. A. (2010). Spatial vision meets spatial cognition: Examining the effect of visual blur on human visually guided route learning. *Perception, 39*(8), 1043-1064. doi:10.1068/p5991 20942357

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, .95*(1), 66-73. doi:10.1037/0022-0663.95.1.66

Trochim, W., & Donnelly, J. (2006). The research knowledge methods base. Cincinatti, OH: Atomic Dog Publishing

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 47A*(3), 631-650. doi:10.1080/14640749408401131

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. *Psychonomic Bulletin & Review, 18*(2), 267-277.

Voyer, D., & Sullivan, A. M. (2003). The relation between spatial and mathematical abilities: Potential factors underlying suppression. *International Journal of Psychology, 38*(1), 11-23. doi:10.1080/00207590244000241

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of Sex-Differences in Spatial Abilities - a Metaanalysis and Consideration of Critical Variables. *Psychological Bulletin, 117*(2), 250-270. doi:10.1037/0033-2909.117.2.250

Watier, N., & Collin, C. (2011). Metamemory for faces, names, and common nouns. *Acta Psychologica, 138*(1), 143-154. doi:10.1016/j.actpsy.2011.05.017

Wraga, M., Duncan, L., Jacobs, E. C., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review, 13*(5), 813-819. doi:10.3758/BF03194002 17328378

CHAPTER 6: EXPERIMENT 2

GENDER DIFFERENCES IN JUDGMENTS OF LEARNING AND RETROSPECTIVE CONFIDENCE IN A FIRST-PERSON VIRTUAL MAZE LEARNING TASK

Chantal L. Lemieux, Charles A. Collin
University of Ottawa

Nicholas N. Watier
Brandon University

Author Note

Chantal L. Lemieux, School of Psychology, University of Ottawa; Charles A. Collin, School of Psychology, University of Ottawa; Nicholas N. Watier, Brandon University.

This work was supported by a grant from the *Canadian Natural Sciences and Engineering Research Council* to CC [Grant # 2015-05067]. We thank Joe MacInnis for creating the computer program used to present the virtual mazes.

Abstract

There are well-established gender differences in spatial navigation ability and global metacognition in spatial navigation tasks (i.e., one's overall confidence in one's navigation ability). However, very few studies have examined whether there are gender differences in local metacognitive accuracy, a form of trial-by-trial self-assessment, during spatial navigation tasks. We assessed trial-by-trial metacognitive performance, based on the Nelson and Narens' (1990) metamemory framework, in a sample of men and women engaging in a first-person virtual maze navigation task. In particular, Judgments of Learning and Retrospective Confidence Judgments were used as prospective and retrospective trial-by-trial self-assessments, respectively. Results showed that women had poorer navigation performance and lower prospective and retrospective confidence ratings than men. Trial-by trial self-assessments made after performance were more accurate for men compared with women. There was also evidence to suggest women and men differed in the types of cues they use to assess their past performance, as women's performance was significantly better correlated with completion times. Women were generally underconfident prior to performance, while men showed accurate absolute metacognition prior to performance. These results demonstrate that the gender differences in self-assessment of spatial navigation ability extend to trial-by-trial metacognitive accuracy. Further investigation in this area could help to better understand the persistent gender difference in spatial navigation.

Keywords: metacognition, gender differences, spatial navigation, virtual reality, confidence, metamemory, wayfinding

Gender Differences in Judgments of Learning and Retrospective Confidence in a First-Person Virtual Maze Learning Task

Despite the closing of the gender gap in many cognitive abilities over recent decades, a difference favoring males in spatial cognition has persisted (for reviews, see Coluccia & Louse, 2004; Linn & Petersen, 1985; Martens & Antonenko, 2012; Voyer, Voyer, & Bryden, 1995). More specifically, men, on average, complete mazes faster and with fewer errors than females. A large body of research has also shown that men and women have a tendency to use different navigation strategies and, correspondingly, to rely on different environmental cues for navigation (for a review, see Coluccia & Louse, 2004). For instance, women have a greater tendency to employ a route-based strategy using landmarks at specific decision points when navigating and giving directions, whereas men have a greater tendency to employ an orientation-based strategy using allocentric cardinal points (e.g., Saucier et al., 2002). Some studies have suggested that this gender difference arises due to biological differences in genetics and hormone levels (e.g., Collaer & Hines, 1995; Halari et al., 2005; Hines et al., 2003; Kimura, 1999; Silverman, Choi, & Peters, 2007), while others have revealed sociocognitive factors to have an influence (Casey, 1996; Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005).

Many studies have suggested that confidence differences favoring men (e.g., Lundeberg, Fox, & Puncochar, 1994) or differences in spatial anxiety disfavoring women (e.g., Lawton, 1994; Lawton & Kallai, 2002; Nowak, Murali, & Driscoll, 2015) could help to explain this performance and strategy difference between the genders. The majority of this line of research has used self-report instruments designed to determine overall confidence in relation to specific navigational contexts, such as the Santa Barbara Sense of Direction Scale (Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002), the Spatial Anxiety Scale (Lawton, 1994), and the

International Wayfinding Strategy Scale (Lawton & Kallai, 2002). As these tools ask participants to rate their abilities, confidence, or comfort during specific types of spatial tasks, they are assessing a form of *global metacognition* (e.g., Kelemen, Frost, & Weaver, 2000; Lieberman, 2004). Studies using global measures have shown differences in metacognition between the genders where women generally underestimate their performance compared to men (e.g., Hertzog, Dixon, & Hultsch, 1990). In contrast to this previous work, the present study is concerned with a more fine-grained form of metacognition that has to do with how well men and women predict their future and assess their past performance on individual trials of an ongoing spatial navigation task. This form of trial-by-trial self-assessment is referred to as *local metacognition* (e.g., Hertzog & Dunlosky, 2011; Kelemen et al., 2000; Lieberman, 2004) to distinguish it from more global measures of overall self-assessment.

The Nelson and Narens (1990) metamemory framework was developed as a model of the acquisition of local metacognitive knowledge. According to their framework, local metacognition critically depends on monitoring and control processes. Monitoring involves acquiring information on the current state of a cognitive process so that it can be evaluated according to some goal, whereas control alters the state of a cognitive process based on the output of monitoring. The dynamic between monitoring and control processes is thought to influence performance. For example, in a wayfinding task, monitoring would involve knowing that by continuing to follow the current route, it will take much longer than anticipated to reach the final destination, whereas control would involve adjusting the pace or following an alternative route. Thus, allowing the individual to adjust their behavior in order to better succeed at a task or goal.

Several studies have supported this model and found an association between metacognitive accuracy and performance in the context of classroom-based tasks (e.g., Dunlosky & Ariel, 2011; Dunlosky & Rawson, 2012; Hacker, Bol, Horgan, & Rakow, 2000; Nietfeld, Cao, & Osborne, 2005; Thiede, Anderson, & Therriault, 2003). As monitoring accuracy increases, students are more effective at regulating their study time and their performance increases (e.g., Hacker et al., 2000). Many studies have shown that participants spend more time studying items they previously judged as difficult compared to items they had previously judged as easy (for a review, see Son and Metcalfe (2000). However, recent research has shown that study time allocation (i.e., control) strategies are dependant on task demands (Son & Metcalfe, 2000). For instance, Son and Metcalfe (2000) found that participants allocated more study time to items they had previously judged as easy when study time was restricted and insufficient. More recently, researchers have argued for a bidirectional link between monitoring and control during learning, whereby a control process (spending extra time studying an item) can feed into monitoring ("that item must have been hard if it took so long to study it, therefore it probably won't be remembered") (e.g., Koriat, 2006; Koriat, Ackerman, Adiv, Lockl, & Schneider, 2014).

If women have greater levels of anxiety and men have greater levels of self-confidence related to spatial tasks, they likely have different levels of self-efficacy related to spatial tasks. Self-efficacy is defined by the belief or confidence in one's own abilities and has been shown to determine how individuals approach different goals, challenges and tasks (Bandura, 1997). Perhaps differences in spatial self-efficacy influence the way each gender approaches these tasks via their monitoring and control processes. Indeed, in one recent modeling study, self-efficacy was shown to be the strongest predictor to a global measure of metacognitive awareness (Coutinho & Neuman, 2008). Although no research to date has linked self-efficacy to local metacognition, this idea is supported by a recent review from Dunlosky, Mueller, and Thiede (2016). This review reports that most of the metacognition research to date suggests metacognitive judgments (measures of local metacognition) are influenced by both epistemological beliefs (e.g., subjective feelings of competency) and perceptions of how easy an item was processed (e.g., Frank & Kuhlmann, 2016; Mueller & Dunlosky, 2017; Mueller, Dunlosky, Tauber, & Rhodes, 2014).

Only a few studies have measured local metacognition for navigation tasks, and these used their own independent methodologies, not directly connected to an existing metacognitive framework (e.g., Pfuhl, Barrera, Living, & Biegler, 2013; Stevens & Carlson, 2016). For example, using a directional pointing task, Stevens and Carlson (2016) were able to measure confidence by asking participants to give pointing confidence intervals (larger intervals representing lower confidence). While this study considered gender to be a covariate in the analyses, the authors did not interpret any data related to gender. However, the fixed factor of gender was significant ($p<.001$) showing women to have larger pointing confidence intervals in one of their linear mixed effects regression models (Stevens & Carlson, 2016). Relatedly, using a well-known mental rotation task and collecting confidence scores after each trial, Cooke-Simpson and Voyer (2007) determined that men had more accurate confidence ratings than women.

In contrast to these studies using independent methodologies, the Nelson and Narens (1990) metamemory framework proposes the use of metacognitive monitoring judgments such as the Retrospective Confidence Judgments (RCJs) and Judgments of Learning (JOLs) to measure retrospective and prospective monitoring on a trial-by-trial basis. JOLs are a reflection of a participant’s subjective sense of how well they learned the items. They are strictly collected after the participant has had a chance to learn each item and before recall/performance. In contrast, RCJs are collected after performance, and are a reflection of a participant's subjective sense of how well they performed. In other words, JOLs are an assessment of the probability that the correct target will be retrieved at test, whereas RCJs are an assessment of the probability that an item retrieved from memory at test was the correct target (e.g., Koriat, Pearlman-Avnion, & Ben-Zur, 1998). Traditionally, both JOLs and RCJs are given on a probabilistic scale where the participant indicates their confidence in either a future or past successful performance. Both judgments can be collected on a trial-by-trial basis and represent the participant's ability to monitor their performance for that particular trial. Metacognitive accuracy can then be measured by comparing each trial's actual performance with the metacognitive judgment. For a recent review of metacognitive methodology, see Dunlosky et al. (2016).

While much of the research on spatial navigation has focused on self-assessed global metacognition, only two studies have examined local metacognition in spatial navigation using a metacognition framework. Adapting the Nelson and Narens' (1990) framework, Schwartz (2006) collected participants' trial-by-trial JOLs for learning maps and verbal directions. Participants were asked to study a map or a verbal direction for a fixed time interval, after which a JOL was solicited. Participants were then asked to rate their confidence in their ability to recall either the layout of the map or verbal directions. Afterwards, participants performed a recall task by either writing or drawing the directions from memory. Schwartz noted that participants were twice as accurate in their ability to self-assess future navigation performance in comparison to typical JOL paradigms that use word-pair associates. He argued this could be because spatial memory tasks may have greater ecological validity than word-pair association tasks. In this preliminary study, Schwartz (2006) was able to successfully adapt the Nelson and Narens' (1990) metamemory framework to a spatial memory task and revealed that memory monitoring for navigational information is well above chance levels.

Continuing to investigate along this path, Lemieux, Collin, and Watier (2017) collected monitoring judgments in the context of a first-person virtual navigation task with the goal of comparing men and women on their prospective and retrospective metacognition. Prospective monitoring was determined by using Nelson and Narens' Ease of Learning (EOL) judgment, which represents a subjective sense of how difficult or easy the participant perceives the item will be to learn; in this case, how difficult or easy a maze will be to successfully remember and navigate without errors. Retrospective monitoring was determined using RCJs, adapted to the same virtual navigation task. The results indicated that men were more accurate than women were when assessing their past performance. Despite the fact that the gender difference in EOL accuracy was not significant, men did have EOL accuracy above chance whereas women's EOL accuracy was not statistically above chance.

While this study showed a gender difference in local metacognitive accuracy using RCJs and EOLs, it remains to be determined whether the gender difference extends to other types of monitoring judgments, in particular JOLs. Considering that correlations among prospective monitoring tasks are weak (e.g., Kelemen et al., 2000; Leonesio & Nelson, 1990), neurological impairments differentially affect performance across monitoring tasks (e.g., Modirrousta & Fellows, 2008; Pannu & Kaszniak, 2005), and each prospective monitoring judgment is thought to rely on different sources of information (Schwartz, 1994), gender differences in the accuracy of EOLs during spatial navigation might not necessarily generalize to JOLs. The purpose of the present study is to test this possibility. Additionally, as Lemieux et al. (2017) is the only study to have looked into the intersection between local metacognition, spatial navigation, and gender difference, we wanted to determine if the gender difference in the relative accuracy of retrospective monitoring could be replicated. To do so, we collected data on trial-by-trial metacognitive performance by having participants make prospective Judgments of Learning (JOL) as well as Retrospective Confidence Judgments (RCJ) during a first-person virtual maze-learning task.

In order to remain as comparable as possible with previous studies examining metacognitive monitoring, data collection and analyses were adapted to a spatial navigation task with as little change as possible to typical metacognition paradigms. Local metacognition is typically evaluated using relative and absolute accuracy between confidence scores and performance. A participant's relative metacognitive accuracy determines how well a participant is able to monitor items relative to each other. It is typically assessed using the Goodman-Kruskal gamma correlation, which provides a statistical index of concordance between two ordinal variables, in this case, between both JOL or RCJ ratings and a dichotomous performance assessment (Nelson & Narens, 1990). Consequently, navigation performance was assessed in a dichotomous fashion, where participants were judged to have "passed" a maze if it was successfully completed without making any wrong turns. Otherwise, the trial was judged as a "fail". This was done for a series of thirty first-person virtual mazes. Moreover, absolute metacognitive accuracy (i.e., how generally over or under confident a participant is) was assessed by a comparison of means between observed and predicted performance probabilities for each gender.

Questions & Hypotheses

The goal of this study is to determine whether the genders differ in their prospective (JOL) and retrospective (RCJ) metacognitive accuracy during a first-person virtual spatial navigation task. Based on the Nelson and Narens (1990) metamemory framework, if a gender difference in metacognitive monitoring accuracy is found, it is likely that differences in control processes between the genders also exist due to the dynamic relationship between monitoring and control. As the metacognitive literature implies, differences in control processes could explain, at least in part, differences in navigation performance (Hacker et al., 2000; Nietfeld et al., 2005; Thiede et al., 2003).

Based on a large number of studies that have shown a gender difference in navigation performance favoring men (for a review, see Coluccia & Louse, 2004), we hypothesized women would have poorer navigation performance than men. That is, we predicted that they would exhibit slower maze completion times with a greater number of errors. Similarly, based on a large number of previous studies showing men to have more confidence (e.g., Lawton, 1994; Lawton & Kallai, 2002; Lundeberg et al., 1994); we hypothesized that women would have lower trial-by-trial metacognitive confidence ratings than men (mean JOLs and RCJs). Finally, we hypothesized that women's trial-by-trial prospective (JOL) and retrospective (RCJ) self-assessments would be less accurate compared to men. Women have less experience with tasks that hone spatial cognition skills compared with men (e.g., Lawton & Morrin, 1999) as well as a poorer visuo-spatial working memory (for a review, see Voyer, Voyer, & Saint-Aubin, 2017). Thus, women should have greater difficulty in distinguishing between mazes that will be (or were) correctly or incorrectly navigated.

Methods

Participants

A total of 67 participants were recruited via either the departmental undergraduate subject pool or recruitment posters. All recruited participants gave informed consent, approved by the University of Ottawa Research Ethics Board. Seven participants could not be included in the analyses. Four women, one man, and one person whose gender was not given had to terminate their participation due to simulator sickness. Technical difficulties prevented one man from completing the experiment. Participants who did not identify on the traditional gender binary (i.e., either male or female) and instead indicated "other" when filling out demographic information were to be excluded. Additionally, individuals who indicated a 7 or above on a 10-point scale where 10 was "I often suffer from motion sickness" and 1 was "I never suffer from motion sickness" were excluded. This was done in order to avoid individuals suceptible to simulatior sickness, which is a type of motion sickness likely to occur in the current virtual navigation task as was seen in Lemieux et al. (2017). The final sample was 30 men and 30 women, for a total sample size of 60. Average age was 22.17±4.32 years.

An a priori power analysis was conducted to determine adequate sample size. The power analysis, using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009), was for a one-tailed independent-samples t-test with $\alpha = .05$, and a moderate effect size ($d = .65$; Cohen, 1988). The effect size used in the power analysis was chosen as a compromise between the very large effect size ($d = 1.49$) from a previous similar study that used virtual Hebb-Williams mazes to examine gender differences in spatial navigation (Shore, Stanford, MacInnes, Klein, & Brown, 2001) [7]

---

[7] While a measure of effect size was not directly reported in this study, it could nonetheless be calculated from the reported main effect of gender collapsing across species by exploiting the fact that $F = t^2$, and $d_s = t\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ (Lakens, 2013).

and the moderate effect size (*d* = .58) seen in a study comparing confidence accuracy between genders after a mental rotation task (Cooke-Simpson & Voyer, 2007). The analysis revealed that a minimum sample size of 60 participants would be required to obtain a power of .80.

Materials

The materials were the same or similar as those used in Lemieux et al. (2017). This includes the maze navigation program, maze video tours, and metacognitive instructions.

Virtual Navigation Task

Thirty first-person perspective virtual mazes were created from 10 Hebb-Williams mazes (Hebb & Williams, 1946). These 10 were shown to be the most difficult in past research (Shore et al., 2001; Therrien & Collin, 2010). In addition to the 10 original maze layouts, 20 more were created by either transposing the maze left-to-right relative to the axis linking the entrance to the exit, or reversing the maze by making the entrance the exit and vice versa. See Figure 1 for a top town view of the mazes. The maze task was programmed in C++ using OpenGL. The virtual maze program recorded an error whenever the participant crossed invisible “trip wires” that were strategically placed in dead ends (Shore et al., 2001). Participants navigated using the arrow keys on a typical computer keyboard.

Video tours of the mazes were also generated. The first person perspective view in each video tour moves through the entire maze environment, following either a right-hand or left-hand wall-follower algorithm. Thus, there were two videos per maze, one left-handed and one right-handed. Participants were shown only one randomly selected video per maze. Each video tour was approximately 45 seconds long, and was shown with a screen resolution of 800 x 600 pixels. Videos were played using Windows Media Player. A 45-degree field-of-view was used in all video tours and maze performances, as a low field-of-view has been shown to be ideal for avoiding simulator sickness (e.g., Allen, 2000; DiZio & Lackner, 1998; Lin, Duh, Parker, Abi-Rached, & Furness, 2002). The experiment was run on a standard desktop PC computer running Windows XP with a 19-inch LCD screen. Every maze was designed with the same texture on all wall, floor and ceiling surfaces, which was a fractal noise pattern (see Figure 2). This pattern mimics the spatial frequency content of an average natural scene, but with all structure randomized. The maze program automatically collected the time taken to complete the maze, error counts per maze, and raw navigation data (x/y position and orientation) every 100 ms.

Metacognitive Materials

For the present study, we modified Nelson and Narens' (1990) prospective Judgment of Learning (JOL) instructions but retained its probabilistic scale:

> *"Clearly indicate on a scale of 0-100 going from "not confident" to "confident" with five options "0-25-50-75-100", how confident you are that you learned the maze's layout such that you could navigate from the entrance to the exit without making any errors."*

Similarly, we retained the probabilistic scale of the Retrospective Confidence Judgment (RCJ) but modified the instructions as follows:

> *"Clearly indicate on a scale of 0-100, going from "not confident" to "confident" with five options "0-25-50-75-100", how confident you are that you successfully navigated the maze without making any errors."*

Additional relevant instructions for both were:

> *"Note that an error is defined as not taking the shortest route between start and finish (e.g., going down a dead end, or taking a wrong turn). Please answer as honestly as possible and use the entire scale."*

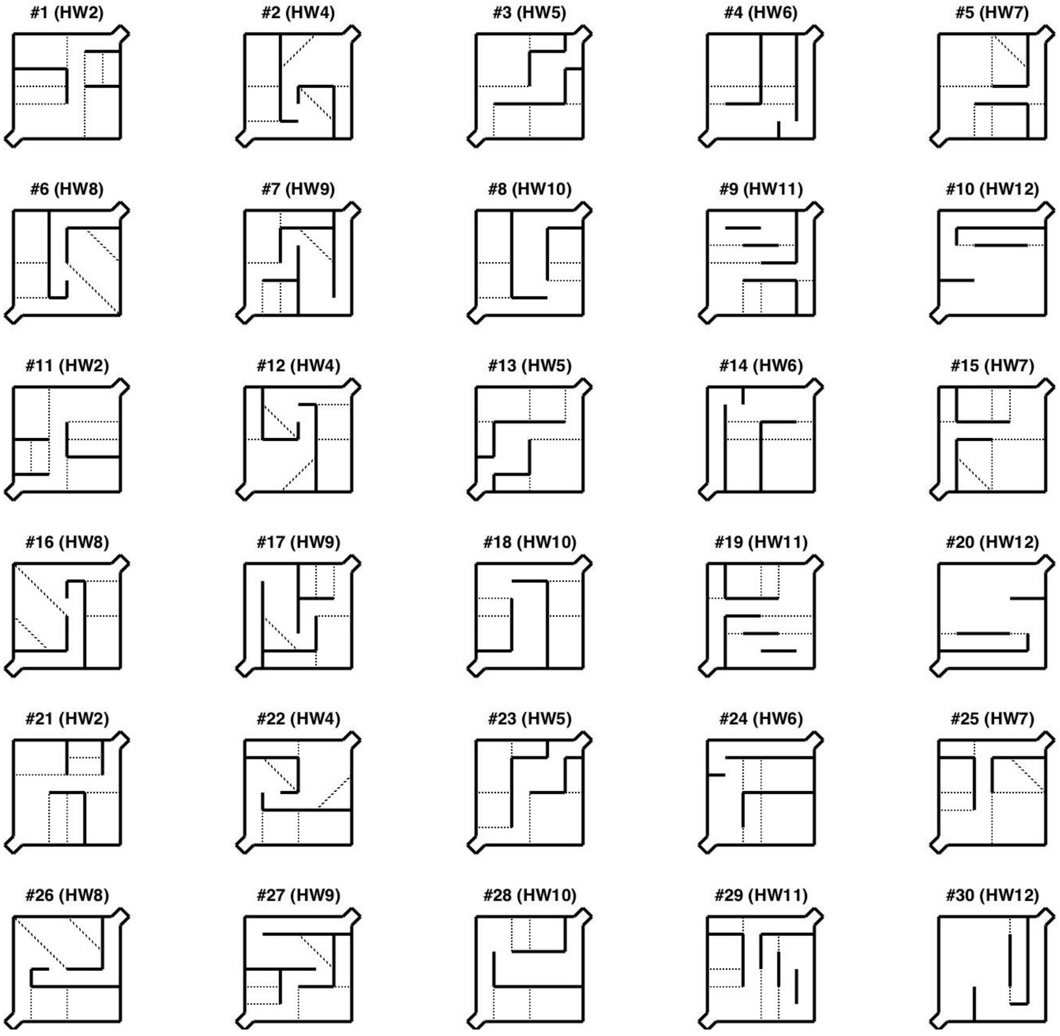


Figure 1. Layouts of the 30 mazes used in this experiment. Mazes 1 to 10 are reproductions of the original 12 Hebb-Williams mazes (Hebb & Williams, 1946), excluding #1 and #3, which were deemed too easy to perform. Mazes 11 to 20 are reversed versions of mazes 1 to 10, where the entrance has been made into the exit, and vice versa. Mazes 21 to 30 are transposed versions of mazes 1-10, where the x and y coordinates of the walls have been exchanged. Dotted lines show virtual “trip wires” replicated from Shore et al. (2001).

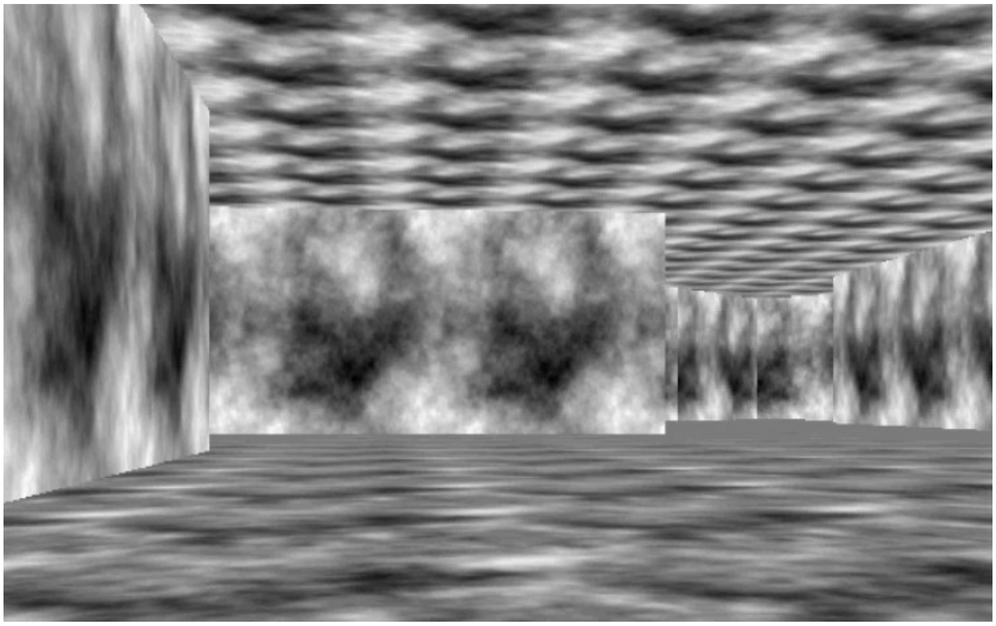


Figure 2. Screen shot demonstrating what participants saw for both the video tours and maze performances.

Procedure

Prior to beginning the experiment, verbal instructions were given. These explained both the learning and testing phases of the experiment, as well the organization of trials into blocks (described below). The metacognitive judgment questions were also presented to the participants and explained. Participants were informed that each maze had a square layout and that the exit was always diagonally opposite to the entrance. Participants performed a single practice trial, consisting of viewing one maze tour video and then attempting to navigate that same maze.

Participants performed 5 blocks of trials, each consisting of 6 learning presentations followed by 6 testing sessions. Each learning presentation involved viewing a video tour of one of the mazes, and giving a JOL judgment. After viewing the 6 tours, participants were tested on how well they could navigate the 6 mazes they had just viewed. After being tested on each maze, participants were asked to make an RCJ judgment. The presentation order of the 6 testing sessions in each block were randomized and different from that of the corresponding learning phase. Consequently, before the beginning of each testing session, in order to help with learning transfer, the experimenter told the participant which maze they were about to perform in relation to the learning phase (e.g., "This was the third maze you saw").

Between blocks, participants were given a 5 minute break (approximately every 15 minutes). This process was repeated for all 5 blocks. Participants were discouraged from taking breaks in the middle of blocks to keep learning conditions consistent. If a participant was ever lost during a maze performance for more than 5 minutes, the experimenter gave them the directions to exit and moved on. This only occurred for two trials (approximately 0.1% of all trials) which were included in the outlier analysis. In total, testing took approximately 80 minutes. The experimenter was in the room for both phases in order to give instructions throughout, assign consistent breaks, play the video tours, start the maze program, and record the metacognitive judgments. Metacognitive judgments were recorded on a paper form by the experimenter as the participant provided their JOL and RCJ scores aloud.

To summarize, the 30 mazes were divided into 5 blocks of 6 mazes each. Each block had a learning phase (6 maze video tours) followed by a performance phase (6 maze performances of the same mazes shown in the learning phase, but in a different order). This block design was done in order to ensure learning and memory transfer from video tour to maze performance without hitting a performance ceiling. Considering JOLs are a determinant of the level of confidence that a trial was learned well enough to succeed at a future performance, our task was organized such that it did encourage a conscious effort to learn the maze layout while viewing the video tours. That is, it was assumed that participants would be able to recall (at least partially) 6 maze layouts.

Results

Univariate outlier analysis was conducted on all dependent variables (see description below for each DV). In accordance with our sample size of 30 individuals per condition, an outlier was defined as having a z score above 2.43 (Van Selst & Jolicoeur, 1994). Outliers were winsorized (Hogg, 1979; Huber, 2011) in order to retain an adequate sample size. In other words, their scores were changed ("tucked") to the closest non-outlying score. Overall, 2% of scores were considered univariate outliers over two iterations and were winsorized. All DVs were considered normally distributed as skewness and kurtosis scores were all near 0 within a ±2 criterion range (Field, 2009; Gravetter & Wallnau, 2016; Trochim & Donnelly, 2006). Cohen's d was calculated as a measure of effect size. Variables with unequal variances, via Levene's Test of Equality of Variances, had their degrees of freedom adjusted as equal variances were not assumed.

In addition to the performance variables recorded by the task program (Error Count and Maze Completion Time), a dependant variable named *Mazes Passed* was created by recoding error count to either a pass or fail (1 or more errors = fail) for each of the 30 mazes. Relative accuracy for JOLs and RCJs was assessed by the Goodman-Kruskal gamma correlation (Nelson, 1984) between performance (Mazes Passed: pass or fail) and metacognitive confidence ratings (confidence score possibilities: 0, 25, 50, 75, 100) for each trial. One limitation of using gamma as a measure of metacognitive accuracy is that it is bounded between -1 to +1, as with most coefficients of association. This inherently violates the assumptions required by parametric hypothesis tests. To remedy this, gamma can be converted into a standardized measure called G*

(Benjamin & Diaz, 2008)[8]. G* is normally distributed and unbounded (Benjamin & Diaz, 2008). For G*, as with gamma, the higher the score the more accurate the participant's relative metacognitive accuracy. Mean JOL and RCJ ratings were also computed as general indicators of metacognitive confidence. Considering JOLs and RCJs are being collected within the same block, it is important to confirm variable independence. Pearson's bivariate correlation was computed between JOL G* and RCJ G* and revealed a small correlation of .252 which accounts for only 6% shared variance. This supports the metacognitive literature, which has suggested that each monitoring judgment measures different metacognitive processes. That is, research has suggested that JOLs and RCJs rely on different sources of information (e.g., Busey, Tunnicliff, Loftus, & Loftus, 2000; Dougherty, Scheck, Nelson, & Narens, 2005) and are differentially affected by neurological impairments (e.g., Modirrousta & Fellows, 2008; Pannu & Kaszniak, 2005) and encoding manipulation (e.g., study time manipulations affect JOLs but not RCJs; Dougherty et al., 2005).

Absolute metacognitive accuracy is typically analyzed by a comparison between observed and predicted performance. In the present study, this was accomplished with two t-tests comparing observed performance (*Percent* Mazes Passed) with predicted performance (mean JOL or RCJ) for each gender. The analyses generally consisted of a series of one-tailed independent sample t-tests comparing gender for each above described DV. Refer to Table 1 for descriptive statistics, t-test results, and effect sizes.

---

[8] The equation used to transform gamma correlations into G* was : $G^* = \log\left(\frac{\gamma+1}{1-\gamma}\right)$ (Benjamin & Diaz, 2008)

Table 1.

One-tailed independent t-test results and descriptive statistics comparing gender

				Women		Men	
DV	t(df)	Sig.	d	$\bar{x}$	s	$\bar{x}$	s
Maze Completion Time (s)	4.107(47.69)[t]	<.001	1.060	40.50	12.60	29.46	7.62
Error Count	3.295(46.84)[t]	.001	.851	38.30	20.84	23.77	12.22
Mazes Passed	-2.699(58)	.005	.697	16.40	4.88	19.53	4.07
JOL mean	-3.810(58)	<.001	.983	49.31	16.54	64.81	14.94
RCJ mean	-3.354(58)	<.001	.866	64.94	14.21	76.61	12.69
JOL gamma				0.03	0.38	0.04	0.26
JOL G*	-.136(51.47)[t]	.446	.033	0.03	0.35	0.04	0.24
RCJ gamma				0.78	0.15	0.81	0.18
RCJ G*	-1.131(58)	.132	.278	1.07	0.52	1.22	0.56
RCJ and Error Count[p]	.090(58)	.465	.000	-0.63	0.13	-0.63	0.17
RCJ and Maze Completion Time[p]	-2.967(58)	.002	.736	-0.72	0.10	-0.65	0.09

Note. [t] = equal variances not assumed. [p] = average Spearman's Roh correlations between RCJ and performance variables (either Error Count of Maze Complation Time) per participant across all 30 mazes. N=60 (30 females and 30 males)

Gender Differences in Performance

In accordance with our hypotheses, results showed women had significantly longer maze completion times than men ($t = 4.107$, $df = 47.69$, $p < .001$) with a large effect size of 1.060.

Women also committed a significantly greater number of errors (*t* = 3.295, *df* = 46.84, *p* = .001) compared to men, with a large effect size of .851. They also averaged fewer successfully completed mazes (*t* = -2.699, *df* = 58, *p* = .005) compared to men, with a moderate effect size of .697. These results confirmed our prediction that women would have poorer maze navigation performance compared to men.

Gender Differences in Metacognitive Confidence and Relative Metacognitive Accuracy

In accordance with our hypotheses, women indicated significantly lower confidence scores on their mean JOLs (*t* = -3.810, *df* = 58, *p* < .001) as well as their mean RCJs (*t* = -3.354, *df* = 58, *p* < .001) in comparison to men, with strong effect sizes (*d* = .983 and *d* = .866, respectively). These findings confirm our predictions that women would have significantly lower trial-by-trial metacognitive confidence compared to men.

Contrary to our hypotheses, men and woman did not significantly differ in their prospective metacognitive accuracy as measured by JOL G* (*t* = -.136, *df* = 51.47, *p* = .446), nor in their retrospective metacognitive accuracy as measured by RCJ G* (*t* = -1.131, *df* =58, *p* = .132). However, the gender difference for RCJ G* still had an effect size of .278, which suggests a small effect favoring men. Interestingly, JOL G* scores among both genders are quite low ($\bar{x}$ = .03 for women and $\bar{x}$ = .04 for men) indicating very poor prospective metacognitive accuracy.

In order to better understand the potential gender difference in retrospective metacognitive accuracy, non-parametric Spearman's Roh correlations were computed between RCJ and the two other performance variables (Error Count and Maze Completion Time) for each participant across all 30 mazes. One-tailed independent samples t-tests were run comparing the average correlation between both genders for each pair of variables (RCJ and Error Count, RCJ and Maze Completion Time). The results show that women have significantly higher correlations

(r_s = -.72) than men (r_s = -.65) between RCJ and Maze Completion Time (t = -2.967, df = 58, p = .002) with a large effect size of .736. However, RCJ and Error Count was not significantly different between the genders (t = .090, df = 58, p = .465).

Gender Differences in Absolute Metacognitive Accuracy

Two-tailed paired sample t-tests were conducted comparing observed performance (Percent Mazes Passed) with predicted performance (RCJ mean or JOL mean) for each gender (see Figure 3). The difference between prospective predicted performance (JOL mean) and observed performance was not statistically significant for either gender (Female: t = -1.302, df = 29, p = .203. d = .329; Male: t = -.098, df = 29, p = .922, d = .052). Nonetheless, looking at the group means, women's observed performance ($\bar{x}$ = 54.67, s = 16.27) was higher than their prospective predicted performance ($\bar{x}$ = 49.31, s = 16.54) showing a small to moderate effect size of .329, while men had similar means between their observed performance ($\bar{x}$ = 65.11, s = 13.58) and their predicted performance ($\bar{x}$ = 64.81, s = 14.94) showing no effect. However, there were gender differences when comparing retrospective predicted performance (RCJ mean) and observed performance (Female: t = 3.050, df = 29, p = .005; Male: t = 5.255, df = 29, p < .001). Both genders were significantly overconfident when assessing past performance, as their observed performance (Women: $\bar{x}$ = 54.67, s = 16.27; Men: $\bar{x}$ = 65.11, s = 13.58) was significantly lower than their predicted performance (Women: $\bar{x}$ = 64.94, s = 14.21; Men: $\bar{x}$ = 76.61, s = 12.69). The effect size was larger for men (d = .847) than for women (d = .631).

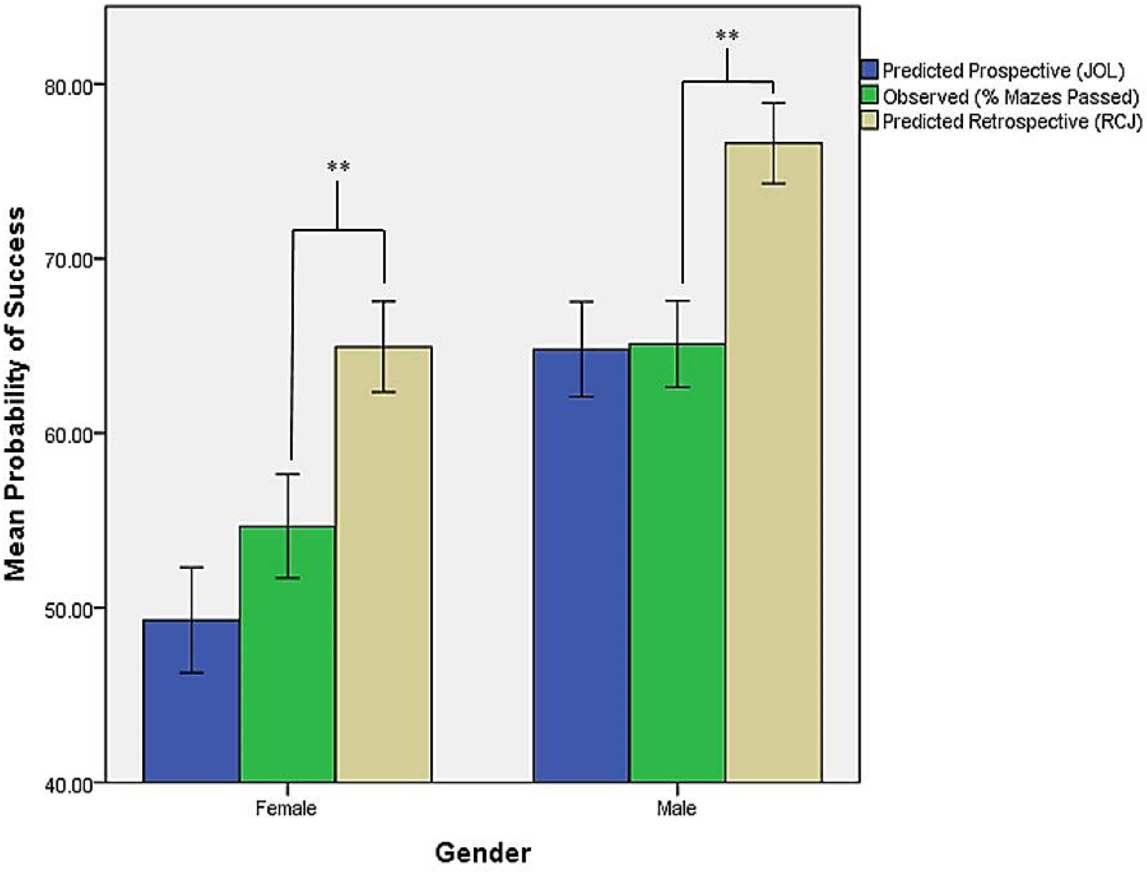


Figure 3. Bar graphs comparing observed performance (% Mazes Passes) with predicted performance (either JOL or RCJ mean) for each gender. **p<.001. One SE is used for error bars. N=60 (30 females and 30 males)

Discussion

In line with previous research on gender differences in spatial navigation (for a review, see Coluccia & Louse, 2004), the results show that women exhibited poorer virtual navigation performance with moderate to large effect sizes. That is to say, women failed a greater number of mazes, took longer to complete the mazes, and made a greater number of navigational errors in comparison to men. The greatest difference was seen with maze completion times ($d = 1.06$), where women took, on average, over 11 seconds more than men, against an average completion time of about 35 seconds. These results replicate that which was found in a similar experiment by Lemieux et al. (2017).

There are several possible explanations for why women may take longer than men when navigating. Several studies have suggested it may take longer for women to change from their default route-based strategy to an orientation-based strategy, and that this explains their longer latencies compared with men (e.g., Barkley & Gabriel, 2007; Lawton, 1994; Sandstrom, Kaufman, & Huettel, 1998). This could be at play in the current study considering our first-person virtual navigation task was designed to force participants to use an orientation-based strategy as it was devoid of overt landmarks. In other words, women may have taken longer to complete the mazes because they needed to shift their strategy whereas the task already favors the default strategy typically used by men.

Other studies have demonstrated that during open exploration tasks, women generally travel slower and display more hesitant behaviors (e.g., Lavenex & Lavenex, 2010; Lawton, Charleston, & Zieles, 1996). It is possible the female participants in our sample generally travelled slower with more hesitance than the male participants. However, it is important to note that, in the current study, movement speed through the maze was fixed, so any difference in travel time across a given distance would necessarily be based on behavioral differences such as hesitance. Hesistance, as a behavior, may be related to a lack of self-confidence and/or to anxiety which has been shown to either influence metacognition or metacognitive monitoring (e.g., Coutinho & Neuman, 2008; Kleitman & Stankov, 2007; McDonald-Miszczak, Gould, & Tychynski, 1999; Stankov, 1999; Tobias & Everson, 1997; Veenman, Kerseboom, & Imthorn, 2000; West, Boatwright, & Schleser, 1984). It is therefore possible that women may exhibit a greater degree of hesitance, which may be related to a different monitoring – control dynamic compared to men. Nonethless, several studies have shown that without a time pressure on spatial tasks the performance difference between genders disapears (e.g., Goldstein, Haldane, & Mitchell, 1990; Voyer & Sullivan, 2003) or is at least reduced (Voyer, 2011). It is still not well understood if this is because women require more time to perform and/or if their performance is impacted by stress or anxiety to a greater degree than men's performance.

Our hypotheses related to confidence were also confirmed as women indicated lower trial-by-trial confidence in their performance compared to men both prospectively (JOL) and retrospectively (RCJ) with large effect sizes. There is a considerable amount of research demonstrating confidence differences between the genders in a variety of fields, including spatial navigation (e.g., Lawton et al., 1996; O'Laughlin & Brubaker, 1998; Picucci, Caffo, & Bosco, 2011; Stevens & Carlson, 2016), and there was no reason to hypothesize that our results would be any different. Though this is one of few studies to confirm confidence differences in a trial-by trial basis. Interestingly, the confidence difference between the genders was greater for JOLs than for RCJs, which is reflected in their corresponding effect sizes (d = .983 and d = .866, respectively). In both this study and Lemieux et al. (2017), the gender difference was greater for judgments made prior to performance compared to judgments made after performance. Considering subjective feelings of competency have been shown to influence metacognitive monitoring (e.g., Frank & Kuhlmann, 2016; Mueller & Dunlosky, 2017; Mueller et al., 2014), and that there is a relatively robust stereotype that women have poor navigational skills (e.g., Harris, 1981; Hausmann, Schoofs, Rosenthal, & Jordan, 2009), it is possible that women have especially lower confidence prior to performance compared to men due to the effects of stereotype threat (Steele & Aronson, 1995) and that this influences their metacognitive monitoring. Indeed, the effects of stereotype have already been shown to influence women's performance on a variety of tasks related to spatial cognition (e.g., Doyle & Voyer, 2016; Massa, Mayer, & Bohon, 2005; Moe & Pazzaglia, 2006; Wraga, Duncan, Jacobs, Helt, & Church, 2006). Another study also found that confidence mediated the gender difference in mental rotation (Estes & Felker, 2012). It is unclear how the confidence differences between the genders influence their navigation performance. However, future research is needed to further investigate whether the effects of stereotype are at play in women's navigation performance and whether this is related to metacognition.

The hypothesis that women would have poorer relative prospective (JOL G*) and retrospective (RCJ G*) metacognitive accuracy than men, could not be entirely confirmed based on the results of the current study. While the difference between genders on retrospective metacognitive accuracy did not reach statistical significance, the effect size ($d = .278$) is non-trivial, and on the same order of magnitude as the effect of gender on other cognitive tasks (Hedges & Nowell, 1995). Indeed, the effect is in line with Lemieux et al. (2017), the only other study to perform such an investigation, which found that men had significantly more accurate retrospective relative metacognition (RCJ) compared to women ($d = .594$). It is possible the difference in retrospective metacognition in the current study did not reach statistical significance due to a lack of power. Our sample size was based on an a priori power analysis using a moderate effect size ($d = .65$). Using the current effect size of gender on *retrospective monitoring accuracy* ($d = .278$) with only 80% power, a sample size of 322 would have been required to detect the effect, compared to 60 in the current analysis. Thus, our analysis is underpowered to detect such a small effect.

As no significant differences were found for retrospective metacognitive accuracy (RCJ G*), we conducted additional analyses of the correlations between RCJ and other performance variables (Error Count and Maze Completion Time). However, it is important to note that interpretation of these correlations is difficult, as RCJs were collected by specifically asking about confidence in relation to successfully performing the mazes without errors or not (relating to Mazes Passed) and not about which mazes would have fewer errors than others (which would relate better to Error Count) or which mazes will be completed more quickly than others (which would relate better to Maze Completion Time). Nonetheless, this analysis found there was a significant gender difference, favoring women, in the correlation between RCJ and Maze Completion Time. That is, the correlation between RCJ scores and the time it took to complete each maze was higher for women than for men. One possible explanation for this is that women had greater variability in Maze Completion Time (s of 12.6 seconds for women and 7.6 seconds for men). They may, therefore, have been better able to use this variable as a cue to distinguish between mazes when assessing their past performance. This is in line with recent studies in metacognition, which have found that control processes such as the time it takes to complete a task can be used as a cue during monitoring processing (Koriat, 2006; Koriat et al., 2014). However, this is in opposition to what was found in Lemieux et al. (2017), who found that men had significantly stronger correlations between RCJ and Maze Completion Time compared to women.

As with the current experiment, Lemieux et al., (2017) did not find a difference between the genders for prospective relative metacognitive accuracy, though this may have been due to a floor effect. Similar to Lemieux et al. (2017) we found that prospective metacognitive accuracy was very poor, with JOL G* scores among both genders being close to 0 ($\bar{x} = .03$ for women and $\bar{x} = .04$ for men). A possible explanation for this finding is that the video tours were viewed passively and involved travelling through the entire maze, including incorrect passages (dead-ends), making it difficult to distinguish between well-learned mazes and poorly learned mazes. Indeed, research has shown lower retention rates in spatial navigation tasks when using passive learning versus active learning (Chrastil & Warren, 2012). Another possible explanation could be that our task was too metacognitively difficult, requiring participants to remember the layouts of 6 different mazes for each block before performing them in random order. The Nelson and Narens' metamemory framework predicts higher accuracy when the information used to make the judgment is more similar to the information needed at time of test. The competition hypothesis also states that as memory interference increases monitoring accuracy decreases (e.g., McGuire & Maki, 2001; Schreiber & Nelson, 1998). Considering JOLs were collected after each video tour, after which there was the potential for other video tours and performances interfering with memory, it is possible our participants lost a lot of what they had learned. In other words, the information they used to make the JOL for a particular maze was lost by the time they performed the maze. It is also possible that at the time of giving a JOL they could not accurately predict learning interference since the performance order was randomized. That is, the task may have been too metacognitively difficult, preventing participants from distinguishing between well learned mazes and not well learned mazes. This, consequently, resulted in a floor effect thereby preventing the variability needed to observe gender differences, if any exist.

Regarding absolute metacognitive accuracy, our results indicate that both genders exhibit overconfidence when assessing a previous performance. This corroborates what was generally found in Lemieux et al. (2017). Despite our results not reaching statistical significance for either gender for prospective absolute accuracy, when examining Figure 3, one can see that women's average predicted performance was lower than their observed performance with a small effect size, implying they were underconfident prior to performance. This is in contrast to men's average predicted performance, which is very close to observed performance, implying their confidence prior to performance was more accurate.

Our findings showed strong gender differences in overall confidence levels before and after performance, though the difference was larger for confidence judgments made prior to performance. This is mirrored by our findings that women have poorer absolute metacognitive accuracy prior to performance compared to men. Despite not finding a difference between the genders for either prospective or retrospective metacognitive accuracy, there was evidence to suggest women and men differed in the types of cues they use to make retrospective judgments. These findings point to the need to assess not only global metacognitive factors, such as overall confidence in one's navigation abilities, but also local metacognitive factors, as the two are not necessarily directly related.

In agreement with results found by Lemieux et al (2017) the relative accuracy gamma correlations for RCJs were slightly high in the present study[9] compared to word-pair association tasks used in most local metacognitive studies (Koriat, 2007). This may be because navigation tasks have greater ecological validity compared to word-pair association tasks and, therefore might yield more developed and more accurate metacognition (Schwartz, 2006). The JOLs collected in the present study are very low compared to the literature and likely suffered from a floor effect, as participants were not able to distinguish between well-learned mazes and poorly learned mazes. Future research should consider the high cognitive load navigation requires to prevent memory interference in order to establish adequate sensitivity when trying to measure metacognitive monitoring.

---

[9] Although G* was used in the analyses, RCJ gamma correlation means were .78 for women and .81 for men. See Table 1 for gamma correlation descriptive statistics by gender.

One limitation of this study is a possible sampling bias. There was a greater number of women who experienced simulator sickness compared to men (4 versus 1) out of those who were recruited. More importantly, recruitment was limited to those who did not indicate a high susceptibility to motion sickness, which may have excluded a greater number of women compared to men. Indeed, women have been shown to have a greater susceptibility to simulator sickness compared to men (Classen, Bewernitz, & Shechtman, 2011). This was also seen in Lemieux et al. (2017), which used the same virtual navigation task as the present study. This means we may have excluded a larger sample of female participants with possible visual motion perception issues, which could contribute to differences in how they navigate in the real world and, therefore, how they would have performed in our experiment. Consequently, women who can tolerate virtual 3D environments are potentially over represented in our sample compared to the actual female population.

Additionally, if men experience less simulator sickness then they may be generally better performers in virtual environments. This would make it very hard to compare their performance with that of women who may not be as comfortable within virtual environments. This is mirrored by the possibility that our sample of men may have had more experience playing video games and/or with computers, leading them to feeling more comfortable using the keypad within the virtual 3D environment. Indeed, studies have shown that men have more experience with computers (Cooper, 2006) and video games (Lucas & Sherry, 2004). However, the findings from the literature are variable, ranging from there being no effect of video game experience on performance (e.g., Stinchcombe, Kadulina, Lemieux, Aljied, & Gagnon, 2017) to a variety of improved cognitive processes such as processing speed and visual skills (e.g., Achtman, Green, & Bavelier, 2008; Dye, Green, & Bavelier, 2009). Because information on video game experience was not collected in the present study, there is no way to determine whether there was a difference in video game experience between genders in our sample and whether this would have an effect on their performance or metacognitive accuracy. However, a previous study, using the same first-person virtual spatial navigation task as the present study, found no effect of video game experience on performance (Therrien & Collin, 2010). Nevertheless, gender differences in navigation have been found among real-world navigation studies (e.g., Galea & Kimura, 1993; Halpern, 2013; Rossano & Moak, 1998), though they are not as strong and robust as seen in virtual environments (for a review, see Coluccia & Louse, 2004). It is important to take into consideration both these limitations when interpreting the results between genders for any study using virtual reality.

In summary, women showed, as expected, poorer navigational performance and lower overall confidence. The gender difference for confidence was larger when the judgment was made before the navigational task compared to when it was made after. There was a small effect favoring men in retrospective relative accuracy, which suggests that self-assessments after navigation performance are more accurate for men compared with women. There was also evidence to suggest that women and men differed in the types of cues they may use to assess their past performance. Women also exhibited underconfidence prior to performance while men showed accurate absolute metacognition prior to performance, and both were overconfident after performance. These results corrobotate the only other study to investigate such hypotheses (Lemieux et al., 2017), and suggest that women are more likely to err in assessing their navigational performance, and thus may be less likely to undertake corrective control actions for a future performance (Nelson & Narens, 1990). Additional research is necessary in order to establish a causal link between gender differences in monitoring and gender differences in control when performing navigation tasks. While a causal connection between metacognitive monitoring and control in spatial navigation tasks still needs to be determined, the current results point to the possibility that the gender differences seen on navigational tasks may, at least in part, be explained by metacognitive factors. Additionally, women's lack of confidence prior to performance could be due to stereotype threat (Steele & Aronson, 1995). Future research is needed to investigate the potential effects of stereotype threat/lift on metacognition in navigation tasks.

References

Achtman, R. L., Green, C. S., & Bavelier, D. (2008). Video games as a tool to train visual skills. *Restorative Neurology and Neuroscience, 26*(4-5), 435-446.

Allen, R. C. (2000). *The effect of restricted field of view on locomotion tasks, head movements, and motion sickness.* (Doctoral Dissertation), University of Central Florida, Retrospective Theses and Dissertations: 1758. Retrieved from http://stars.library.ucf.edu/rtd/1758

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W H Freeman/Times Books/Henry Holt & Co.

Barkley, C. L., & Gabriel, K. I. (2007). Sex differences in cue perception in a visual scene: Investigation of cue type. *Behavioral Neuroscience, 121*(2), 291-300. doi:10.1037/0735-7044.121.2.291 17469918

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 73-94). New York, NY: Psychology Press.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7*(1), 26-48. doi:10.3758/Bf03210724

Casey, M. B. (1996). Gender, sex, and cognition: Considering the interrelationship between biological and environmental factors. *Learning and Individual Differences, 8*(1), 39-53. doi:10.1016/S1041-6080%2896%2990006-0

Chrastil, E. R., & Warren, W. H. (2012). Active and passive contributions to spatial learning. *Psychon Bull Rev, 19*(1), 1-23. doi:10.3758/s13423-011-0182-x

Classen, S., Bewernitz, M., & Shechtman, O. (2011). Driving simulator sickness: an evidence-based review of the literature. *Am J Occup Ther, 65*(2), 179-188.

Cohen, J. (1988). *Statistical analysis for the behavioral sciences* (2nd ed.). Hillsdale: NJ: Lawrance Erlbaum Associates.

Collaer, M. L., & Hines, M. (1995). Human behavioral sex differences: a role for gonadal hormones during early development? *Psychological Bulletin, 118*(1), 55-107.

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology, 24*(3), 329-340. doi:10.1016/j.jenvp.2004.08.006

Cooke-Simpson, A., & Voyer, D. (2007). Confidence and gender differences on the mental rotations test. *Learning and Individual Differences, 17*(2), 181-186. doi:10.1016/j.lindif.2007.03.009

Cooper, J. (2006). The digital divide: The special case of gender. *Journal of Computer Assisted Learning, 22*(5), 320-334.

Coutinho, S. A., & Neuman, G. (2008). A model of metacognition, achievement goal orientation, learning style and self-efficacy. *Learning Environments Research, 11*(2), 131-151. doi:10.1007/s10984-008-9042-7

DiZio, P., & Lackner, J. (1998). *Alleviation of motion sickness and postural instability during and after virtual environment exposure*. Retrieved from Naval Air Warfare Center Training Systems Division: Orlando, FL

Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Mem Cognit, 33*(6), 1096-1115.

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences, 47*, 103-116. doi:10.1016/j.lindif.2015.12.018

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 103-140). San Diego, CA: Elsevier Academic Press.

Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2016). Methodology for investigating human metamemory: Problems and pitfalls *The Oxford handbook of metamemory* (pp. 23-37). New York, NY: Oxford University Press.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271-280. doi:10.1016/j.learninstruc.2011.08.003

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science, 18*(6), 321-326. doi:10.1111/j.1467-8721.2009.01660.x 20485453

Estes, Z., & Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. *Arch Sex Behav, 41*(3), 550-570. doi:10.1007/s10508-011-9875-5 22130691

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods, 41*(4), 1149-1160.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, England: Sage Publications.

Frank, D. J., & Kuhlmann, B. G. (2016). More Than Just Beliefs: Experience and Beliefs Jointly Contribute to Volume Effects on Metacognitive Judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*(680-693). doi:10.1037/xlm0000332 27709983

Galea, L. A., & Kimura, D. (1993). Sex differences in route-learning. *Personality and Individual Differences, 14*(1), 53-65. doi:10.1016/0191-8869%2893%2990174-2

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition, 18*(5), 546-550. doi:10.3758/BF03198487 2233267

Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the behavioral sciences* (10th ed.). Belmont, California: Wadsworth Publishing.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*(1), 160-170. doi:10.1037/0022-0663.92.1.160

Halari, R., Hines, M., Kumari, V., Mehrotra, R., Wheeler, M., Ng, V., & Sharma, T. (2005). Sex Differences and Individual Differences in Cognitive Performance and Their Relationship to Endogenous Gonadal Hormones and Gonadotropins. *Behavioral Neuroscience, 119*(1), 104-117. doi:10.1037/0735-7044.119.1.104 15727517

Halpern, D. F. (2013). *Sex differences in cognitive abilities* (4th ed.). New York, NY: Psychology Press.

Harris, L. J. (1981). Sex-related variations in spatial skill. In L. S. Liben, A. H. Patterson, & N. Newcombe (Eds.), *Spatial representation and behavior across the life span: theory and application* (pp. 83-111). Ann Arbor, MI: Academic Press.

Hausmann, M., Schoofs, D., Rosenthal, H. E. S., & Jordan, K. (2009). Interactive effects of sex hormones and gender stereotypes on cognitive sex differences-A psychobiosocial approach. *Psychoneuroendocrinology, 34*(3), 389-401. doi:10.1016/j.psyneuen.2008.09.019

Hebb, D. O., & Williams, K. (1946). A method of rating animal intelligence. *Journal of General Psychology, 34*, 59-65. doi:10.1080/00221309.1946.10544520 21015350

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220), 41-45.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30*(5), 425-447.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Metamemory in adulthood: differentiating knowledge, belief, and behavior. In T. M. Hess (Ed.), *Aging and cognition: knowledge organization and utilization* (pp. 161-212). Oxford, England: Elsevier.

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. *Current Directions in Psychological Science, 20*(3), 167-173. doi:10.1177/0963721411409026 24478539

Hines, M., Fane, B. A., Pasterski, V. L., Mathews, G. A., Conway, G. S., & Brook, C. (2003). Spatial abilities following prenatal androgen abnormality: targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. *Psychoneuroendocrinology, 28*(8), 1010-1026.

Hogg, R. V. (1979). Statistical robustness: One view of its use in applications today. *The American Statistician, 33*(3), 108-115.

Huber, P. J. (2011). *Robust statistics*. Hoboken, NJ: John Wiley & Sons.

Kelemen, W. L., Frost, P. J., & Weaver, C. A., III. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition, 28*(1), 92-107. doi:10.3758/BF03211579 10714142

Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: The MIT Press.

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences, 17*(2), 161-173. doi:10.1016/j.lindif.2007.03.004

Koriat, A. (2006). Are we frightened because we run away? Some evidence from metacognitive feelings *Memory and emotion: Interdisciplinary perspectives* (pp. 83-103). Malden: Blackwell Publishing.

Koriat, A. (2007). Metacognition and consciousness *The Cambridge handbook of consciousness* (pp. 289-325). New York, NY: Cambridge University Press.

Koriat, A., Ackerman, R., Adiv, S., Lockl, K., & Schneider, W. (2014). The effects of goal-driven and data-driven regulation on metacognitive monitoring during learning: A developmental perspective. *Journal of Experimental Psychology: General, 143*(1), 386-403. doi:10.1037/a0031768 23421442

Koriat, A., Pearlman-Avnion, S., & Ben-Zur, H. (1998). The subjective organization of input and output events in memory. *Psychological Research-Psychologische Forschung, 61*(4), 295-307. doi:10.1007/s004260050034

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol, 4*, 1-12. doi:10.3389/fpsyg.2013.00863

Lavenex, P. B., & Lavenex, P. (2010). Spatial relational learning and memory abilities do not differ between men and women in a real-world, open-field environment. *Behavioral Brain Research, 207*(1), 125-137. doi:10.1016/j.bbr.2009.09.046

Lawton, C. A. (1994). Gender Differences in Way-Finding Strategies - Relationship to Spatial Ability and Spatial Anxiety. *Sex Roles, 30*(11-12), 765-779. doi:10.1007/Bf01544230

Lawton, C. A., Charleston, S. I., & Zieles, A. S. (1996). Individual- and gender-related differences in indoor wayfinding. *Environment and Behavior, 28*(2), 204-219. doi:10.1177/0013916596282003

Lawton, C. A., & Kallai, J. (2002). Gender differences in wayfinding strategies and anxiety about wayfinding: A cross-cultural comparison. *Sex Roles, 47*(9-10), 389-401. doi:10.1023/A:1021668724970

Lawton, C. A., & Morrin, K. A. (1999). Gender differences in pointing accuracy in computer-simulated 3D mazes. *Sex Roles, 40*(1-2), 73-92. doi:10.1023/A:1018830401088

Lemieux, C. L., Collin, C. A., & Watier, N. N. (2017). *Gender Differences in Metacognitive Judgments and Performance on a First-Person Virtual Maze Learning Task*. Manuscript submitted for publication.

Leonesio, R. J., & Nelson, T. O. (1990). Do Different Metamemory Judgments Tap the Same Underlying Aspects of Memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 16*(3), 464-470. doi:10.1037//0278-7393.16.3.464

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic Status Modifies the Sex Difference in Spatial Skill. *Psychological Science, 16*(11), 841-845. doi:10.1111/j.1467-9280.2005.01623.x 16262766

Lieberman, D. A. (2004). *Learning and Memory: An Integrative Approach*. Belmont, United States: Thomson/Wadsworth.

Lin, J. J. W., Duh, H. B. L., Parker, D. E., Abi-Rached, H., & Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings IEEE Virtual Reality*, 164-171. doi:10.1109/Vr.2002.996519

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*(6), 1479-1498. doi:10.2307/1130467 4075870

Lucas, K., & Sherry, J. L. (2004). Sex Differences in Video Game Play: A Communication-Based Explanation. *Communication Research, 31*(5), 499-523. doi:10.1177/0093650204267930

Lundeberg, M. A., Fox, P. W., & Puncochar, J. (1994). Highly Confident but Wrong - Gender Differences and Similarities in Confidence Judgments. *Journal of Educational Psychology, 86*(1), 114-121. doi:10.1037/0022-0663.86.1.114

Martens, J., & Antonenko, P. D. (2012). Narrowing gender-based performance gaps in virtual environment navigation. *Computers in Human Behavior, 28*(3), 809-819. doi:10.1016/j.chb.2012.01.008

Massa, L. J., Mayer, R. E., & Bohon, L. M. (2005). Individual differences in gender role beliefs influence spatial ability test performance. *Learning and Individual Differences, 15*(2), 99-111. doi:10.1016/j.lindif.2004.11.002

McDonald-Miszczak, L., Gould, O. N., & Tychynski, D. (1999). Metamemory predictors of prospective and retrospective memory performance. *Journal of General Psychology, 126*(1), 37-52. doi:10.1080/00221309909595350

McGuire, M. J., & Maki, R. H. (2001). When knowing more means less: The effects of fan on metamemory judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1172-1179. doi:10.1037/0278-7393.27.5.1172 11550745

Modirrousta, M., & Fellows, L. K. (2008). Medial prefrontal cortex plays a critical and selective role in 'feeling of knowing' meta-memory judgments. *Neuropsychologia, 46*(12), 2958-2965. doi:10.1016/j.neuropsychologia.2008.06.011 18606176

Moe, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences, 16*(4), 369-377. doi:10.1016/j.lindif.2007.01.002

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language, 93*, 245-258. doi:10.1016/j.jml.2016.10.008

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language, 70*, 1-12. doi:10.1016/j.jml.2013.09.007

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109-133. doi:10.1037/0033-2909.95.1.109 6544431

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The psychology of learning and motivation, 26*, 125-141.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education, 74*(1), 7-28.

Nowak, N. T., Murali, A., & Driscoll, I. (2015). Factors related to sex differences in navigating a computerized maze. *Journal of Environmental Psychology, 43*, 136-144. doi:10.1016/j.jenvp.2015.06.007

O'Laughlin, E. M., & Brubaker, B. S. (1998). Use of landmarks in cognitive mapping: Gender differences in self report versus performance. *Personality and Individual Differences, 24*(5), 595-601. doi:10.1016/S0191-8869%2897%2900237-7

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychol Rev, 15*(3), 105-130. doi:10.1007/s11065-005-7091-6 16328731

Pfuhl, G., Barrera, L. B. G., Living, M., & Biegler, R. (2013). Do adjustments in search behavior depend on the precision of spatial memory? *Learning & Behavior, 41*(1), 77-93. doi:10.3758/s13420-012-0087-1 22886715

Picucci, L., Caffo, A. O., & Bosco, A. (2011). Besides navigation accuracy: Gender differences in strategy selection and level of spatial confidence. *Journal of Environmental Psychology, 31*(4), 430-438. doi:10.1016/j.jenvp.2011.01.005

Rossano, M. J., & Moak, J. (1998). Spatial representations acquired from computer models: Cognitive load, orientation specificity and the acquisition of survey knowledge. *British Journal of Psychology, 89*(3), 481-497. doi:10.1111/j.2044-8295.1998.tb02698.x

Sandstrom, N. J., Kaufman, J., & Huettel, S. A. (1998). Males and females use different distal cues in a virtual environment navigation task. *Brain Res Cogn Brain Res, 6*(4), 351-360.

Saucier, D. M., Green, S. M., Leason, J., MacFadden, A., Bell, S., & Elias, L. J. (2002). Are sex differences in navigation caused by sexually dimorphic strategies or by differences in the ability to use the strategies? *Behavioral Neuroscience, 116*(3), 403-410. doi:10.1037/0735-7044.116.3.403 12049321

Schreiber, T. A., & Nelson, D. L. (1998). The relation between feelings of knowing and the number of neighboring concepts linked to the test cue. *Memory & Cognition, 26*(5), 869-883. doi:10.3758/BF03201170 9796223

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review, 1*(3), 357-375. doi:10.3758/BF03213977 24203520

Schwartz, B. L. (2006). The feeling of going: Judgments of learning (JOLs) for maps and directions. *The International Journal of Creativity & Problem Solving, 16*(2), 5-16.

Shore, D. I., Stanford, L., MacInnes, W. J., Klein, R. M., & Brown, R. E. (2001). Of mice and men: Virtual Hebb-Williams mazes permit comparison of spatial learning across species. *Cognitive Affective & Behavioral Neuroscience, 1*(1), 83-89. doi:10.3758/Cabn.1.1.83

Silverman, I., Choi, J., & Peters, M. (2007). The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Arch Sex Behav, 36*(2), 261-268. doi:10.1007/s10508-006-9168-6 17351740

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 204-221. doi:10.1037/0278-7393.26.1.204 10682298

Stankov, L. (1999). Mining on the “no man’s land” between intelligence and personality. In P. L. Ackerman, Kyllonen, P. C. & Roberts R. D. (Ed.), *Learning and Individual Differences: Process, Trait, and Content Determinants.* (pp. 315-337). Washington, DC: American Psychological Association.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol, 69*(5), 797-811. doi:10.1037/0022-3514.69.5.797 7473032

Stevens, C. A., & Carlson, R. A. (2016). Are you sure the library is that way? Metacognitive monitoring of spatial judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(7), 1034-1049. doi:10.1037/xlm0000229 26866660

Stinchcombe, A., Kadulina, Y., Lemieux, C. L., Aljied, R., & Gagnon, S. (2017). Driving is not a game: Video game experience is associated with risk-taking behaviors in the driving simulator. *Computers in Human Behavior, 69*, 415-420. doi:10.1016/j.chb.2016.12.006

Therrien, M. E., & Collin, C. A. (2010). Spatial vision meets spatial cognition: Examining the effect of visual blur on human visually guided route learning. *Perception, 39*(8), 1043-1064. doi:10.1068/p5991 20942357

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, .95*(1), 66-73. doi:10.1037/0022-0663.95.1.66

Tobias, S., & Everson, H. T. (1997). Studying the relationship between affective and metacognitive variables. *Anxiety, Stress & Coping: An International Journal, 10*(1), 59-81. doi:10.1080/10615809708249295

Trochim, W., & Donnelly, J. (2006). The research knowledge methods base. Cincinatti, OH: Atomic Dog Publishing

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 47A*(3), 631-650. doi:10.1080/14640749408401131

Veenman, M. V. J., Kerseboom, L., & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: Availability versus production deficiencies. *Anxiety, Stress & Coping: An International Journal, 13*(4), 391-412. doi:10.1080/10615800008248343

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. *Psychonomic Bulletin & Review, 18*(2), 267-277.

Voyer, D., & Sullivan, A. M. (2003). The relation between spatial and mathematical abilities: Potential factors underlying suppression. *International Journal of Psychology, 38*(1), 11-23. doi:10.1080/00207590244000241

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of Sex-Differences in Spatial Abilities - a Metaanalysis and Consideration of Critical Variables. *Psychological Bulletin, 117*(2), 250-270. doi:10.1037/0033-2909.117.2.250

Voyer, D., Voyer, S., & Saint-Aubin, J. (2017). Sex differences in visual-spatial working memory: A meta-analysis. *Psychonomic Bulletin & Review, 24*(2), 307-334. doi:10.3758/s13423-016-1085-7 27357955

West, R. L., Boatwright, L. K., & Schleser, R. (1984). The link between memory performance, self-assessment, and affective status. *Exp Aging Res, 10*(4), 197-200. doi:10.1080/03610738408258464 6535732

Wraga, M., Duncan, L., Jacobs, E. C., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review, 13*(5), 813-819. doi:10.3758/BF03194002 17328378

CHAPTER 7: EXPERIMENT 3

EFFECTS OF STEREOTYPE ACTIVATION ON GENDER DIFFERENCES IN METACOGNITIVE ACCURACY AND NAVIGATION PERFORMANCE

Chantal L. Lemieux, Charles A. Collin

University of Ottawa

Nicholas N. Watier

Brandon University

Author Note

Chantal L. Lemieux, School of Psychology, University of Ottawa; Charles A. Collin, School of Psychology, University of Ottawa; Nicholas Watier, Brandon University.

This work was supported by a grant from the *Canadian Natural Sciences an*d *Engineering Research Council* to CC. We thank Stephane Rainville for creating the computer program used to present the virtual mazes.

Abstract

Many studies have shown a gender difference in spatial navigation ability, including a related gender difference in global metacognitive self-assessment and spatial anxiety. Recently, two studies by Lemieux and colleagues (2017a, 2017b) determined there are gender differences in the accuracy of trial-by-trial self-assessment, which they term *local* metacognition. In addition, many studies have shown the effects of stereotype threat/lift to influence confidence and performance between the genders in mental rotation but very few have studied this effect using a navigation task. Thus, we investigated whether actively manipulating confidence, related to stereotype threat/lift, may modulate local metacognitive accuracy and performance in a spatial navigation task. We assessed trial-by-trial metacognitive performance during a first-person virtual maze navigation task under three stereotype facilitation conditions where participants were told that either: 1) men perform better on this particular task than women, 2) women perform better on this particular task than men, or 3) the genders perform equally. Results showed that stereotype facilitation had an effect on navigation performance and confidence, but no effect on local metacognition. Contrary to our hypotheses, men only had higher metacognitive accuracy than women did in the Pro-Female condition. Women generally had poorer absolute metacognition prior to performance compared to men. These results demonstrate an influence of stereotype threat/lift on performance and confidence, but not on local metacognitive performance. These findings may help to better understand gender differences in navigation.

Keywords: stereotype threat, metacognition, gender differences, spatial navigation, virtual maze, confidence

Effects of Stereotype Activation on Gender Differences in Metacognitive Accuracy and Navigation Performance

There is a persistent gender gap in performance on spatial cognition tasks (for reviews, see Coluccia & Louse, 2004; Linn & Petersen, 1985; Martens & Antonenko, 2012; Voyer, Voyer, & Bryden, 1995). Some studies have suggested that this arises due to biological factors such as differences in genetics and hormones (e.g., Collaer & Hines, 1995; Halari et al., 2005; Hines et al., 2003; Kimura, 1999; Silverman, Choi, & Peters, 2007), and others sociocognitive ones such as differences in experience and gender stereotypes (Casey, 1996; Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005). Regardless of the origin of these differences, it is of interest to understand the mechanisms underlying them. For instance, many studies suggest that confidence differences may mediate the relationship between gender and spatial cognition performance (e.g., Estes & Felker, 2012).

These studies have shown that women generally underestimate their performance and have less confidence compared to men when performing a variety of tasks including spatial tasks (e.g., Cooke-Simpson & Voyer, 2007; Hertzog, Dixon, & Hultsch, 1990; Lundeberg, Fox, & Puncochar, 1994; Nori & Piccardi, 2015). Moreover, women have been shown to have more spatial anxiety compared to men (e.g., Lawton, 1994; Lawton & Kallai, 2002; Nowak, Murali, & Driscoll, 2015). However, most of this research uses self-reported scales designed to determine a participant's *overall* confidence or anxiety in specific navigational contexts, such as the Santa Barbara Sense of Direction Scale (Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002), and the Spatial Anxiety Scale (Lawton, 1994). These tools assess a form of *global metacognition* (e.g., Kelemen, Frost, & Weaver, 2000; Lieberman, 2004) because they ask participants to rate their overall abilities or comfort levels during a variety of common daily spatial tasks.

In contrast to global metacognition, very little research has examined gender differences in *local metacognition* during spatial navigation tasks. This is a more fine-grained form of metacognition that has to do with how well one predicts their future and assesses their past performance on individual trials of an ongoing task (e.g., Hertzog & Dunlosky, 2011; Kelemen et al., 2000; Lieberman, 2004). The accuracy with which one can predict and assess one's performance is important in determining how effective one will be at a task, because such assessments allow strategic allocation of time and cognitive resources. Gender differences in local metacognitive accuracy during a navigation task have been previously demonstrated, a finding compatible with the idea that they, at least partly, explain gender differences in spatial cognition (Lemieux, Collin, & Watier, 2017a, 2017b).

In 1990, Nelson and Narens developed a framework for the collection of local metacognitive knowledge. According to their framework, local metacognitive knowledge critically depends on *monitoring* and *control* processes. Monitoring consists of acquiring information on the current state of a cognitive process and evaluating this information according to some goal. For example, in a wayfinding task, monitoring would involve knowing that by continuing to follow the current route, it will take much longer than anticipated to reach the final destination. Control changes the state of a cognitive process based on the output of monitoring; for example, choosing an alternative route in order to better succeed at a task or goal. In other words, monitoring and control processes work together to adjust behavior and improve performance.

This model has been supported by several classroom-based studies that found an association between metacognitive accuracy and academic performance (e.g., Dunlosky & Ariel, 2011; Dunlosky & Rawson, 2012; Hacker, Bol, Horgan, & Rakow, 2000; Nietfeld, Cao, & Osborne, 2005; Thiede, Anderson, & Therriault, 2003). The general findings are that students are more effective at regulating their study time as monitoring accuracy improves, and their performance increases (e.g., Hacker et al., 2000). For example, participants spend less time studying items they previously judged as easy and focused more so on items they had previously judged as difficult (for a review, see Son & Metcalfe, 2000). However, recent findings have suggested that this traditional view of the relationship between metacognitive and cognitive performance is overly simple. For instance, Son and Metcalfe (2000) found that participants allocated more study time (i.e., control) to items they had previously judged as easy when study time was restricted and insufficient, showing that the nature of metacognitive control strategies may depend on task demands. More recently, it has been suggested that a bidirectional link between monitoring and control may exist, whereby a control process (spending extra time studying an item) can feed into monitoring ("that item was hard if it took so long, therefore it probably won't be remembered") (Koriat, 2006; Koriat, Ackerman, Adiv, Lockl, & Schneider, 2014).

The Nelson and Narens (1990) framework proposes the use of several different metacognitive monitoring judgments as measures of local metacognition (for a review, see Dunlosky, Mueller, & Thiede, 2016). Among these is the Ease of Learning Judgment (EOL), collected before study, which is a reflection of how difficult or easy the participant perceives the item will be to learn. Judgments of Learning (JOLs), collected after study but before performance, are a reflection of a participant's subjective sense of how well they learned the item. Retrospective Confidence Judgments (RCJs), collected after performance, are a reflection of participants' subjective sense of how well they performed. All judgments are collected on a trial-by-trial basis and represent the participants' ability to monitor their performance for that particular trial. Metacognitive accuracy can then be measured by comparing each trial's actual performance with the various monitoring judgments collected on that trial. Each monitoring judgment is thought to measure different metacognitive processes. For instance, JOLs and RCJs seem to rely on different sources of information (e.g., Busey, Tunnicliff, Loftus, & Loftus, 2000; Dougherty, Scheck, Nelson, & Narens, 2005), and are differentially affected by encoding manipulation (e.g., study time manipulations affects JOLs but not RCJs; Dougherty et al., 2005). In addition, each monitoring judgment is differentially affected by neurological impairments (e.g., Modirrousta & Fellows, 2008; Pannu & Kaszniak, 2005).

Given that men and women have different levels of anxiety and confidence on spatial tasks (e.g., Cooke-Simpson & Voyer, 2007; Hertzog et al., 1990; Lawton, 1994; Lawton & Kallai, 2002; Lundeberg et al., 1994; Nori & Piccardi, 2015; Nowak et al., 2015), it is reasonable to suggest that they would have different perceptions of self-efficacy on such tasks. Self-efficacy is defined by the belief or confidence in one's own abilities and has been shown to determine how individuals approach various goals, challenges and tasks (Bandura, 1997). A recent study found that gender did not predict navigation performance on three different tasks, whereas it was predicted by the participants' subjective feelings of competency on spatial tasks (Nori & Piccardi, 2015). Another study found that confidence mediated the gender difference in a spatial task across four experiments (Estes & Felker, 2012). Thus, differences in self-efficacy related to spatial tasks could influence the way each gender monitors and controls their performance during a spatial navigation task. This idea is supported by a recent review of metacognitive research

(Dunlosky et al., 2016) suggesting that measures of local metacognition are influenced by both epistemological beliefs (e.g., subjective feelings of competency) and perceptions of how easy an item was processed (e.g., Frank & Kuhlmann, 2016; Mueller & Dunlosky, 2017; Mueller, Dunlosky, Tauber, & Rhodes, 2014). Moreover, self-efficacy was shown to be a strong predictor of global metacognitive self-assessment (Coutinho & Neuman, 2008).

This possibility has some empirical support. Using the same materials and first-person virtual navigation task as the present study, Lemieux et al. (2017a, 2017b) showed that men outperform women across all performance measures and have greater levels of trial-by-trial confidence, especially prior to performance. More critically, they showed that the genders differ in their metacognitive accuracy. In their first study, Lemieux et al. (2017b) collected EOLs and RCJs and found that judgments made prior to performance (EOLs) were significantly correlated with performance for men but not women, indicating men to have a slightly better ability to predict their navigational performance. Additionally, men were significantly more accurate in assessing their past performance (RCJ) compared with women. In a second study, JOLs and RCJs were collected. Results corroborated the finding that men were more accurate in assessing their past performance (RCJ) compared to women (Lemieux et al., 2017a). They also found that women were generally underconfident prior to performance, while men showed no confidence bias prior to performance.

Women's lack of self-efficacy related to spatal tasks could be explained by the existence of a robust stereotype that women do poorly on navigation tasks (e.g., Harris, 1981; Hausmann, Schoofs, Rosenthal, & Jordan, 2009). Because subjective feelings of competency are influenced by beliefs in stereotypes, we decided to examine whether women's underconfidence prior to performance and poor metacognitive monitoring could be, at least partially, caused by *stereotype* *threat* (Steele & Aronson, 1995). Stereotype threat is produced when reminders of stereotyped inferiority cause an individual to feel self-conscious and anxious, leading to disrupted emotion regulation, working memory and executive function, which in turn prevent the individual from performing at their best (e.g., Beilock, Jellison, Rydell, McConnell, & Carr, 2006; Inzlicht, McKay, & Aronson, 2006; Johns, Inzlicht, & Schmader, 2008; Schmader & Johns, 2003; Schmader, Johns, & Forbes, 2008). The complement to stereotype threat is the concept of stereotype lift, which occurs when an individual's sense of self-efficacy is improved by downward comparisons made to a social group that is negatively stereotyped, causing a boost in performance (for a review, see Walton & Cohen, 2003). While several studies confirmed the effects of stereotype threat/lift in spatial tasks (e.g., Delgado & Prieto, 2008; Heil, Jansen, Quaiser-Pohl, & Neuburger, 2012; Massa, Mayer, & Bohon, 2005; McGlone & Aronson, 2006; Moe & Pazzaglia, 2006; Sharps, Welton, & Price, 1993; Wraga, Helt, Jacobs, & Sullivan, 2007), a recent meta-analysis on studies investigating stereotype threat/lift effect among spatial tasks concluded that there was no consistent effect of stereotype threat on women's performance, but there was evidence for stereotype lift (Doyle & Voyer, 2016). However, this meta-analysis mainly included studies using mental rotation tasks and no navigation or wayfinding tasks. Thus, it is difficult to know how generalizable these results are to wayfinding/navigation tasks.

Despite a large body of research on gender differences in spatial cognition, only two studies specifically looked at the effects of stereotype on women and men on a wayfinding task, both of which used first-person 3-dimensional virtual environments (Allison, Redhead, & Chan, 2017; H. E. S. Rosenthal, Norman, Smith, & McGregor, 2012). H. E. S. Rosenthal et al. (2012) measured the time it took participants, across three trials, to find a hidden goal using either geometric or landmark-based cues. They found that stereotype lift was able to affect men's performance in both preferential (geometric) and not-preferential (landmark) cue types. In contrast, stereotype threat did not significantly affect women's performance for either cue type. This last finding may be because the task was either too difficult or too easy, as Allison et al. (2017) found that stereotype threat only had an effect at an appropriate level of task difficulty. That is, when the task was especially hard for women, stereotype threat did not have an effect, but it did in a second experiment when participants were given orientation cues in order to make the task easier (Allison et al., 2017). Interestingly, men were affected by stereotype threat at both difficulty levels, likely because they had not yet reached a performance floor, as was likely the case for women in the harder task. However, they did not find any significant effects for stereotype lift across either gender and task difficulty.

Schmader et al. (2008) have proposed a process model to explain how stereotype threat affects performance. Their model suggests that during stereotype threat activation, individuals experience an increase in performance monitoring and stress arousal due to increased self-consciousness and anxiety; this in turn depletes working memory resources and thus limits performance. In the same way, they consider stereotype lift to alleviate stress, which allows working memory resources to fully support performance. In addition to performance monitoring and stress arousal, other variables related to self-efficacy have been linked to metacognition (e.g., Coutinho & Neuman, 2008; Kleitman & Stankov, 2007; McDonald-Miszczak, Gould, & Tychynski, 1999; Stankov; Tobias & Everson, 1997; Veenman, Kerseboom, & Imthorn, 2000; West, Boatwright, & Schleser, 1984). Thus, it is possible that stereotype threat has an effect on performance via the intermediary of metacognitive performance. That is, it may increase anxiety and reduce self-confidence, which in turn impairs metacognitive monitoring and control processes, which ultimately limits performance. The present study sought to investigate this hypothesis by determining if metacognitive monitoring accuracy and performance are affected by stereotype threat/lift.

The stereotype threat manipulation protocol used in the present study was inspired by Moe and Pazzaglia (2006). In the context of a mental rotation task, Moe and Pazzaglia (2006) tested the effects of stereotype threat/lift using three different deceptive instructions at the onset of testing. Participants were simply told either that men are better at the task, that women are better at the task, or there was no reference to gender. They found that women in the Pro Female condition and men in the Pro Male condition outperformed their counterparts, confirming the effect of stereotype. The present study sought to extend this research to gender differences in navigation and expand on the findings from H. E. S. Rosenthal et al. (2012) and Allison et al. (2017), as well as to investigate the role of local metacognition in the Schmader et al. (2008) stereotype threat model.

Questions & Hypotheses

The primary goal of this study was to apply a stereotype threat manipulation to male and female participants to see how this influences navigational and metacognitive performance. In order to determine the effects of stereotype threat, we had groups of men and women perform a virtual maze learning task under three different instructional Stereotype Facilitation conditions. These conditions included statements to subjects to the effect that: 1) men perform better at the task than women (Pro Male Facilitation condition), 2) women perform better at the task than men (Pro Female Facilitation condition), or 3) the genders tend to perform equally on the task (Neutral Facilitation condition).

The hypotheses relating to this goal are based on the stereotype threat research conducted on spatial tasks (Allison et al., 2017; Heil et al., 2012; Moe & Pazzaglia, 2006; H. E. S. Rosenthal et al., 2012). The women in the Pro Female Facilitation condition were expected to experience stereotype lift and therefore exhibit less anxiety and more confidence compared to both the Neutral Facilitation condition and the Pro Male Facilitation condition. This was in turn expected to improve task performance as well as metacognitive accuracy by increasing cognitive resources in women in this condition. An effect of stereotype threat was expected to reduce women's performance and metacognitive accuracy in the Pro Male Facilitation condition. The reverse pattern was expected for men in all corresponding conditions. Consequently, the effect of gender was expected to be at its smallest in the Pro Female Facilitation condition. The opposite was expected in the Pro Male Facilitation condition; with stereotype threat experienced among women and stereotype lift experienced among men such that the gender difference should be at its greatest in this condition. The gender difference for the Neutral Facilitation condition was expected to have effect sizes in between those from the corresponding Pro Female and Pro Male Facilitation conditions.

In addition to measures of local metacognition, two scales were used to evaluate global metacognition. This was done in order to explore other relevant factors that might be influenced by stereotype threat. The Santa Barbara Sense of Direction Scale (SBSOD; Hegarty et al., 2002) was used to evaluate overall navigation metacognitive knowledge. The Spatial Anxiety Scale (SAS; Lawton, 1994) was used to determine if spatial anxiety was influenced by stereotype threat. Our hypotheses for global metacognitive measures follow the same pattern as for all other variables; the largest difference between genders was expected to be seen in the Pro Male condition while the smallest was expected to be seen in the Pro Female condition.

The secondary goal of this study is to confirm the findings of our previous studies and to better understand the gender difference in metacognitive monitoring during a spatial navigation task. This will be done by comparing the genders across three different monitoring judgments: EOL, JOL, and RCJ. As the metacognitive literature implies, if a gender difference in metacognitive monitoring accuracy is found, this could imply differences in control processes, which could in turn help to explain differences in navigation performance (Hacker et al., 2000; Nietfeld et al., 2005; Thiede et al., 2003). Consistent with Lemieux et al. (2017a, 2017b), we expected women's metacognitive monitoring judgments to be less well correlated with their performance on a trial-by-trial basis.

A third goal of this study was to better understand gender differences in confidence and navigation performance. We hypothesized that men would outperform women on all performance measures across all three Stereotype Facilitation conditions, based on the results seen in Lemieux et al. (2017a, 2017b). We expected the gender difference to be at its greatest for Maze Completion Time, as this was a very consistent finding in Lemieux et al. (2017a, 2017b), who found that women took on average approximately double the time it took men to complete each maze. Similarly, consistent with Lemieux et al. (2017a, 2017b), we hypothesized that women would have lower trial-by-trial metacognitive confidence ratings than men and that this difference would be greatest for retrospective confidence.

Methods

Participants

A total of 311 participants were recruited via the departmental subject pool made up of first and second year undergraduates. Participants who did not identify on the traditional gender binary (i.e., either male or female) and instead indicated "other" when filling out demographic information were screened out of recruitment. Additionally, individuals who indicated a 7 or above on a 10-point scale where 10 was "I often suffer from motion sickness" and 1 was "I never suffer from motion sickness" were excluded from recruitment. This was done in order to avoid individuals susceptible to simulator/motion sickness, which is likely to occur in all 3-dimensional virtual environment, including the current virtual navigation task (e.g., Draper, Viirre, Furness, & Gawron, 2001; Lemieux et al., 2017a; Lemieux et al., 2017b; Stinchcombe, 2014). Sixty-four participants could not be included in the analyses: 44 women and 10 men had to terminate their participation due to simulator sickness, 1 woman and 2 men had to terminate their participation due to fatigue, and technical difficulties prevented 1 woman and 3 men from completing the experiment. Additionally, 1 woman and 2 men were excluded because they gave the same metacognitive judgment of confidence across all trials, despite being asked to use the entire range, which would have prevented us from calculating relative metacognitive accuracy. The final sample was 247 with numbers in Gender and Stereotype Facilitation conditions approximately balanced throughout testing (see Table 1 for number of participants in each group). Average age was 19.53 ± 2.69 years.

Sample size was initially determined by conducting an a priori power analysis using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) for an independent sample 2 X 3 (Gender X Facilitation condition) ANOVA with $\alpha = .05$, and a moderate to low effect size of $f = .20$ ($d = .40$). This effect size was chosen as a compromise between the large effect sizes (e.g., $d \leq .80$) seen in gender differences in spatial navigation performance using virtual Hebb-Williams mazes (e.g., Shore, Stanford, MacInnes, Klein, & Brown, 2001)[10] and the small effects (e.g., d = .29) found during gender stereotype threat manipulations for math and spatial tasks (for a meta-analysis see Doyle & Voyer, 2016). The outcome of the analysis revealed that a minimum sample size of 244 participants would be required to obtain a power of .80. Recruitment of subsequent participants was done in order to equalize the number of participants of each gender across all three Stereotype Facilitation conditions.

Materials

The materials were the same or similar as those used in Lemieux et al. (2017a, 2017b). This includes the maze navigation program, maze video tours, and metacognitive instructions.

Virtual Navigation Task

Thirty virtual mazes were created from 10 Hebb-Williams mazes (Hebb & Williams, 1946). The 10 chosen were shown to be the most difficult in past research (Shore et al., 2001; Therrien & Collin, 2010). In addition, 20 more were created by either transposing the maze left-to-right relative to the axis linking the entrance to the exit, or reversing the maze by making the entrance the exit and vice versa. The virtual maze program recorded an error whenever the participant crossed invisible "trip wires" that were strategically placed in dead ends (Shore et al., 2001). See Figure 1 for an example of field of view and maze environment textures, and Figure 2 for layouts of all 30 mazes including "trip wires".

---

[10] While a measure of effect size was not directly reported in this study, it could nonetheless be calculated from the reported main effect of gender collapsing across species by exploiting the fact that $F = t^2$, and $d_s = t\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ (Lakens, 2013).

Virtual video tours of the mazes were generated in addition to a virtual first person navigation task. The video tours had a first person perspective view; the same view as the navigation task. Each video tour moves through the entire maze environment, following either a right-hand or left-hand wall-follower algorithm. Thus, there were two videos per maze, one left-handed and one right-handed. Participants were shown only one randomly selected video per maze. Each video tour was approximately 45 seconds long.

The experiment was programmed, using Matlab and OpenGL, to include the instructions, automatic collection of metacognitive judgments, presentation of video tours, and maze navigation performances. The experiment was run on Windows 7 and shown in a 800x600 aspect ratio with on a 21-inch screen. A 45 degree field of view was used in all video tours and maze performances, as a low field-of-view has been shown to be ideal for avoiding simulator sickness (e.g., Allen, 2000; DiZio & Lackner, 1998; Lin, Duh, Parker, Abi-Rached, & Furness, 2002).

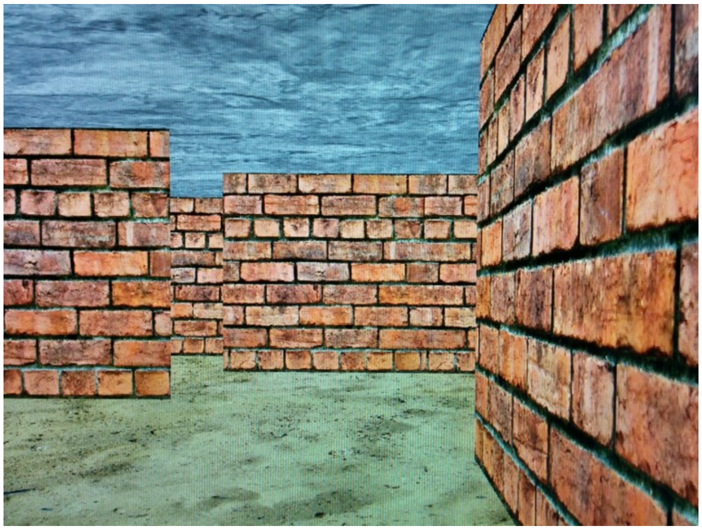
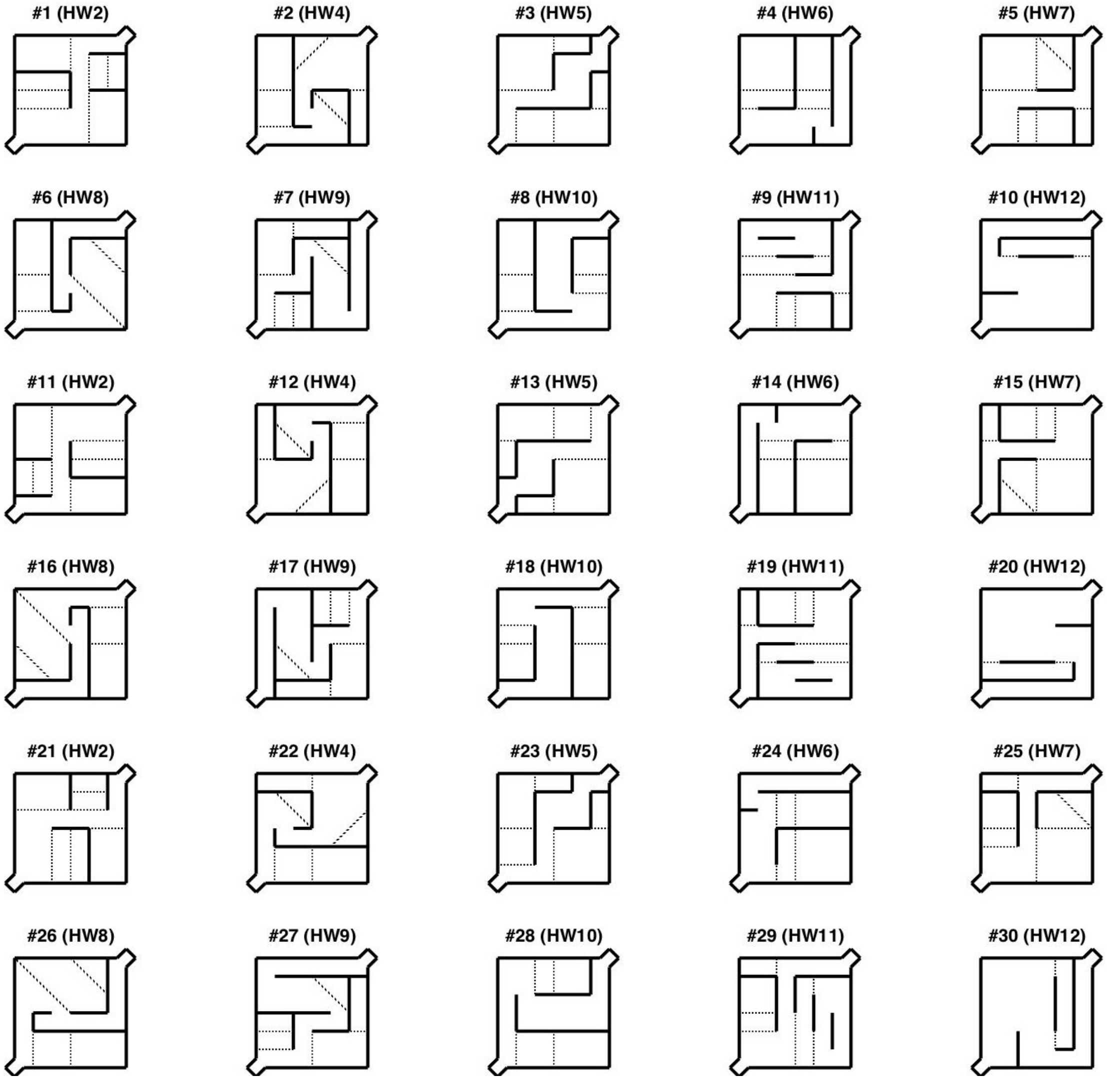


Figure 1. Screen shot demonstrating what participants saw for both the video tours and maze performances.



#1 (HW2) #2 (HW4) #3 (HW5) #4 (HW6) #5 (HW7)
#6 (HW8) #7 (HW9) #8 (HW10) #9 (HW11) #10 (HW12)
#11 (HW2) #12 (HW4) #13 (HW5) #14 (HW6) #15 (HW7)
#16 (HW8) #17 (HW9) #18 (HW10) #19 (HW11) #20 (HW12)
#21 (HW2) #22 (HW4) #23 (HW5) #24 (HW6) #25 (HW7)
#26 (HW8) #27 (HW9) #28 (HW10) #29 (HW11) #30 (HW12)

Figure 2. Layouts of the 30 mazes used in this experiment, as well as in Lemieux et al. (2017a, 2017b). Mazes 1 to 10 are reproductions of the original 12 Hebb-Williams mazes (Hebb & Williams, 1946), excluding #1 and #3, which were deemed too easy to perform. Mazes 11 to 20 are reversed versions of mazes 1 to 10, where the entrance has been made into the exit, and vice versa. Mazes 21 to 30 are transposed versions of mazes 1-10, where the x and y coordinates of the walls have been exchanged. Dotted lines show virtual “trip wires” replicated from Shore et al. (2001).

Local Metacognitive Materials

For the present study, we used the same instructions for EOLs as seen in Lemieux et al. (2017b), which were adapted from Nelson and Narens' (1990) framework. The adapted EOL judgment instructions are as follows:

> *"Clearly indicate on a scale of 0-100, going from "difficult" to "easy" with five options "0-25-50-75-100", how easy you believe it would be to learn the maze's layout such that you could navigate from the entrance to the exit without making any errors."*

We also used modified Nelson and Narens' (1990) prospective Judgment of Learning (JOL) instructions, as seen in Lemieux et al. (2017a), and retained its probabilistic scale:

> *"Clearly indicate on a scale of 0-100 going from "not confident" to "confident" with five options "0-25-50-75-100", how confident you are that you learned the maze's layout such that you could navigate from the entrance to the exit without making any errors."*

Similarly, we retained the probabilistic scale of the Retrospective Confidence Judgment (RCJ) but used the same modified instructions as seen in Lemieux et al. (2017a, 2017b):

> *"Clearly indicate on a scale of 0-100, going from "not confident" to "confident" with five options "0-25-50-75-100", how confident you are that you successfully navigated the maze without making any errors."*

Additional relevant instructions for all were:

> *"Note that an error is defined as not taking the shortest route between start and finish (e.g., going down a dead end, or taking a wrong turn). Please answer as honestly as possible and use the entire scale."*

Global Metacognitive Materials

Two questionnaires were used in order to measure aspects of global metacognition. The Santa Barbara Sense of Direction Scale (SBSOD; Hegarty et al., 2002) is a self-report measure of spatial ability. That is, it measures participants' judgments about their own spatial abilities. It is made up of 15 items and uses a 7-point rating scale ranging from "Strongly Agree" to

"Strongly Disagree", where the higher the score the better the individual's perceived navigation abilities. Additionally, the Spatial Anxiety Scale (SAS; Lawton, 1994) measures perceived anxiety during wayfinding with 8 items on a 5-point rating scale ranging from "Not at all anxious" to "Very anxious", where the higher the score the more spatially anxious the individual perceives themselves to be.

Stereotype Facilitation Condition Instructions

There were three possible Stereotype Facilitation condition groups where different verbal deceptive introductions were given: Pro Male, Pro Female, and Neutral. The deceptive introduction was as follows with specific differences for each facilitation condition in *italics*:

> "This study concerns the metacognitive aspects of spatial navigation. That is, we're interested in knowing how aware you are of your own performance when navigating. Metacognition will be measured with a series of questions, which we will go over soon. Our task is a virtual 3D spatial navigation task that is widely used in the literature. Previous studies have consistently shown that in this version of the task: *1) men outperform women, 2) women outperform men, 2) both genders perform equally*."

Once the corresponding Stereotype Facilitation introduction was given, general task instructions were then specified and explained. Participants were able to ask questions at the end of the verbal instructions. Note that task instructions and protocol were the same across all three Stereotype Facilitation conditions with the only difference being the above-described introduction.

Procedure

All testing sessions were done at the INSPIRE shared departmental laboratory at the School of Psychology, University of Ottawa. Between two and four participants were tested alongside each other. The groups of participants were in one room, but each had their own carrel and computer. The groups were always of mixed gender, never exclusively male nor exclusively female. Verbal instructions, always given by the same female experimenter, were provided prior to testing, including the deception introduction corresponding to the appropriate Stereotype Facilitation condition. That is, all participants in the room were under the same Facilitation condition.

Once instructions were given, the task was deployed on each participant's computer and the experimenter moved to the adjacent control room. From there she was able to observe the participants via camera feeds, as well as communicate with them via headphones if needed. There were two parts to this task. Part 1 consisted of viewing a series of 30 randomly ordered video tours of mazes, each followed by an EOL judgment. Part 2 consisted of a series of 30 paired video tours and maze performances, where the video tours were followed by a JOL judgment and performances were followed by an RCJ judgment. Part 1 did not require learning of the video tours, but only an assessment of their difficulty, as defined by the EOL, whereas participants were instructed in Part 2 to learn the layout of the maze in the video tour in order to navigate through it immediately after.

Part 1 began with a review of the overall instructions, the instructions specific to Part 1, and a practice trial involving a video tour and EOL judgment. Once experimental trials began, participants viewed a video tour and then gave their EOL judgment using the arrow keys to select the appropriate level on the scale (0-25-50-75-100) and space bar to confirm selection. EOL judgment instructions were displayed on screen, and participants were instructed to read the judgment instructions, throughout the experiment, before giving their response in order to not confuse one judgment type for another (e.g., EOL vs JOL). This was repeated until all 30 video tours were shown and all 30 EOLs collected. Presentation of maze video tours was randomized for each participant. Participants were allowed breaks between any two trials as desired, and were told to take a 2-minute break between Part 1 and Part 2.

Part 2 began with a review of the instructions specific to Part 2 and a practice trial consisting of a video tour, JOL judgment, maze performance, and RCJ judgment. Mazes were randomized relative to Part 1 for each participant. On each experimental trial, participants viewed a video tour and tried to learn the maze layout as best as they could in preparation for the upcoming navigation performance. Once the video tour was over, they gave their JOL and moved on to the performance of that same maze. Once they reached the exit of the maze, they gave their RCJ. This was repeated until all 30 mazes were done and all 30 JOLs and RCJs were collected. At this point, the last window told them to fill out the printed questionnaire booklet on their desk in the order that it appears using the pen provided and, when finished, to come out to the waiting room for a debriefing. The questionnaire booklet included the SBSOD, the SAS, and the debriefing form explaining to them the deception used and the true purpose of the study. The experimenter waited outside the room to debrief exiting participants.

To summarize, Part 1 consisted of 30 video tours each followed by an EOL judgment, Part 2 consisted of 30 trials, each consisting of a video tour, a JOL judgment, a maze performance, and an RCJ judgment. Global metacognitive measures were collected at the end.

Results

Local metacognition is typically evaluated using relative and absolute accuracy between monitoring judgment scores and performance. Relative accuracy is typically assessed by the Goodman-Kruskal gamma correlation (Nelson, 1984), which provides a statistical index of concordance between two ordinal variables. Performance is traditionally a dichotomous variable (in our case *Mazes Passed*: pass or fail) and confidence an ordinal variable (confidence score 0-25-50-75-100) for each trial. However, as gamma is bounded between -1 to +1 it inherently violates the assumptions required by parametric hypothesis tests. One solution is to convert gamma into a standardized measure that is normally distributed and unbounded called G* (Benjamin & Diaz, 2008)[11]. For G*, as with gamma, the higher the score the more accurate the participant's metacognitive accuracy. As each monitoring judgment was collected for each trail, pearson's bivariate correlations were computed across all three judgments in order to assess variable independence. The correlation between EOL G* and RCJ G* was not significant. The correlation between EOL G* and JOL G* was -.184 ($p<.05$), and the correlation between JOL and RCJ was .226 ($p<.001$). Despite these correlations reaching statistical significance, they were nonetheless small and likely inconsequential to the interpretation of the below findings.

Absolute metacognitive accuracy typically compares the probability of predicted performance to the probability of observed performance to determine overall over or under confidence. Thus, absolute metacognitive accuracy was analyzed by two one-tailed paired t-tests comparing observed performance (Percent Mazes Passed) with predicted performance (mean JOL or RCJ) for each gender in each condition.

Along with local metacognitive accuracy measures, mean EOL, JOL, and RCJ ratings were also computed as general indicators of confidence. This is in addition to performance variables, which were recorded by the maze navigation program.

---

[11] The equation used to transform gamma correlations into G* was : $G^{*}=\log\left(\frac{\gamma+1}{1-\gamma}\right)$ (Benjamin & Diaz, 2008)

In accordance with the structure of our hypotheses, the analyses consisted of planned simple contrasts using the omnibus error terms of a series of 2x3 factorial ANOVAs (R. Rosenthal & Rosnow, 1985), with Gender and Facilitation condition as independent variables. Separate ANOVAs were run for each DV, including performance DVs (Maze Completion Time, Error Count, Mazes Passed), metacognitive confidence DVs (mean EOL, JOL and RCJ), average relative metacognitive accuracy DVs (i.e., G*) for EOLs, JOLs and RCJs, and global metacognitive scales (SBSOD and SAS). For each DV, three planned simple contrasts comparing across gender within each facilitation condition were performed. As group sample sizes differ only slightly, the harmonic mean between sample sizes across all 6 groups was used in all contrast analyses (R. Rosenthal & Rosnow, 1985, p.17). Considering a small number of planned contrasts are being conducted relative to the total number of possible comparisons, the alpha level was not adjusted and remains at .05 (R. Rosenthal & Rosnow, 1985, p. 45). Effect sizes for contrasts were measured as Cohen's *d*.

Univariate outlier analyses was conducted on all dependent variables. An outlier was defined as having a z score above 2.46 in accordance with our sample size of approximately 40 individuals per condition (Van Selst & Jolicoeur, 1994). In order to maintain sample size, outliers were winsorized (Hogg, 1979; Huber, 2011). In other words, their scores were changed ("tucked") to the closest non-outlying score. Overall, 0.02% of scores were considered univariate outliers and were winsorized. Normality was defined as skewness and kurtosis scores near 0 within a ±2 criterion range (Field, 2009; Gravetter & Wallnau, 2016; Trochim & Donnelly, 2006). Normality was confirmed across all conditions for all variables except JOL G* which had a kurtosis of 2.52 among the men in the Neutral Facilitation condition and 3.06 among the men in the Pro Male condition. Nevertheless, considering ANOVAs are relatively robust against nonnormality, parametric analyses were also performed on this variable in order to maintain consistency in the analyses across all variables. Levene's test of homogeneity of variance was significant for a few variables. Nevertheless, to provide consistency, we opted to proceed with parametric analyses with careful consideration to differences in variance between samples (see Table 1 for descriptive statistics for all DVs across all conditions).

Navigation Performance

We hypothesized that the gender difference would be greatest in the Pro Male condition where men would take less time to complete the mazes, have fewer errors, as well as fewer failed mazes. Conversely, in the Pro Female condition, we expected women's performance to be boosted and for gender differences to be reduced resulting in the smallest effects sizes. The Neutral Facilitation condition was expected to have effect sizes in between those seen in the Pro Male and Pro Female conditions.

The contrasts showed men to significantly outperform women on all three measures across all three facilitation conditions, with large effect sizes (see Table 1 for descriptive statistics and Table 2 for contrast statistics). In other words, men had significantly lower Error Counts than women in the Pro Male condition ($F = 36.37$, $df = 1, 241$, $p < .001$, $d = 1.199$), Pro Female condition ($F = 16.88$, $df = 1, 241$, $p = .001$, $d = .901$), and Neutral Facilitation condition ($F = 11.58$, $df = 1,241$, $p < .001$, $d = .946$). Men also had shorter Maze Completion Times than women in the Pro Male condition ($F = 38.16$, $df = 1, 241$, $p < .001$, $d = 1.360$), Pro Female condition ($F = 40.48$, $df = 1, 241$, $p < .001$, $d = 1.394$), and Neutral Facilitation condition ($F = 52.67$, $df = 1,241$, $p < .001$, $d = 1.725$). Additionally, men also had a greater number of Mazes Passed than did women in the Pro Male condition ($F = 26.11$, $df = 1, 241$, $p < .001$, $d = 1.142$), Pro Female condition ($F = 10.61$, $df = 1, 241$, $p = .001$, $d = .726$), and Neutral Facilitation condition (*F* = 14.96, *df* = 1,241, *p*< .001, *d* = .868). By comparing effect sizes, one can see that the effect for gender is greatest in the Pro Male condition for Error Counts (*d* = 1.199) and Mazes Passed (*d* = 1.142) compared to the other conditions which have effect sizes ranging from .726 to .946, which corroborate our hypotheses. Contrary to our hypotheses, the effect size of Maze Completion Time is greatest in the Neutral Facilitation condition (*d* = 1.725), though the effect sizes in both the Pro Male and Pro Female conditions are similar and still quite large (1.360 and 1.394, respectively). It is important to note that women have significantly greater variability in their performance variables compared to men, with women's standard deviations sometimes doubling what is seen in men (see Table 1). This is a violation of the assumption of homogeneity of variance and could decrease statistical power by inflating the omnibus error term for these contrasts.

Metacognitive Confidence

We hypothesized the same stereotype facilitation trend for all metacognitive confidence analyses. The contrasts generally show men to have significantly more confidence than women on all three confidence measures across all three facilitation conditions, with mostly strong effect sizes (see Table 1 for descriptive statistics and Table 2 for contrast statistics). Women had lower mean EOL confidence scores compared to men in the Pro Male Facilitation condition (*F* = 11.37, *df* = 1, 241, *p* = .001, *d* = .796), and Neutral Facilitation condition (*F* = 8.58, *df* = 1,241, *p* = .004, *d* = .666), but there was no difference in the Pro Female Facilitation condition (*F* = .66, *df* = 1, 241, *p* = .419, *d* = .164). Additionally, women had significantly lower mean JOL confidence scores than men in the Pro Male Facilitation condition (*F* = 20.43, *df* = 1, 241, *p*< .001, *d* = .959), Pro Female Facilitation condition (*F* = 17.17, *df* = 1, 241, *p*< .001, *d* = .884), and Neutral Facilitation condition (*F* = 13.12, *df* = 1,241, *p*< .001, *d* = .878). Women also exhibited lower mean RCJ confidence scores in the Pro Male Facilitation condition ($F = 19.34$, $df = 1, 241$, $p < .001$, $d = 1.045$), Pro Female Facilitation condition ($F = 20.09$, $df = 1, 241$, $p < .001$, $d = .945$), and Neutral Facilitation condition ($F = 19.86$, $df = 1{,}241$, $p < .001$, $d = 1.009$). By comparing effect sizes, one can again see that the effect for gender is consistently greatest in the Pro Male condition across all three confidence scores. Additionally, as hypothesized, the effect sizes across all confidence DVs in the Pro Female condition were the smallest.

Relative Metacognitive Accuracy

We hypothesized the same stereotype facilitation trend for all relative metacognitive accuracy analyses. Contrary to our hypotheses, the contrasts generally show few significant differences between the genders on any of the three relative metacognitive accuracy parameters (see Table 1 for descriptive statistics and Table 2 for contrast statistics). The only significant differences were seen in the Pro Female condition. Paradoxically, women were significantly less accurate in their retrospective metacognitive accuracy than men in the Pro Female condition ($F = 4.62$, $df = 1{,}241$, $p = .033$) with a moderate effect size of .474. A trend towards a similar gender difference was seen for EOL G* ($F = 2.74$, $df = 1{,}241$, $p = .099$) with a slightly lower moderate effect size of .433.

Table 1.

Descriptive Statistics for all DVs across Stereotype Facilitation Conditions

	Women						Men					
	Pro Male (n = 43)		Pro Female (n = 46)		Gender Neutral (n = 43)		Pro Male (n = 39)		Pro Female (n = 40)		Gender Neutral (n = 36)	
DV	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
Maze Completion Time (s)	17.97	6.45	18.71	6.10	19.74	5.61	11.13	2.99	11.67	3.71	11.72	3.45
Error Count	22.49	14.46	20.02	12.71	18.40	8.42	9.28	5.79	11.03	6.15	10.94	7.30
Mazes Passed	17.14	5.89	18.09	6.07	17.79	5.31	22.85	3.90	21.73	3.65	22.11	4.62
% Mazes Passed	57.13	19.65	60.29	20.24	59.30	17.69	76.15	13.01	72.42	12.17	73.70	15.41
EOL mean	57.23	12.80	59.80	14.38	58.82	13.51	68.21	14.71	62.44	17.69	68.36	15.08
JOL mean	64.59	16.77	63.08	15.81	64.57	12.48	79.59	14.42	76.83	15.29	76.60	14.81
RCJ mean	77.13	11.64	76.34	15.12	75.12	11.90	88.14	9.30	87.56	7.29	86.27	10.15
EOL gamma	0.26	0.31	0.10	0.31	0.17	0.21	0.15	0.45	0.23	0.29	0.06	0.31
EOL G*	0.25	0.31	0.10	0.30	0.16	0.20	0.19	0.58	0.22	0.29	0.07	0.30
JOL gamma	0.23	0.38	0.21	0.40	0.29	0.29	0.18	0.53	0.20	0.36	0.31	0.41
JOL G*	0.25	0.41	0.24	0.45	0.29	0.29	0.22	0.61	0.20	0.36	0.42	0.66
RCJ gamma	0.73	0.21	0.66	0.19	0.74	0.19	0.66	0.24	0.75	0.18	0.70	0.22
RCJ G*	0.96	0.56	0.82	0.48	0.99	0.53	0.90	0.64	1.10	0.67	0.95	0.60
SBSOD	4.05	1.08	4.17	1.22	4.26	1.12	4.65	1.00	5.11	0.87	4.80	0.98
SAS	22.51	5.50	22.17	6.44	23.40	5.58	21.36	6.71	19.15	5.92	20.25	6.05

Table 2.
All Planned Contrasts Comparing Gender across all Stereotype Facilitation Conditions

DV	Simple Contrasts (Men vs Women)	F	p	d
Error Count	Pro Male Facilitation	36.37	0.001	1.199
	Pro Female Facilitation	16.88	0.001	0.901
	Gender Neutral Condition	11.58	0.001	0.946
Maze Completion Time	Pro Male Facilitation	38.16	0.001	1.360
	Pro Female Facilitation	40.48	0.001	1.394
	Gender Neutral Condition	52.67	0.001	1.725
Mazes Passed	Pro Male Facilitation	26.11	0.001	1.142
	Pro Female Facilitation	10.61	0.001	0.726
	Gender Neutral Condition	14.96	0.001	0.868
EOL Confidence	Pro Male Facilitation	11.37	0.001	0.796
	Pro Female Facilitation	0.66	0.419	0.164
	Gender Neutral Condition	8.58	0.004	0.666
JOL Confidence	Pro Male Facilitation	20.43	0.001	0.959
	Pro Female Facilitation	17.17	0.001	0.884
	Gender Neutral Condition	13.12	0.001	0.878
RCJ Confidence	Pro Male Facilitation	19.34	0.001	1.045
	Pro Female Facilitation	20.09	0.001	0.945
	Gender Neutral Condition	19.86	0.001	1.009
EOL G*	Pro Male Facilitation	0.66	0.417	0.134
	Pro Female Facilitation	2.74	0.099	0.433
	Gender Neutral Condition	1.37	0.242	0.349
JOL G*	Pro Male Facilitation	0.05	0.819	0.046
	Pro Female Facilitation	0.14	0.705	0.097
	Gender Neutral Condition	1.59	0.208	0.260
RCJ G*	Pro Male Facilitation	0.21	0.651	0.096
	Pro Female Facilitation	4.62	0.033	0.474
	Gender Neutral Condition	0.09	0.762	0.069
SAS	Pro Male Facilitation	0.74	0.389	0.188
	Pro Female Facilitation	5.12	0.025	0.489
	Gender Neutral Condition	5.54	0.019	0.541
SBSOD	Pro Male Facilitation	6.56	0.011	0.574
	Pro Female Facilitation	15.94	0.001	0.882
	Gender Neutral Condition	5.44	0.021	0.519

Note: df were 1, 241 for all contrasts as they were all from the same 2x3 ANOVAs. All contrasts are one-tailed.

Global Metacognition

In reference to the SBSOD scale, we hypothesized the same stereotype facilitation trend, and the contrasts demonstrate men to have significantly higher SBSOD scores than women across all three conditions (see Table 1 for descriptive statistics and Table 2 for contrast statistics): Pro Male Facilitation ($F = 6.56$, $df = 1, 241$, $p = .011$, $d = .574$), Pro Female Facilitation ($F = 15.94$, $df = 1, 241$, $p<.001$, $d = .882$), and Neutral ($F = 5.44$, $df = 1, 241$, $p = .021$, $d = .519$). Contrary to our hypotheses, the effect size is largest for the Female Facilitation condition ($d = .882$) while the other two are equally moderate ($d = .574$ and $d = .519$).

In reference to the SAS scale, we hypothesized the same trend, and again contrary to our hypotheses, the contrasts show men to have significantly lower spatial anxiety in the Pro Female condition ($F = 5.12$, $df = 1, 241$, $p = .025$, $d = .489$) and Neutral Facilitation condition ($F = 5.54$, $df = 1, 241$, $p = .019$, $d = .541$) compared to women but no difference was found in the Pro Male condition ($F = .74$, $df = 1, 241$, $p = .389$, $d = .188$).

Absolute Metacognitive Accuracy

One-tailed paired sample t-tests were conducted comparing observed performance (% Mazes Passed) with predicted performance (JOL mean) and retrospectively rated performance (RCJ mean) for each gender across all three Stereotype Facilitation conditions (see Figure 3). For the sake of parsimony, only the statistically significant t-tests will be reported. The difference between prospective predicted performance (JOL mean) and observed performance (% Mazes Passed) was only statistically significant for women among the Pro Male ($t = 2.006$, $df = 42$, $p = .026$. $d = .379$) and Neutral Facilitation conditions ($t = 1.722$, $df = 42$, $p = .046$, $d = .298$). In both cases, women were overconfident as their prospective predicted performance means ($\bar{x} = 64.59$, $s = 16.77$, and $\bar{x} = 64.57$, $s = 12.48$, respectively) were higher than their observed performance means ($\bar{x} = 57.13$, $s = 19.65$, and $\bar{x} = 59.30$, $s = 17.69$, respectively). The differences between retrospective predicted performance (RCJ mean) and observed performance were significant across all six conditions (Women Pro Male: $t = 6.705$, $df = 42$, $p < .001$, $d = 1.018$; Women Pro Female: $t = 7.577$, $df = 45$, $p < .001$, $d = .793$; Women Neutral: $t = 5.537$, $df = 42$, $p < .001$, $d = .894$; Men Pro Male: $t = 6.285$, $df = 38$, $p < .001$, $d = .922$; Men Pro Female: $t = 8.969$, $df = 39$, $p < .001$, $d = 1.244$; Men Neutral: $t = 5.399$, $df = 35$, $p < .001$, $d = .816$). Both genders were significantly overconfident when assessing past performance, with large effect sizes, as their observed performance means (% Mazes Passed) were significantly lower than their predicted performance means (RCJ mean; see Figure 3). Interestingly, the largest effect sizes were among Women in the Pro Male condition ($d = 1.018$) and among Men in the Pro Female condition ($d = 1.244$). This indicates that each gender was most overconfident when assessing past performance in the opposite Stereotype Threat Facilitation condition than expected.

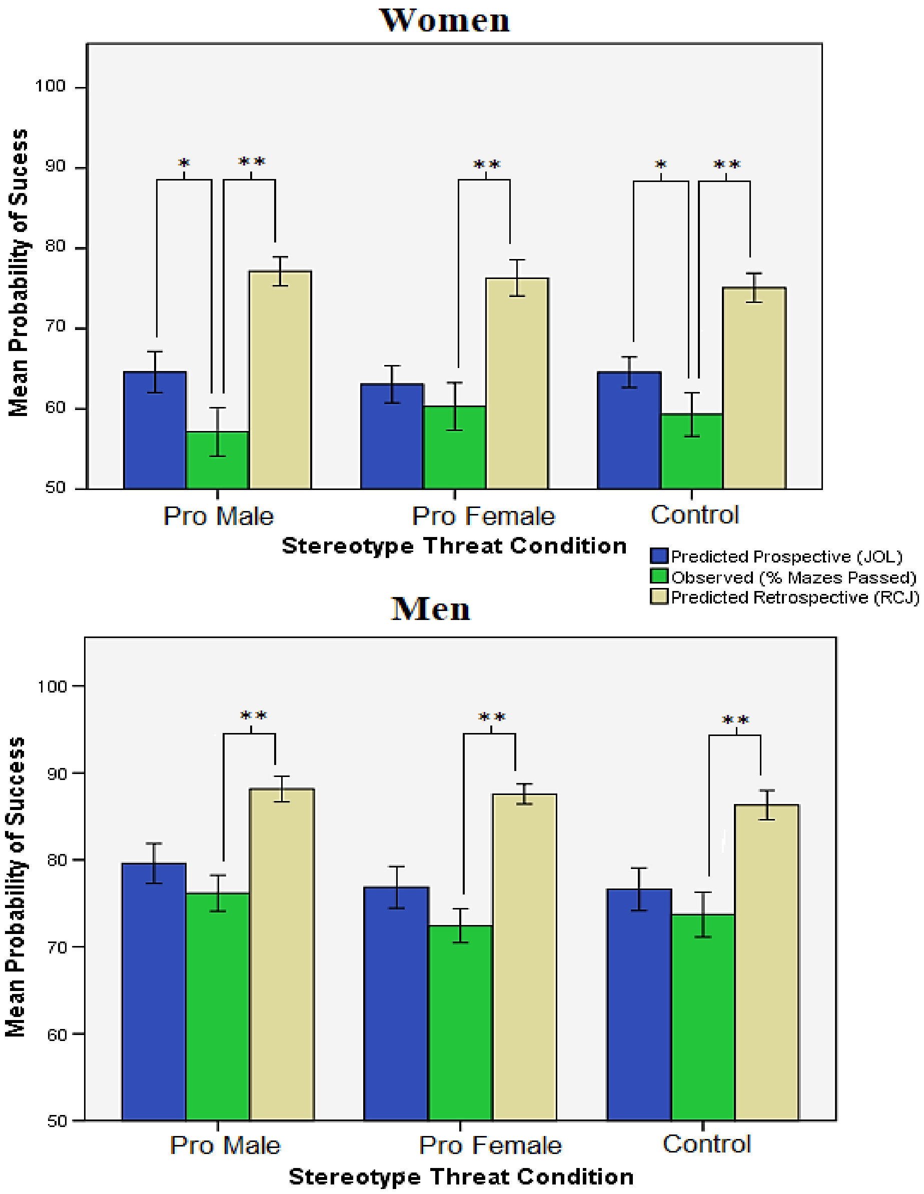


Figure 3. Bar graphs comparing observed performance (% Mazes Passed) with predicted (JOL mean) and retrospectively assessed performance (RCJ mean) for each gender across all three Stereotype Threat conditions. **p*<.05 ***p*<.001. One SE is used for error bars.

Discussion

The results of the current study are consistent with previous research on gender differences in spatial navigation performance (for a review, see Coluccia & Louse, 2004). With large effect sizes, women demonstrated poorer virtual navigation performance compared to men across all three Stereotype Facilitation conditions. More specifically, women took, on average, almost twice as long to complete the mazes compared to men, made almost twice as many navigational errors in comparison to men, and overall failed a greater number of mazes than did men. Additionally, our hypotheses on the influence of the Stereotype Facilitation conditions on navigation performance are supported, in that the gender difference in most performance variables showed the greatest effect size in the Pro Male condition and the smallest effect size in the Pro Female condition. That is, the gender difference was greatest when participants were told that men outperform women on the task. Similarly, the gender difference was reduced when participants are told that women outperform men on the task. To clarify, men still outperformed women across all conditions, however both genders' performance was influenced by stereotype threat/lift.

It is important to note that women had greater variability in their performance compared to men, with women's standard deviations sometimes doubling what was seen in men. This is contrary to what is typically seen with regards to gender diffrences, as men are known for having greater variability in performance data for most cognitive tasks, including spatial tasks (for a review, see Hedges & Nowell, 1995). One possibility is that some women in our study may have been better able to shift from a landmark-based navigation strategy, which women typical use as a default, to an orientation-based strategy, allowing them to perform better. Indeed, several studies have shown that while men are easily capable of changing from their default orientation-based navigation strategy to a landmark-based strategy, women have been shown to either take longer to shift strategies or are only capable of doing so with training (e.g., Astur, Purton, Zaniewski, Cimadevilla, & Markus, 2016; Livingstone-Lee, Zeman, Gillingham, & Skelton, 2014). It is possible a subset of women were better or sooner able to shift to a more apropriate navigation strategy creating large variations in performance means across women. Indeed, the current task was specifically designed to favor an orientation-based strategy in that there are no landmarks, on which the female-typical route-based navigation strategy depends (for a review, see Coluccia & Louse, 2004; Saucier et al., 2002). It is also important to note that the gender differences in performance on our task may have been augmented due to the fact that our task favor's a male-typical strategy. Thus, these performance differences may not generalize to a landmark-based navigation task nor to day-to-day navigation which often includes appropriate information for either strategy.

The results of the current study are also consistent with previous research on gender differences in confidence on spatial tasks. Women generally exhibited lower confidence scores compared to men with large effects for all three confidence variables (EOL, JOL, and RCJ mean) across two of the three Stereotype Facilitation conditions. This demonstrates a robust effect of gender on confidence in navigation tasks, similar to what has previously been found (e.g., Lawton, Charleston, & Zieles, 1996; Lemieux et al., 2017a, 2017b; O'Laughlin & Brubaker, 1998; Picucci, Caffo, & Bosco, 2011; Stevens & Carlson, 2016).

The only case in which there was no difference in confidence between men and women was in mean prospective EOL confidence scores in the Pro Female condition. This is compatible with our hypotheses related to the influence of Stereotype Threat. That is, the effect size of the gender difference was expected to be greatest in the Pro Male condition and smallest in the Pro

Female condition. This hypothesized trend was generally observed across all three confidence variables in that the effect sizes were greatest in the Pro Male condition and smallest in the Pro-Female condition. In the case of mean EOL scores, the gender effect was small enough to not be statistically significant in the Pro Female condition. As with the performance variables, we can conclude that metacognitive confidence among the genders can be influenced by stereotype threat/lift.

We hypothesized, along the same lines as with the other variables, that the gender difference in metacognitive accuracy would be greatest in the Pro Male condition, whereas in the Pro Female condition, we expected women's accuracy to be boosted and for gender differences to be reduced. However, contrary to our hypotheses, the contrasts generally show few significant differences between the genders on relative accuracy scores for any of the three metacognitive judgments. The only significant differences were seen in the Pro Female condition, where women were significantly less accurate in their retrospective metacognitive accuracy and near significantly less accurate in their ease-of-learning prospective metacognitive accuracy than men, both with moderate effect sizes. Thus, despite finding effects of stereotype threat/lift on performance and confidence, we found no consistent evidence of such effects on relative metacognitive accuracy.

The means for RCJ G* across all conditions are consistently high, ranging from 0.82 to 1.10. These correspond to mean gamma correlations ranging from .66 to .75[12], indicating both men and women have good retrospective metacognitive accuracy regardless of Stereotype Facilitation condition. This is in line with what is seen in the literature on the retrospective confidence-accuracy relationship for general knowledge questions, where the magnitude of the

---

[12] See Table 1 for gamma correlation descriptive statistics by condition.

within-subject correlations is typically moderate-to-high (Koriat, 2007). However, the means for JOL G* across all conditions are marginally low, ranging from .20 to .42 (mean gamma correlations ranging from .18 to .31) compared to the metamemory literature using word-pair associates which averages around .20 to .54 in their gamma correlations (Koriat, 2007). The means for EOL G* across all conditions range from .07 to .22 (mean gamma correlations range from .06 to .26), which is comparable to the metamemory literature using word-pair associates, which average approximately .20 in their gamma correlations (Dunlosky & Matvey, 2001; Leonesio & Nelson, 1990).

The absolute metacognitive accuracy analysis determined that women were overconfident prior to performance in the Pro Male and Neutral Facilitation conditions with small effect sizes, while men were neither over nor under confident prior to performance in any of the Stereotype Facilitation conditions. This is in contrast to Lemieux et al. (2017a, 2017b) who found that women were generally *under*confident prior to performance. Both genders were significantly overconfident when assessing past performance, with large effect sizes. Interestingly, the largest effect size among women was in the Pro Male condition and among men, it was in the Pro Female condition. This demonstrates that each gender was most overconfident when assessing past performance in the opposite Stereotype Threat Facilitation condition than expected. That is, absolute retrospective metacognitive accuracy seems to have been impaired by stereotype threat, in the form of an exaggeration of over-confidence, in both genders. This overconfidence when assessing past performance has been seen before in a series of similar studies by Lemieux et al. (2017a, 2017b), as well as in the broader metacognitive literature as the “overconfidence effect” (Koriat, 2007).

Regarding our global metacognitive measures, the SBSOD and SAS, men rated themselves significantly higher in spatial ability and lower in spatial anxiety than did women across all three Stereotype Facilitation conditions. Contrary to what we were expecting, the largest effect sizes were for the Pro Female condition which does not support the present hypothesized stereotype facilitation trend. This could be because the measures were administered at the end of testing when stereotype facilitation may have lost salience, and/or because these global measures are be more stable. It is also possible that the Stereotype Facilitation conditions set the participants up with an expectation regarding how they would perform that was then violated by their actual performance. That is, women in the Pro Female condition may have had high expectations for their performance, and if their actual performance did not live up to their expectations they may have rated their navigation abilities and spatial anxiety as worse than they otherwise would have. The opposite is true of men, creating a larger gender gap in the Pro Female condition. However, this explanation is made more complicated when taking into consideration the absolute metacognitive analysis as both genders were overconfident after each performance regardless of Stereotype Facilitation condition.

One limitation of this research is a potential sampling bias due to the fact that one of the requirements for recruitment was answering a question about motion sickness susceptibility. Previous studies have shown that there are a greater number of women who experience simulator/motion sickness (Classen, Bewernitz, & Shechtman, 2011; Lemieux et al., 2017a, 2017b) including the current experiment which had four times more women who could not complete the task due to simulator sickness compared to men. Therefore, the sample we collected may have an overrepresentation of women who have a high tolerance for visual motion displays.

Another potential source of bias is that our sample of men may have had more experience playing video games, leading them to feeling more comfortable using the keypad within the virtual 3D environment compared to women. Studies have shown that men have more experience with video games (Lucas & Sherry, 2004) and computers (Cooper, 2006) compared to women. However, the literature is inconsistent regarding whether this type of experience results in any cognitive improvements. For instance, video game experience was not related to performance during a simulated driving assessment (e.g., Stinchcombe, Kadulina, Lemieux, Aljied, & Gagnon, 2017), but, in other studies, it was shown to improve a variety of cognitive processes such as processing speed and visual skills (e.g., Achtman, Green, & Bavelier, 2008; Dye, Green, & Bavelier, 2009). As the present study did not collect information on video game experience there was no way to determine whether there was a difference in video game experience between genders and whether this would have an effect on performance or metacognitive accuracy. However, in a previous study, using a very similar first-person virtual spatial navigation task, video game experience was shown to have no effect on performance (Therrien & Collin, 2010). Moreover, gender differences have been documented in real-world navigation studies (e.g., Galea & Kimura, 1993; Halpern, 2013; Rossano & Moak, 1998), though they are not as strong and robust in virtual environments (for a review, see Coluccia & Louse, 2004).

Overall, the current results confirm that stereotype threat/lift can influence men’s and women’s navigation performance and trial-by-trial confidence. Studies have shown that a lack of confidence in spatial cognition from a young age results in fewer girls having an interest in STEM (Science, Technology, Engineering, and Mathematics) fields through school and for women in their careers (for reviews, see Ceci & Williams, 2007; Halpern et al., 2007). These results, along with other studies of the effects of stereotype threat/lift on spatial cognition, have important implications regarding the development of programs to help girls and women improve their confidence and performance in spatial cognition by preventing and counteracting the sociocognitive effects of damaging stereotypes. This is especially important considering the continuing under-representation of women in STEM fields where spatial cognition skills are often required.

An important negative finding in the current data is that stereotype threat/lift had no effect on local metacognitive accuracy. Thus, our results are not compatible with the idea of including local metacognitive accuracy in the Schmader et al. (2008) process model for stereotype threat. That is, we find no evidence that stereotype threat reduces the accuracy of self-assessing future or past performance, and thus there is no suggestion that it is this factor which limits performance. Rather, it seems there is either a more direct link between self-confidence and performance, or other factors act as intermediaries between them.

The results demonstrate a more complex structure at play between local metacognitive monitoring and navigation performance between the genders as women occasionally demonstrated poorer monitoring accuracy compared to men but not always and not on all measures. This finding is especially important considering most research on the subject focuses on global measures of metacognition, where an underlying practical theoretical model does not exist. Thus, understanding differences in local metacognition between the genders better allows for theoretical depth compared to only studying global metacognitive measures. Future research is necessary to further determine how local metacognition differs between the genders, how these differences translate to differences in control, and, subsequently, how these differences affect performance.

References

Achtman, R. L., Green, C. S., & Bavelier, D. (2008). Video games as a tool to train visual skills. *Restorative Neurology and Neuroscience, 26*(4-5), 435-446.

Allen, R. C. (2000). *The effect of restricted field of view on locomotion tasks, head movements, and motion sickness.* (Doctoral Dissertation), University of Central Florida, Retrospective Theses and Dissertations: 1758. Retrieved from http://stars.library.ucf.edu/rtd/1758

Allison, C., Redhead, E. S., & Chan, W. (2017). Interaction of task difficulty and gender stereotype threat with a spatial orientation task in a virtual nested environment. *Learning and Motivation, 57*, 22-35. doi:10.1016/j.lmot.2017.01.005

Astur, R. S., Purton, A. J., Zaniewski, M. J., Cimadevilla, J., & Markus, E. J. (2016). Human sex differences in solving a virtual navigation problem. *Behavioral Brain Research, 308*, 236-243. doi:10.1016/j.bbr.2016.04.037 27108050

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W H Freeman/Times Books/Henry Holt & Co.

Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: can skills that don't rely heavily on working memory still be threatened? *Pers Soc Psychol Bull, 32*(8), 1059-1071. doi:10.1177/0146167206288489

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 73-94). New York, NY: Psychology Press.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7*(1), 26-48. doi:10.3758/BF03210724 10780019

Casey, M. B. (1996). Gender, sex, and cognition: Considering the interrelationship between biological and environmental factors. *Learning and Individual Differences, 8*(1), 39-53. doi:10.1016/S1041-6080%2896%2990006-0

Ceci, S. J., & Williams, W. M. (2007). *Why aren't more women in science?: Top researchers debate the evidence*. Washington, DC: American Psychological Association.

Classen, S., Bewernitz, M., & Shechtman, O. (2011). Driving simulator sickness: an evidence-based review of the literature. *Am J Occup Ther, 65*(2), 179-188.

Collaer, M. L., & Hines, M. (1995). Human behavioral sex differences: a role for gonadal hormones during early development? *Psychological Bulletin, 118*(1), 55-107.

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology, 24*(3), 329-340. doi:10.1016/j.jenvp.2004.08.006

Cooke-Simpson, A., & Voyer, D. (2007). Confidence and gender differences on the mental rotations test. *Learning and Individual Differences, 17*(2), 181-186. doi:10.1016/j.lindif.2007.03.009

Cooper, J. (2006). The digital divide: The special case of gender. *Journal of Computer Assisted Learning, 22*(5), 320-334.

Coutinho, S. A., & Neuman, G. (2008). A model of metacognition, achievement goal orientation, learning style and self-efficacy. *Learning Environments Research, 11*(2), 131-151. doi:10.1007/s10984-008-9042-7

Delgado, A. R., & Prieto, G. (2008). Stereotype threat as validity threat: The anxiety-sex-threat interaction. *Intelligence, 36*(6), 635-640. doi:10.1016/j.intell.2008.01.008

DiZio, P., & Lackner, J. (1998). *Alleviation of motion sickness and postural instability during and after virtual environment exposure*. Retrieved from Naval Air Warfare Center Training Systems Division: Orlando, FL

Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory & Cognition, 33*(6), 1096-1115. doi:10.3758/BF03193216 16496729

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences, 47*, 103-116. doi:10.1016/j.lindif.2015.12.018

Draper, M. H., Viirre, E. S., Furness, T. A., & Gawron, V. J. (2001). Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Human Factors, 43*(1), 129-146.

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 103-140). San Diego, CA: Elsevier Academic Press.

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology-Learning Memory and Cognition, 27*(5), 1180-1191. doi:10.1037/0278-7393.27.5.1180

Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2016). Methodology for investigating human metamemory: Problems and pitfalls *The Oxford handbook of metamemory* (pp. 23-37). New York, NY: Oxford University Press.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271-280. doi:10.1016/j.learninstruc.2011.08.003

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science, 18*(6), 321-326. doi:10.1111/j.1467-8721.2009.01660.x 20485453

Estes, Z., & Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. *Arch Sex Behav, 41*(3), 550-570. doi:10.1007/s10508-011-9875-5 22130691

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods, 41*(4), 1149-1160.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, England: Sage Publications.

Frank, D. J., & Kuhlmann, B. G. (2016). More Than Just Beliefs: Experience and Beliefs Jointly Contribute to Volume Effects on Metacognitive Judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*(680-693). doi:10.1037/xlm0000332 27709983

Galea, L. A., & Kimura, D. (1993). Sex differences in route-learning. *Personality and Individual Differences, 14*(1), 53-65. doi:10.1016/0191-8869%2893%2990174-2

Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the behavioral sciences* (10th ed.). Belmont, California: Wadsworth Publishing.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*(1), 160-170. doi:10.1037/0022-0663.92.1.160

Halari, R., Hines, M., Kumari, V., Mehrotra, R., Wheeler, M., Ng, V., & Sharma, T. (2005). Sex Differences and Individual Differences in Cognitive Performance and Their Relationship to Endogenous Gonadal Hormones and Gonadotropins. *Behavioral Neuroscience, 119*(1), 104-117. doi:10.1037/0735-7044.119.1.104 15727517

Halpern, D. F. (2013). *Sex differences in cognitive abilities* (4th ed.). New York, NY; Psycholgy Press.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological science in the public interest, 8*(1), 1-51.

Harris, L. J. (1981). Sex-related variations in spatial skill. In L. S. Liben, A. H. Patterson, & N. Newcombe (Eds.), *Spatial representation and behavior across the life span: theory and application* (pp. 83-111). Ann Arbor, MI: Academic Press.

Hausmann, M., Schoofs, D., Rosenthal, H. E. S., & Jordan, K. (2009). Interactive effects of sex hormones and gender stereotypes on cognitive sex differences-A psychobiosocial approach. *Psychoneuroendocrinology, 34*(3), 389-401. doi:10.1016/j.psyneuen.2008.09.019

Hebb, D. O., & Williams, K. (1946). A method of rating animal intelligence. *Journal of General Psychology, 34*, 59-65. doi:10.1080/00221309.1946.10544520 21015350

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220), 41-45.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30*(5), 425-447.

Heil, M., Jansen, P., Quaiser-Pohl, C., & Neuburger, S. (2012). Gender-specific effects of artificially induced gender beliefs in mental rotation. *Learning and Individual Differences, 22*(3), 350-353. doi:10.1016/j.lindif.2012.01.004

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Metamemory in adulthood: differentiating knowledge, belief, and behavior. In T. M. Hess (Ed.), *Aging and cognition: knowledge organization and utilization* (pp. 161-212). Oxford, England: Elsevier.

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. *Current Directions in Psychological Science, 20*(3), 167-173. doi:10.1177/0963721411409026 24478539

Hines, M., Fane, B. A., Pasterski, V. L., Mathews, G. A., Conway, G. S., & Brook, C. (2003). Spatial abilities following prenatal androgen abnormality: targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. *Psychoneuroendocrinology, 28*(8), 1010-1026.

Hogg, R. V. (1979). Statistical robustness: One view of its use in applications today. *The American Statistician, 33*(3), 108-115.

Huber, P. J. (2011). *Robust statistics*. Hoboken, NJ: John Wiley & Sons.

Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigma as ego depletion: how being the target of prejudice affects self-control. *Psychol Sci, 17*(3), 262-269. doi:10.1111/j.1467-9280.2006.01695.x

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: examining the influence of emotion regulation. *J Exp Psychol Gen, 137*(4), 691-705. doi:10.1037/a0013834

Kelemen, W. L., Frost, P. J., & Weaver, C. A., III. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition, 28*(1), 92-107. doi:10.3758/BF03211579 10714142

Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: The MIT Press.

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences, 17*(2), 161-173. doi:10.1016/j.lindif.2007.03.004

Koriat, A. (2006). Are we frightened because we run away? Some evidence from metacognitive feelings *Memory and emotion: Interdisciplinary perspectives* (pp. 83-103). Malden: Blackwell Publishing.

Koriat, A. (2007). Metacognition and consciousness *The Cambridge handbook of consciousness* (pp. 289-325). New York, NY: Cambridge University Press.

Koriat, A., Ackerman, R., Adiv, S., Lockl, K., & Schneider, W. (2014). The effects of goal-driven and data-driven regulation on metacognitive monitoring during learning: A developmental perspective. *Journal of Experimental Psychology: General, 143*(1), 386-403. doi:10.1037/a0031768 23421442

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol, 4*, 1-12. doi:10.3389/fpsyg.2013.00863

Lawton, C. A. (1994). Gender Differences in Way-Finding Strategies - Relationship to Spatial Ability and Spatial Anxiety. *Sex Roles, 30*(11-12), 765-779. doi:10.1007/Bf01544230

Lawton, C. A., Charleston, S. I., & Zieles, A. S. (1996). Individual- and gender-related differences in indoor wayfinding. *Environment and Behavior, 28*(2), 204-219. doi:10.1177/0013916596282003

Lawton, C. A., & Kallai, J. (2002). Gender differences in wayfinding strategies and anxiety about wayfinding: A cross-cultural comparison. *Sex Roles, 47*(9-10), 389-401. doi:10.1023/A:1021668724970

Lemieux, C. L., Collin, C. A., & Watier, N. N. (2017a). *Gender Differences in Judgments of Learning and Retrospective Confidence in a First-Person Virtual Maze Learning Task*. Manuscript in preparation.

Lemieux, C. L., Collin, C. A., & Watier, N. N. (2017b). *Gender Differences in Metacognitive Judgments and Performance on a First-Person Virtual Maze Learning Task.* Manuscript submitted for publication.

Leonesio, R. J., & Nelson, T. O. (1990). Do Different Metamemory Judgments Tap the Same Underlying Aspects of Memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 16*(3), 464-470. doi:10.1037//0278-7393.16.3.464

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic Status Modifies the Sex Difference in Spatial Skill. *Psychological Science, 16*(11), 841-845. doi:10.1111/j.1467-9280.2005.01623.x 16262766

Lieberman, D. A. (2004). *Learning and Memory: An Integrative Approach*. Belmont, United States: Thomson/Wadsworth.

Lin, J. J. W., Duh, H. B. L., Parker, D. E., Abi-Rached, H., & Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings IEEE Virtual Reality*, 164-171. doi:10.1109/Vr.2002.996519

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*(6), 1479-1498. doi:10.2307/1130467 4075870

Livingstone-Lee, S. A., Zeman, P. M., Gillingham, S. T., & Skelton, R. W. (2014). Navigational strategy may be more a matter of environment and experience than gender. *Learning and Motivation, 45*, 30-43. doi:10.1016/j.lmot.2013.09.003

Lucas, K., & Sherry, J. L. (2004). Sex Differences in Video Game Play: A Communication-Based Explanation. *Communication Research, 31*(5), 499-523. doi:10.1177/0093650204267930

Lundeberg, M. A., Fox, P. W., & Puncochar, J. (1994). Highly Confident but Wrong - Gender Differences and Similarities in Confidence Judgments. *Journal of Educational Psychology, 86*(1), 114-121. doi:10.1037/0022-0663.86.1.114

Martens, J., & Antonenko, P. D. (2012). Narrowing gender-based performance gaps in virtual environment navigation. *Computers in Human Behavior, 28*(3), 809-819. doi:10.1016/j.chb.2012.01.008

Massa, L. J., Mayer, R. E., & Bohon, L. M. (2005). Individual differences in gender role beliefs influence spatial ability test performance. *Learning and Individual Differences, 15*(2), 99-111. doi:10.1016/j.lindif.2004.11.002

McDonald-Miszczak, L., Gould, O. N., & Tychynski, D. (1999). Metamemory predictors of prospective and retrospective memory performance. *Journal of General Psychology, 126*(1), 37-52. doi:10.1080/00221309909595350

McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology, 27*(5), 486-493. doi:10.1016/j.appdev.2006.06.003

Modirrousta, M., & Fellows, L. K. (2008). Dorsal medial prefrontal cortex plays a necessary role in rapid error prediction in humans. *The Journal of Neuroscience, 28*(51), 14000-14005. doi:10.1523/JNEUROSCI.4450-08.2008 19091989

Moe, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences, 16*(4), 369-377. doi:10.1016/j.lindif.2007.01.002

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language, 93*, 245-258. doi:10.1016/j.jml.2016.10.008

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language, 70*, 1-12. doi:10.1016/j.jml.2013.09.007

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109-133. doi:10.1037/0033-2909.95.1.109 6544431

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education, 74*(1), 7-28.

Nori, R., & Piccardi, L. (2015). I believe I'm good at orienting myself... But is that true? *Cognitive Processing, 16*(3), 301-307. doi:10.1007/s10339-015-0655-3

Nowak, N. T., Murali, A., & Driscoll, I. (2015). Factors related to sex differences in navigating a computerized maze. *Journal of Environmental Psychology, 43*, 136-144. doi:10.1016/j.jenvp.2015.06.007

O'Laughlin, E. M., & Brubaker, B. S. (1998). Use of landmarks in cognitive mapping: Gender differences in self report versus performance. *Personality and Individual Differences, 24*(5), 595-601. doi:10.1016/S0191-8869%2897%2900237-7

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychol Rev, 15*(3), 105-130. doi:10.1007/s11065-005-7091-6 16328731

Picucci, L., Caffo, A. O., & Bosco, A. (2011). Besides navigation accuracy: Gender differences in strategy selection and level of spatial confidence. *Journal of Environmental Psychology, 31*(4), 430-438. doi:10.1016/j.jenvp.2011.01.005

Rosenthal, H. E. S., Norman, L., Smith, S. P., & McGregor, A. (2012). Gender-Based Navigation Stereotype Improves Men's Search for a Hidden Goal. *Sex Roles, 67*(11-12), 682-695. doi:10.1007/s11199-012-0205-8

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis : focused comparisons in the analysis of variance*. Cambridge Cambridgeshire ; New York: Cambridge University Press.

Rossano, M. J., & Moak, J. (1998). Spatial representations acquired from computer models: Cognitive load, orientation specificity and the acquisition of survey knowledge. *British Journal of Psychology, 89*(3), 481-497. doi:10.1111/j.2044-8295.1998.tb02698.x

Saucier, D. M., Green, S. M., Leason, J., MacFadden, A., Bell, S., & Elias, L. J. (2002). Are sex differences in navigation caused by sexually dimorphic strategies or by differences in the ability to use the strategies? *Behavioral Neuroscience, 116*(3), 403-410. doi:10.1037/0735-7044.116.3.403 12049321

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *J Pers Soc Psychol, 85*(3), 440-452. doi:10.1037/0022-3514.85.3.440

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336-356. doi:10.1037/0033-295X.115.2.336

Sharps, M. J., Welton, A. L., & Price, J. L. (1993). Gender and Task in the Determination of Spatial Cognitive Performance. *Psychology of Women Quarterly, 17*(1), 71-83. doi:10.1111/j.1471-6402.1993.tb00677.x

Shore, D. I., Stanford, L., MacInnes, W. J., Klein, R. M., & Brown, R. E. (2001). Of mice and men: Virtual Hebb-Williams mazes permit comparison of spatial learning across species. *Cognitive Affective & Behavioral Neuroscience, 1*(1), 83-89. doi:10.3758/Cabn.1.1.83

Silverman, I., Choi, J., & Peters, M. (2007). The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Arch Sex Behav, 36*(2), 261-268. doi:10.1007/s10508-006-9168-6 17351740

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 204-221. doi:10.1037/0278-7393.26.1.204 10682298

Stankov, L. (1999). Mining on the “no man’s land” between intelligence and personality. In P. L. Ackerman, Kyllonen, P. C. & Roberts R. D. (Ed.), *Learning and Individual Differences: Process, Trait, and Content Determinants.* (pp. 315-337). Washington, DC: American Psychological Association.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol, 69*(5), 797-811.

Stevens, C. A., & Carlson, R. A. (2016). Are you sure the library is that way? Metacognitive monitoring of spatial judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(7), 1034-1049. doi:10.1037/xlm0000229 26866660

Stinchcombe, A. (2014). *Understanding the challenges of the older driver: Attention, road complexity and assessment.* (Doctoral Dissertation), University of Ottawa, Available from Dissertation Abstracts International, Section B: The Sciences and Engineering.

Stinchcombe, A., Kadulina, Y., Lemieux, C. L., Aljied, R., & Gagnon, S. (2017). Driving is not a game: Video game experience is associated with risk-taking behaviors in the driving simulator. *Computers in Human Behavior, 69*, 415-420. doi:10.1016/j.chb.2016.12.006

Therrien, M. E., & Collin, C. A. (2010). Spatial vision meets spatial cognition: Examining the effect of visual blur on human visually guided route learning. *Perception, 39*(8), 1043-1064. doi:10.1068/p5991 20942357

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66-73. doi:10.1037/0022-0663.95.1.66

Tobias, S., & Everson, H. T. (1997). Studying the relationship between affective and metacognitive variables. *Anxiety, Stress & Coping: An International Journal, 10*(1), 59-81. doi:10.1080/10615809708249295

Trochim, W., & Donnelly, J. (2006). The research knowledge methods base. Cincinatti, OH: Atomic Dog Publishing

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 47A*(3), 631-650. doi:10.1080/14640749408401131

Veenman, M. V. J., Kerseboom, L., & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: Availability versus production deficiencies. *Anxiety, Stress & Coping: An International Journal, 13*(4), 391-412. doi:10.1080/10615800008248343

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of Sex-Differences in Spatial Abilities - a Metaanalysis and Consideration of Critical Variables. *Psychological Bulletin, 117*(2), 250-270. doi:10.1037/0033-2909.117.2.250

Walton, G. M., & Cohen, G. L. (2003). Stereotype Lift. *Journal of Experimental Social Psychology, 39*(5), 456-467. doi:10.1016/S0022-1031(03)00019-2

West, R. L., Boatwright, L. K., & Schleser, R. (1984). The link between memory performance, self-assessment, and affective status. *Exp Aging Res, 10*(4), 197-200. doi:10.1080/03610738408258464 6535732

Wraga, M., Helt, M., Jacobs, E., & Sullivan, K. (2007). Neural basis of stereotype-induced shifts in women's mental rotation performance. *Social Cognitive and Affective Neuroscience, 2*(1), 12-19. doi:10.1093/scan/nsl041 18985116

CHAPTER 8: GENERAL DISCUSSION

Across three experiments, the goal of this research project was to determine if gender differences exist in the local metacognition of a spatial navigation task, and if this is related to gender differences in spatial navigation performance. The last experiment also investigated whether these differences are influenced by stereotype threat/lift. An ancillary goal was to determine the feasibility of adapting the Nelson and Narens (1990) metamemory framework in the context of a spatial navigation task. Each of these topics will be discussed below, including the theoretical and practical implications of the results from this research as well as the study limitations.

CROSS STUDY COMPARISON TO METACOGNITION LITERATURE

In order to help structure our hypotheses and operationally define local metacognition, the three experiments presented herein were developed around the adaptation of the Nelson and Narens (1990) metamemory framework to a first-person virtual maze-learning task. This required the instructions for each monitoring judgment to be adapted to a navigation task while still preserving the intended nature of the judgments. This was done relatively easily as the judgments chosen (EOL, JOL, and RCJ) simply require task performance to be defined as a dichotomy. In the current setting, a successful performance was defined as having completed the maze without any errors (i.e., without having crossed any of the invisible "trip wires" that defined dead-ends and wrong turns). Relative and absolute metacognitive accuracy analyses could be easily carried out with the same kind of variables typically collected on any cognitive task. In addition to slightly changing monitoring judgment instructions, task protocols were different across experiments in order to reflect the needs of each monitoring judgment. For instance, EOLs were collected in Experiment 1 and Experiment 3, and as they reflect a subjective sense of how easy or difficult one finds each maze before learning they require the participant to view the video tours without the intention to learn the layout for a future performance. It is for this reason that the 30 video tours were presented entirely before performance, to help ensure the participants focused on assessing how difficult a maze was instead of trying to learn its layout in anticipation for a future performance. Altogether, it was very easy to adapt the Nelson and Nerens (1990) metamemory framework to a first-person virtual maze-learning task and the current study supports this approach for future research.

The results from each monitoring judgment across all three experiments are generally consistent with other studies that have examined performance using word-pair associates (e.g., Koriat, 2007; Leonesio & Nelson, 1990) as well as face-name associates (e.g., Watier & Collin, 2011). That is, the metacognitive literature has shown EOLs to have the poorest correlation with performance compared to all other metacognitive judgments, with averages around .20 (Dunlosky & Matvey, 2001; Leonesio & Nelson, 1990). In the present study, gamma correlations[13] for EOLs across experiments were quite low: ranging from chance level for women to slightly above chance for men in Experiment 1 and between .06 and .26 in Experiment 3. Relative accuracy gamma correlations for RCJs across all experiments were quite high: ranging from .65 to .83 in Experiment 1, from .78 to .81 in Experiment 2, and from .66 to .75 in Experiment 3. This is in line with what is seen in the literature on the retrospective confidence-accuracy relationship for general knowledge questions, where the magnitude of the within-subject correlations is typically moderate-to-high (Koriat, 2007).

---

[13] Note that gamma correlations were transformed to G* in the statistical analyses across experiments, but for the sake of comparability with the literature, gamma correlations are discussed here.

However, the accuracy results for JOLs in Experiment 2 and 3 were quite poor compared to the literature, which reports averages from .20 to .54 (Koriat, 2007). The JOL is a judgment of learning where the participant indicates, after having seen the video tour, how confident they are that they learned the maze layout such that they will successfully complete that maze during performance. In Experiment 2, JOL gamma correlations were near 0; a floor effect occurred likely because the task was too metacognitively difficult, preventing participants from distinguishing between well learned mazes and poorly learned mazes. This is because the task required participants to remember the layouts of 6 different mazes for each block before performing them in random order. The Nelson and Narens' metamemory framework predicts higher accuracy when the information used to make the judgment is more similar to the information needed at time of test. The competition hypothesis also states that as memory interference increases monitoring decreases (e.g., McGuire & Maki, 2001; Schreiber & Nelson, 1998). Considering JOLs were collected after each video tour, after which there was the potential for other video tours and performances interfering with memory, it is possible our participants lost a lot of what they had learned. In other words, the information they used to make the JOL for a particular maze was lost by the time they performed the maze. It is also possible that at the time of giving a JOL they could not accurately predict learning interference since the performance order was randomized.

In Experiment 3, JOL gamma correlations ranged from .18 to .31, which is still relatively low. To avoid the same floor effect, the task in Experiment 3 was designed to be easier than in Experiment 2. Specifically, in Experiment 3, participants performed the maze immediately after seeing a single video tour, as opposed to having to remember several maze layouts at once as in Experiment 2. That is, there was no interference between maze video tours, which could have reduced JOL accuracy. Indeed, JOL accuracy was improved, though it was still lower than what is typically seen in the metacognitive literature. Generally speaking, JOLs are typically more accurate when they are collected soon before performance such that the monitoring processes reflect items in memory that will be used at performance (Koriat, 2007). In Experiment 3, this should be the case, as participants performed the mazes immediately after giving their JOL.

A possible explanation for both poor JOL and poor EOL accuracy could be that the video tours were viewed passively and involved travelling through the entire maze, including incorrect passages (dead-ends). This could have made it more difficult to accurately assess the difficulty of the maze layout, as research has shown lower retention rates in spatial navigation tasks when using passive learning versus active learning (see Chapter 1 for a review). It is also possible that navigation tasks are simply more complex than word-pair association tasks, making it more difficult to predict future performance. However, in contrast to our results, Schwartz (2006) showed very high relative accuracy gamma correlations for JOLs in his experiments using two navigation related memory tasks. He argued that navigation tasks may have greater ecological validity compared to word-pair association tasks and therefore might yield more developed and more accurate metacognition. The rest of our results point to this being a possibility, but perhaps future research should focus on using active learning tools as opposed to passive learning, as this may be the reason for our poor JOL accuracy results. In addition, future research should consider the high cognitive load navigation requires and prevent memory interference in order to establish adequate sensitivity when trying to measure metacognitive monitoring.

Because several metacognitive judgments were collected during each trial, it is possible that they may have influenced each other, rendering the interpretation of our analyses difficult. Therefore, in order to assess variable independence across metacognitive judgments pearson’s bivariate correlations were computed across metacognitive judgment types in each experiment. In Experiment 1, EOL G* and RCJ G* were not significantly correlated. In Experiment 2, the correlation between JOL G* and RCJ G* was .253 and significant (p<.001), which is quite small and only accounts for 6% of shared variance. In Experiment 3, which collected all three monitoring judgments for each trial, the correlation between EOL G* and RCJ G* was not significant. However, the correlation between EOL G* and JOL G* was significant (p<.05) at -.184, as was the correlation between JOL G* and RCJ G*, at .226 (p<.001). The finding that these correlations are either small or not significant, is in agreement with the metacognitive literature, which generally suggests that each monitoring judgment measures independent metacognitive processes (see Chapter 2 for a review).

In conclusion, it was quite straightforward to adapt the Nelson and Narens (1990) metamemory framework to a first-person virtual maze-learning task. As previously noted, research on metacognition, and more specifically the Nelson and Narens' framework, has been largely based around verbal tasks such as word-pair association learning. It was previously unclear to what degree this work would be applicable to spatial navigation tasks considering there is considerable evidence and theory supporting separate verbal and visual processing systems (e.g., Paivio, 1991; Perfect & Hollins, 1996, 1999; Tulving & Schacter, 1990). However, the results from the current study give a preliminary picture as to how local metacognition during navigation fits into the Nelson and Narens (1990) framework for metamemory. Indeed, this framework allowed for a theoretical structure to help with preliminary hypotheses and interpretation on the local metacognition of spatial navigation. In addition, this research has helped to begin expanding the metacognitive literature to cognitive realms outside of traditional word-pair associates, as has been done with other fields of research such as face recognition

(Watier & Collin, 2010, 2011, 2012). Future research should continue to explore ways of adapting the Nelson and Nerens (1990) metamemory framework to different types of spatial tasks. This will be useful for both research in metacognition, which will benefit from having its models tested across a broader range of cognitive tasks, and research in spatial cognition, which will benefit from a better understanding of the mechanisms that underlie it.

Gender Differences in Navigation Performance and Confidence

The present work investigated whether gender differences in performance on navigation tasks continue to exist. The first-person virtual maze learning task used in all three experiments is considered challenging, but perhaps especially for women, as it is devoid of landmarks and thus requires participants to use an orientation-based navigation strategy, which is more typically used by men. Participants need to extract graph knowledge presented during the video tours, and then extract route knowledge by finding the correct path to the exit, and remember this information until they perform the maze. While in the maze, for each turn taken they need to mentally rotate their cognitive maps and keep track of where they have been and where they are going – it is this that involves an orientation-based navigation strategy. Strong gender differences in performance on such a task are often reported (for a review, see Coluccia & Louse, 2004), likely because men, compared to women, are particularly good at using an orientation-based navigation strategy (see Chapter 3 for a review). It was therefore as expected when the current analyses revealed large and strong gender differences across all performance variables consistently for all experiments. Indeed, women, on average, failed a greater number of mazes, took twice as much time on average to complete the mazes and sometimes made up to twice as many errors compared to men.

There are several possible explanations for why women may take longer than men when navigating. Several studies have suggested it may take longer for women to change from their default route-based strategy to an orientation-based strategy, and that this explains their longer latencies compared with men (e.g., Barkley & Gabriel, 2007; Lawton, 1994; Sandstrom, Kaufman, & Huettel, 1998). In other words, women may have taken longer to complete the mazes because they needed to consciously shift their strategy for every maze whereas the task already favors the default strategy typically used by men. Other studies have demonstrated that during open exploration tasks, women generally travel slower and display more hesitant behaviors (e.g., Lavenex & Lavenex, 2010; Lawton et al., 1996). It is possible the female participants in our sample generally travelled slower with more hesitance than the male participants. However, it is important to note that, in the current study, movement speed through the maze was fixed, so any difference in travel time across a given distance would necessarily be based on behavioral differences such as hesitance. Hesistance, as a behavior, may be related to a lack of self-confidence and/or to anxiety which has been shown to either influence metacognition or metacognitive monitoring (e.g., Coutinho & Neuman, 2008; Kleitman & Stankov, 2007; McDonald-Miszczak et al., 1999; Stankov; Tobias & Everson, 1997; Veenman et al., 2000; West et al., 1984). Nonethless, several studies have shown that without a time pressure on spatial tasks the performance difference between genders disappears (e.g., Goldstein, Haldane, & Mitchell, 1990; Voyer & Sullivan, 2003) or is at least reduced (Voyer, 2011). It is still not well understood if this is because women require more time to perform and/or if their performance is impacted by stress or anxiety to a greater degree than men's performance.

Although it was not possible to know which strategy either gender was using at any given time in these experiments, the fact that women made a significantly greater number of errors indicates that they were likely not effectively using an orientation-based navigation strategy. A greater number of errors suggests that women were less able to solve or remember the correct route during the video tour, and/or keep track of where they had been as well as where they were going during navigation. Identifying the most efficient route as well as accurately identifying one's position on this route in real time both require the ability to continuously update and mentally manipulate one's environment in visual-spatial working memory. It is this ability that is central to an orientation-based navigation strategy. In the meantime, women have been shown to have difficulties with mental rotation (for a review, see Voyer & Doyle, 2010) and visual-spatial working memory compared to men (for a review, see Voyer et al., 2017). In fact, it has been proposed that the ratio of visual spatial working memory (VSWM) and verbal working memory capacity may determine the spatial strategy difference between the genders (L. Wang & Carr, 2014). Thus, the women in the present study may not have been able to identify and use the appropriate strategy due to a lack of VSWM capacity relative to verbal working memory, causing them to take a greater number of wrong turns (see Chapter 3 for a review).

Correspondingly, one recent explanation for the gender differences in spatial cognition is that generally women have weaker VSWM (for a review, see Voyer et al., 2017). Based on current evidence, it is not possible to know to what degree this difference in VSWM is due to biological differences or sociocognitive differences. For instance, it is possible that men, in general, have superior VSWM because they have more experience with spatial tasks, allowing them the chance to better develop this cognitive ability. This idea is corroborated by a recent meta-analysis that showed that age is a moderator between gender and VSWM, such that the gender difference in VSWM favoring men only emerges in young adulthood (Voyer et al., 2017). There are several sociocognitive factors that could explain why boys and men may have more experience with navigation-related tasks, which would help to develop their VSWM skills as they age (see Chapter 3 for a review). In addition, girls and women are more socialized to fear for their safety, and concern for their own safety is associated with higher levels of anxiety related to wayfinding (Lawton & Kallai, 2002). More importantly, women's higher wayfinding anxiety is related to their lower tendency to use an orientation-based navigation strategy (Lawton, 1994; Lawton & Kallai, 2002), which has been associated with women's underperformance compared to men on orientation-based navigation tasks (see Chapter 3 for a review). Thus, it is possible that spatial anxiety arises due to a lack of experience with spatial tasks, and without this experience, women are not capable of developing their VSWM capacity, which in turn prevents them from using effective holistic spatial strategies (L. Wang & Carr, 2014).

In the present studies, in addition to large gender differences in navigation performance, there were also large gender differences in prospective and retrospective confidence across all experiments. In Experiments 1 and 2, the gender difference was greatest among the prospective confidence judgments compared to the retrospective judgments. That is, generally, women were especially low in confidence prior to performance. In Experiment 3, men had significantly higher mean trial-by-trial confidence among all three confidence judgments compared to women. A number of researchers have attempted to explain this gender difference in overall confidence (see Chapter 3 for a review). Perhaps the most relevant potential explanation is that women, compared to men, attribute success more to external (e.g., Meehan & Overton, 1986) and unstable causes (e.g., LaNoue & Curtis, 1985; Parsons et al., 1982). Indeed, previous research has demonstrated that women exhibit a self-derogatory bias (e.g., Berg, Stephan, & Dodson, 1981; Erkut, 1983; R. Levine, Gillman, & Reis, 1982).

In summary, our findings clearly corroborate the considerable amount of research demonstrating confidence and performance differences between the genders in a variety of fields, including spatial navigation (e.g., Coluccia & Louse, 2004; Galea & Kimura, 1993; Halpern, 2013; Lawton et al., 1996; O'Laughlin & Brubaker, 1998; Picucci et al., 2011; Rossano & Moak, 1998; C. A. Stevens & Carlson, 2016).What is novel here is that, this is one of only a few studies to confirm confidence differences in spatial cognition on a trial-by trial basis, as opposed to global confidence measured by a self-reported questionnaire.

GENDER DIFFERENCES IN LOCAL METACOGNITION

This research supports the idea that men and women have different local metacognitive accuracy performance during a navigation task, though the results point to a more complex story than originally hypothesized. Across all three studies, men were shown to have more accurate local metacognition than women did on several variables. However, not all hypothesized metacognitive monitoring variables showed a gender difference. A gender difference was consistently seen on relative retrospective monitoring accuracy across all three experiments, though the effect sizes were all moderate to small. That is, women were consistently less accurate at assessing a previous performance compared to men. When comparing experiments, Experiment 1 showed the greatest gender difference in retrospective accuracy, with a moderate effect size. This effect was only significant in the Pro Female facilitation condition in Experiment 3 with a moderate effect size. Finally, this same effect size for Experiment 2 was small, showing men to be more accurate, though, it did not reach statistical significance.

Subsequent analyses were carried out in both Experiments 1 and 2, which determined a gender difference in the correlation means between retrospective judgments and Maze Completion Time as well as retrospective judgments and Error Count. The results for

Experiment 1 showed that with moderate effect sizes, men's Maze Completion Time and Error Count was, on average, better correlated with their retrospective judgments compared to that of women. This indicates that men may have a better ability to use both the time it took to complete the maze as well as how many errors they committed to assess their previous performance. That is, well and poorly performed mazes could be better distinguished if the participant used these measures as a cue for metacognitive monitoring. Intriguingly, the reverse effect was significant in Experiment 2, where women's Maze Completion Time was, on average, better correlated with their retrospective judgments compared to that of men. This reverse effect may be due to task differences between Experiment 1 and 2. The task in Experiment 2 was, generally, more difficult compared to Experiment 1, which increased the time it took to complete the mazes, as well as the variability in maze completion times. This was true for both genders but more so for women than for men. It is possible that this cue was better used by women in Experiment 2 because women actually had much greater variability in their maze completion times, allowing them to better use this cue for distinguishing between good and poor past performances. The gender difference in correlation means between Error Count and retrospective judgments was not significant in Experiment 2, though both genders had relatively strong correlations suggesting that they both used errors as a cue for monitoring. Considering these were the only obvious measures of performance in the current task, it is possible that the use of these cues follows the accessibility hypothesis put forth by Koriat (1995, 1997), whereby participants use whatever performance information they can best remember to make monitoring judgments. Although cue utilization theories are not new, this is the first study to find gender differences in metacognitive monitoring cue utilization during a navigation task. Such gender differences are expected, however, as men and women are known to generally use different types of spatial cues to help with navigation performance (see Chapter 3 for a review). This may result in different types of cues being readily available in memory to be used for metacognitive monitoring. While future research is necessary in order to better understand the gender differences in cue utilization for metacognitive monitoring, we can conclude that these performance cues are strongly related to retrospective monitoring judgments irrespective of gender. Future research is needed in order to help determine if there are other performance cues strongly related to metacognitive monitoring.

Although there were no consistent findings indicating a gender difference for prospective relative metacognitive accuracy (EOL and JOL), it is not possible to discount a potential gender difference. The results for these judgments demonstrate quite poor metacognitive accuracy in both men and women. This may indicate the measures were not sensitive enough to detect a gender difference as results between both genders were at floor. This floor effect may have arisen due to methodological limitations. These limitations include a lack of statistical power for small effect sizes and the use of passive video tours as learning stimuli. That is, as passive learning tools are less engaging than active ones, this may have prevented participants from accurately assessing the difficulty of the mazes (see Chapter 1 for a review).

Regarding the absolute accuracy analyses, we consistently see that women have poorer absolute metacognitive accuracy prior to performance compared to men. That is, men’s mean confidence levels match their overall performance levels consistently, but this is not the case with women. However, the direction in which women were inaccurate varied. In Experiments 1 and 2 women were under confident prior to performance whereas the women in Experiment 3 were overconfident prior to performance for 2 of the 3 Facilitation conditions. This inconsistency for women across experiments could be related to the fact that Experiment 3 was obviously a gender-based study, which could have elicited stereotype-based stress in women, regardless of facilitation condition, and thus impaired their absolute metacognitive accuracy. This may have incited women to self-report greater levels of confidence after performance no matter the facilitation condition. Nonetheless, women were clearly less well able to predict how well they would perform compared to men. Generally speaking, both genders were equally overconfident when assessing past performance across all experiments. These results give further support for the “overconfidence effect" (see Chapter 2 for a review), whereby most individuals judge their performance or abilities to be better than they actually are.

According to the Nelson and Narens (1990) metamemory framework, accurate monitoring accuracy allows an individual to better strategize corrective control behaviors, which is related to better performance. It is therefore possible that women’s lower monitoring accuracy leads them to poorer control strategies, which may be related to their poor performance, relative to men. That is, with poorer monitoring accuracy one does not know which items to study better or which strategies are best used to maximize performance; this could be the case for women during navigation tasks. This idea is consistent with previous research showing women to have difficulty in switching from their default landmark-based navigation strategy to a more suitable orientation-based navigation strategy (for a review, see Coluccia & Louse, 2004; but also see Chapter 3).

A possible explanation for this gender difference in metacognitive accuracy is that as men generally have more experience with spatial tasks and/or virtual environments (see Chapter 3 for a review), this may have allowed them to better develop these cognitive skills including the ability to discriminate between low and high performances. That is, with experience comes the ability to better recognize monitoring cues and resources, which in turn can help to identify more effective navigation strategies.

In Experiment 3, global metacognition measures (i.e., SAS and SBSOD) were collected at the end of testing. The results corroborate the literature in that women rated themselves lower in their spatial abilities than did men across all three facilitation conditions. In addition, women also indicated greater levels of spatial anxiety than did men across all three facilitation conditions. Indeed, these findings are consistent with the literature in showing women to have greater levels of spatial anxiety and lower self-rated spatial abilities (see Chapter 3 for a review). The distinction between this global form of metacognition and the local form is based on the object of the metacognitive assessment. Global metacognition involves assessing one's general cognitive abilities in the absence of an actual task or stimulus; local metacognition is more task-specific and relates to a particular situation. Despite these differences, the memory representations that are used to make these different kinds of self-assessments likely overlap. For example, global assessments of cognitive ability might be affected by recent experiences with cognitive tasks, and local assessments of task performance might be affected by general representations of self-efficacy (see Chapter 2 for a review).

Overall, the gender difference in local metacognition is complex and depends on several factors such as task demands and type of metacognitive monitoring. We have shown that men, compared to women, have more accurate absolute local metacognition prior to performance as well as more accurate relative local metacognition after performance. With less accurate metacognition, women may not be making appropriate control behaviors such as changing to a more appropriate navigation strategy. With generally more experience in spatial tasks, men have developed better metacognition and likely know which navigation strategies they can best use based on the available information. This finding is especially important considering most research on the subject focuses on global measures of metacognition, where an underlying practical theoretical model does not exist. Thus, understanding differences in local metacognition between the genders allows for greater theoretical depth compared to only studying global metacognitive measures. Future research is necessary to further determine how local metacognition differs between the genders on spatial tasks, how these differences translate to differences in control, and, subsequently, how these differences affect spatial performance.

THE EFFECTS OF STEREOTYPE ACTIVATION

Schmader et al. (2008) proposed that stereotype threat increases stress arousal and performance monitoring, which are related to self-doubt, and that this takes away resources from working memory and executive function (see Chapter 3 for a review on stereotype threat). The current study hypothesized a connection between this model and the Nelson and Narens (1990) metamemory framework in order to examine the role that metacognitive monitoring accuracy and confidence might play in stereotype threat. We hypothesized that stereotype threat would decrease task performance by reducing metacognitive monitoring accuracy via changes in self-confidence. The opposite effect is expected (stereotype lift) for the gender being positively stereotyped. In order to test this hypothesis, we used three different stereotype facilitation instructions in Experiment 3 to determine if confidence, metacognitive monitoring accuracy and navigation performance are affected by stereotype activation. For all these variables, the gender difference was expected to be greatest in the Pro Male facilitation condition as stereotype activation was hypothesized to further exagerate this gender difference by boosting men's and/or threatening women's confidence, monitoring accuracy, and performance. Following this same line of reasoning, the gender difference was expected to be smallest in the Pro Female facilitation condition, as the reverse stereotype activation should have the reverse effects by boosting women's and/or threatening men's confidence, monitoring accuracy, and performance. Thus, the

Pro Female facilitation condition was expected to counteract and therefore reduce the existing gender difference. In addition, the size of the gender difference in the Neutral faciliation condition was expected to be in between these two conditions.

The results of Experiment 3 confirmed that stereotype activation had an effect on most navigation performance and metacognitive confidence variables. Specifically, the effect size for gender difference was largest in the Pro Male facilitation condition and smallest in the Pro Female facilitation condition for Error Count, Mazes Passed, EOL confidence, JOL confidence, and RCJ confidence. However, stereotype activation did not have a consistent effect on any of the metacognitive monitoring accuracy variables. That is, the hypothesized trend for stereotype activation was observed for most performance and confidence variables but not the monitoring accuracy variables. This demonstrates that metacognitive monitoring accuracy is not affected by stereotype threat and therefore should not be added into the stereotype threat process model proposed by Schmader et al. (2008). However, the current study was able to extend the effects of stereotype activation on women and men's navigation performance as well as confidence on a first-person virtual spatial navigation task.

The only performance or confidence variable to not show this trend was Maze Completion Time, which showed the greatest effect size among the Neutral facilitation condition by far at 1.725. The smallest effect size was in the Pro Male facilitation condition at 1.360, with the Pro Female facilitation condition in the middle at 1.394. All three of these effect sizes are very large; in fact, they are the largest across the entire experiment. This indicates a very large main effect whereby men take significantly less time, on average, to complete a maze compared to women no matter the stereotype facilitation condition. In light of research showing that women exhibit more hesitant and anxious navigational behaviors (see Chapter 3 for a review), it is possible that the Neutral facilitation condition had the largest effect because it was the condition that induced the most amount of spatial anxiety for women according to SAS scores.

Contrary to what was expected, men rated their general spatial abilities highest in the Pro Female facilitation condition while women rated themselves highest in the Neutral facilitation condition. It is unclear why the condition in which women rated themselves the most spatially anxious was also the condition they rated themselves as having the best general spatial abilities. Nonetheless, the hypothesized stereotype activation trend was not found in either global metacognition measure (general spatial ability via SBSOD and spatial anxiety via SAS).

As was previously mentioned in the introduction, a recent meta-analysis investigating the effects of stereotype activation in the context of gender differences in spatial tasks found there was no effect of stereotype threat in either men or women, and only a small effect of stereotype lift on women only (Doyle & Voyer, 2016). These findings appear to be at odds with the present work. However, there are a number of caveats to be considered. For instance, their analysis examined very little research on spatial tasks. Specifically, they compared the effect sizes from 18 manuscripts that collected data on spatial tasks, none of which involved a navigation or wayfinding related task. Therefore, the findings may apply to the broader concept of spatial tasks but not to the more specific concept of navigation abilities. Along similar lines, the authors suggest that, compared to the gender stereotype regarding math performance, there may not be a strong and pervasive stereotype that men outperform women on “spatial tasks”. However, there may nevertheless be a strong and pervasive stereotype related to men’s advantage in navigation specifically. One reasons to suspect this is that navigation is a more concrete and relatable day-to-day task compared to the more abstract and general concept of “spatial ability”. Indeed, it may actually be the spatial ability with the most pervasive stereotype, making it more susceptible to the effects of stereotype activation. However, it should be noted that this must remain speculation at present because there is no data, to our knowledge, about the relative prevalence and strength of gender-based stereotypes regarding navigation and spatial abilities. In summary, it is not possible to know if the results from this meta-analysis can be compared to stereotype activation effects in navigation/wayfinding. Future research should continue to investigate the prevalence of such gender-based stereotypes and their effects on navigation performance.

Overall, the current results confirm that stereotype activation can influence men's and women's navigation performance and trial-by-trial confidence. An important negative finding in the current data is that stereotype activation had no effect on local metacognitive accuracy. Thus, our results are not compatible with the idea of including local metacognitive accuracy in the Schmader et al. (2008) process model for stereotype threat. That is, we find no evidence that stereotype threat reduces the accuracy of self-assessing future or past performance, and thus there is no suggestion that it is this factor that limits performance. Rather, it seems there is either a more direct link between self-confidence and performance, or other factors act as intermediaries between them.

STUDY LIMITATIONS

As with any research, there are a number of study limitations that must be taken into consideration when interpreting the data presented here. One example of these is related to the lack of a true control condition in Experiment 3. The Pro Male and Pro Female facilitation conditions were explicitly designed to manipulate the existing implicit stereotypes within each participant, by either exaggerating it as in the Pro Male facilitation condition or counteracting it as in the Pro Female facilitation condition. Although the Neutral facilitation condition did not activate a stereotype in any particular direction, it still revealed to participants that the experiment was about gender by explicitly stating that the genders performs equally. It is difficult to know whether this manipulation was strong enough to have attenuated the existing internalized stereotype, and therefore truly reflects a gender equal activation, or if its effect was to prime participants into their existing internalized stereotype, which, in this case, favors men's performance. In either case, the result would be a gender difference with an effect size in between that seen in the Pro Male and Pro Female facilitation conditions. Therefore, including the Neutral facilitation condition into the hypothesized trend would further improve the validity of confirming that stereotype activation had the expected effect by increasing the specificity of the predicted outcomes. Cross facilitation condition comparisons were not carried out due to the fact that statistical power would have needed to be greatly improved in order to perform such cross facilitation condition comparisons as the effect sizes would be much smaller (for a review, see Doyle & Voyer, 2016). By considering the effects of both stereotype threat and lift simultaneously, more overall information is being used in the analysis to determine if stereotype activation had an effect at all.

Another limitation of the present work, one that exists across all three experiments, is a potential sampling bias whereby each sample may have had an over representation of women who can better tolerate visual motion displays compared to women in the general population. Because one of the inclusion criteria for recruitment was having a low susceptibility to simulator/motion sickness and women generally have a greater susceptibility to it compared to men (Classen, Bewernitz, & Shechtman, 2011), it is reasonable to assume that there may have been a greater number of women compared to men who could not be recruited due to their susceptibility to simulator/motion sickness. Additionally, once participants were recruited there were up to 4 times more women than men who had to terminate their participation due to simulator/motion sickness symptoms across all experiments. This would exacerbate the over representation of women who can tolerate visual motion displays across all our experiments. However, to our knowledge, there is no evidence that tolerance for visual displays correlates with spatial or navigation abilities, so the degree to which this would affect the results is unclear. In connection with the following limitation, women's greater susceptibility to simulator sickness may be related their lack of video game experience.

Another important potential source of bias is related to the fact that women have less experience with video games (Lucas & Sherry, 2004) and computers (Cooper, 2006) compared to men. This difference in experience likely leads men to feel more comfortable using the keypad within the virtual 3D environment. It may also, at least partially, explain the performance difference between the genders on any virtual tasks, including the one used in the present study. Indeed, experience playing these video games, which are often three dimensional and have a first person perspective just like the current task, has been shown to result in better performance on research-based first-person navigation tasks likely due to learning transfer and platform familiarity (e.g., Murias et al., 2016; but see Chapter 1 for more on video game experience). Conversely, there is also research that did not find video game experience to affect performance in a virtual environment (e.g., Stinchcombe et al., 2017). Additionally, two previous studies, using very similar first-person virtual spatial navigation tasks as the current study, showed video game experience to have no significant effect on performance (Shore et al., 2001; Therrien & Collin, 2010). Moreover, gender differences have been documented in real-world navigation studies (e.g., Galea & Kimura, 1993; Halpern, 2013; Rossano & Moak, 1998), though they are not as strong and robust in virtual environments (for a review, see Coluccia & Louse, 2004). Finally, while experience with video games may improve performance, it is not clear that it would necessarily affect metacognitive accuracy for a different spatial task. Nevertheless, it is important to take into consideration these limitations when interpreting the results of a gender-based analysis for any study using virtual reality.

Social desirability is a problem with any self-reported measure, and this includes the metacognitive monitoring judgments collected in this study. Although Experiments 1 and 2 made no reference to gender, it was clear to participants that Experiment 3 was related to gender differences. This may have incited participants to give responses they considered to be socially desirable with regards to gender. For instance, men may have given lower confidence and performance scores in an effort to be socially desirable and to help counteract the existing stereotype that women do not navigate as well as men. Conversely, men may have given higher ratings in an attempt to fit the gender stereotype that men are confident. In either case, the effect may have been exacerbated by the fact that the only experimenter to give the stereotype facilitation instructions was female. Studies have shown that the gender of the experimenter influences stereotype activation effects (e.g., Boutcher, Fleischer-Curtian, & Gines, 1988; F. M. Levine & de Simone, 1991; A. L. Nichols & Maner, 2008). Indeed, a recent meta-analysis on stereotype threat/lift for math and spatial tasks determined that gender of the experimenter was the only significant moderator for stereotype threat in women. The findings showed that having both male and female experimenters simultaneously present showed the greatest threat activation effects (Doyle & Voyer, 2016). It is difficult to hypothesize how this would have influenced both male and female participants' behavior and performance in the current study, though it seems that including a male experimenter alongside the female one would have likely enhanced stereotype activation.

Despite having conducted a priori power analyses using a moderate effect size for all experiments, some analyses were underpowered as some effect sizes were small. As there was no previous research allowing us to estimate an effect size for gender differences in local metacognition, it was not possible to perform a fully informed power analysis. Future research on gender differences in local metacognition should expect small to moderate effect sizes. It is important to note that without adequate power to detect gender differences in local metacognition it would not be possible to detect the effects of stereotype threat either. It is possible that this was the case in Experiment 3; differences in local metacognition as a result of stereotype activation may have been too small to be detected with the current dataset.

PRACTICAL IMPLICATIONS OF THIS RESEARCH

The studies presented in this thesis have contributed to research showing a persistent gender difference on spatial tasks. Spatial cognition is widely recognized as an important skill. Among other things, it is needed to not only maintain interest in science and math courses in school, but is generally considered an essential skill in STEM fields (Kell, Lubinski, Benbow, & Steiger, 2013). Whether it is for visualizing the trajectory of a piston in a complex mechanical system or imagining how a complex set of molecules interact with each other, spatial skills have also been linked to creativity and technological innovation (Kell et al., 2013). It is clear that spatial cognition is an ability that helps individuals succeed in many of today's most profitable and prestigious employment markets. Various organizations have argued that there is a need to better identify and develop skills and interest in STEM fields in order to further promote intellectual innovation (Friedman, 2007; National Science Board, 2010; U.S. Department of Commerce, 2012). At the same time, the persistent wage gap between the genders is partially due to a lack of female employment in STEM fields (Buffington, Cerf, Jones, & Weinberg, 2016).

Thus, increasing girl's and women's skill and confidence in spatial cognition would likely result in an increase in women's involvement in STEM, which would in turn help to close the gender pay gap as well as further increase global intellectual innovation.

The current research helps to illuminate some of the sociocognitive contributors to gender differences in spatial tasks as well as some potential solutions to improving women's interest and involvement in STEM fields. Although there is research demonstrating that biological factors may partially explain performance and interest differences between the genders in spatial cognition (see Chapter 3 for a review), there have also been several sociocognitive factors identified, including, among many others, differences in childhood and adult socialization as well as gender roles. In addition, several studies have shown that girls' interest in STEM falls steeply at the onset of puberty, with both biological and sociocognitive factors thought to be at play (for reviews, see Ceci & Williams, 2007; Halpern et al., 2007). While one can argue over the relative importance of the effects from nature and nurture in determining the gender difference in STEM fields, the current research does not specifically support either side. That is, both biological and sociocognitive factors likely contribute to this gender difference in a complex way. However, contrary to biological determinants, the sociocognitive factors preventing girls and women from having interest and skills in STEM can be reformed. Studies as early as the 1960s have indicated that the gender difference in spatial abilities likely contributes to the gender disparity in mathematics-based fields that draw on spatial skills, such as physics and engineering (e.g., Fennema, 1975; Maccoby & Jacklin, 1974; Sherman, 1967). While such findings might have been used as a justification for the underrepresentation of women in STEM fields, researchers more recently have argued that the discussion would be better served by focusing on the aspects that are holding women back from entering these fields (e.g., Hines, 2007; Newcombe, 2007). As previously mentioned, there are several hypotheses about how to support and improve girls' and women's spatial skills (e.g., S. C. Levine et al., 2016; Miller et al., 2015).

In addition to training programs which have been shown, via two meta-analyses, to be effective at improving women's spatial skills (Baenninger & Newcombe, 1989; Uttal et al., 2013), gender stereotypes may be a potential locus of intervention and prevention; by working to abolish stereotype threat, and thus improving confidence in women, we may be able to further improve their spatial skills and thus their success in STEM fields. Studies have shown that the gender stereotype associating science with men exist across cultures (e.g., Miller et al., 2015; Nosek et al., 2009) and emerges early in development (e.g., Chambers, 1983; Steffens, Jelenec, & Noack, 2010). Over 40 years ago, nearly 5,000 North American children were asked to a draw a picture of a scientist, and only 0.6% depicted a woman scientist (Chambers, 1983). Thankfully, these associations have weakened over time with 35% of American children depicting a woman scientist (e.g., Fralick, Kearn, Thompson, & Lyons, 2009; Tippett & Milford, 2013). Studies have shown that these changes in stereotype magnitude across cultures mirror women's increasing participation in science (e.g., Hill, Corbett, & St Rose, 2010; Miller et al., 2015). Interestingly, even nations with high overall gender equity still have strong gender-science stereotypes (Miller et al., 2015). More specifically, studies have shown that exposure to successful women scientists and mathematicians can weaken stereotypes related to STEM fields among both young girls and adult women (e.g., Galdi, Cadinu, & Tomasetto, 2014; Mason, Kahle, & Gardner, 1991; Young, Rudman, Buettner, & McLean, 2013). Thus, one way to combat the effects of stereotype could be to support social programs aimed at further exposing children and the public to the great work done by existing female scientists.

The results from the current study provide evidence that stereotype activation had an effect on both men and women's performance and confidence on a spatial task. That is, one simple sentence prior to performance was able to change the way men and women performed on a spatial task such that the gender difference in performance was reduced. If the gender gap in performance on a spatial task can be so easily reduced, one can just imagine the beneficial effects of creating early childhood programs whose purpose is to improve girls' confidence and interest in spatial tasks and prevent girls' and women's attrition in STEM fields by eliminating the damaging effects of stereotypes. In addition, further research on the ways in which the genders differ in their local metacognition would likely help future social programs to better target cognitive tools that could better assist in women's spatial cognition performance.

CONCLUSIONS

Taken together, the work presented in this thesis serves to illuminate possible explanations for the persistent gender difference in navigation tasks by studying differences in metacognition. This was accomplished by adapting a pre-existing methodology to studying metacognition, the Nelson and Narens (1990) framework, to the context of a first-person virtual spatial navigation task. Doing so showed that there are indeed strong gender differences in first-person virtual maze navigation performance as well as in confidence on such tasks, both of which favored men. Gender differences were not found for prospective local metacognitive relative accuracy, which could have been due to poor statistical power and/or methodological limitations. Gender differences were found for retrospective local metacognitive relative accuracy, showing men to be more accurate in their assessment of their own prior performance. This finding is especially important considering most research on the subject focuses on global measures of metacognition, where an underlying practical theoretical model does not exist. Thus, understanding differences in local metacognition between the genders better allows for theoretical depth compared to only studying global metacognitive measures.

Although this research cannot provide proof of a causal relationship between gender differences in metacognitive monitoring accuracy and navigation performance, the current results do point to the possibility of gender differences in metacognitive control, which could be an important factor in determining gender differences in navigation performance. Improved local metacognition could help to improve performance via more appropriate control behaviors (Nelson & Narens, 1990). Future research is necessary to further determine how local metacognition differs between the genders, how these differences translate to differences in control, and, subsequently, how these differences affect performance. In addition, research on the cognitive differences between the genders related to spatial tasks, such as navigation and metacognitive strategy, could help us to better identify sources for improving women's performance in these skills.

Most previous work employing this framework has used word-pair associates and the generalizability of the model to other tasks was thus unclear. However, the Nelson and Narens (1990) framework was recently shown to be effective in better understand the metacognition of face recognition (Watier & Collin, 2010, 2011, 2012). In addition to this work, the current study shows preliminary evidence that this framework is applicable outside of word-pair associates and can be applied to the metacognition of spatial navigation. Further research is needed to determine how well it applies to other spatial tasks as well as to other cognitive tasks in general. A better understanding of the metacognitive factors underlying navigation performance will provide an important missing component to models of this complex cognitive task.

The current results also confirm that stereotype activation can influence men's and women's navigation performance and trial-by-trial confidence. These conclusions have important theoretical implications that will help to better understand the underlying causes of gender differences in navigation performance. They also have important practical implications related to improving women's involvement in STEM fields, which require mastery of spatial processing. Intriguingly, stereotype threat/lift had no effect on local metacognition. Thus, our results are not compatible with the idea to include local metacognition to the Schmader et al. (2008) process model for stereotype threat. That is, it is not possible to confirm that stereotype threat reduces self-confidence, which limits performance via changes in metacognitive accuracy. Rather, it seems there is either a more direct link between self-confidence and performance, or other factors act as intermediaries between them.

In conclusion, this study supports the literature demonstrating a gender difference favoring men in performance and confidence on an orientation-based navigation task. This study also provides preliminary evidence, using the Nelson and Narens (1990) metamemory framework, of a gender difference, again favoring men, in the local metacognition of spatial navigation. In addition, this study determined that stereotype activation was able to influence men's and women's navigation performance and confidence, but not their local metacognition. Future research is necessary to determine where local metacognition and the effects of stereotype fit into the mechanisms behind the gender difference in spatial navigation performance. The practical applications of this new field of research are to find sociocognitive solutions to improving women's performance in spatial tasks in order to help support women's involvement in STEM fields.

REFERENCES

Achtman, R. L., Green, C. S., & Bavelier, D. (2008). Video games as a tool to train visual skills. *Restorative Neurology and Neuroscience, 26*(4-5), 435-446.

Alban, M. W., & Kelley, C. M. (2013). Embodiment meets metamemory: Weight as a cue for metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(5), 1628-1634. doi:10.1037/a0032420 23565793

Alexander, G. M. (2003). An evolutionary perspective of sex-typed toy preferences: Pink, blue, and the brain. *Arch Sex Behav, 32*(1), 7-14.

Allen, G. L. (2000). Men and women, maps and minds - Cognitive bases of sex-related differences in reading and interpreting maps. *Spatial Cognition: Foundations and Applications, 26*, 3-18.

Allen, G. L., Kirasic, K. C., Dobson, S. H., Long, R. G., & Beck, S. (1996). Predicting environmental learning from spatial abilities: An indirect route. *Intelligence, 22*(3), 327-355. doi:10.1016/S0160-2896%2896%2990026-4

Allen, R. C. (2000). *The effect of restricted field of view on locomotion tasks, head movements, and motion sickness.* (Doctoral Dissertation), University of Central Florida, Retrospective Theses and Dissertations: 1758. Retrieved from http://stars.library.ucf.edu/rtd/1758

Amaral, D., & Lavenex, P. (2007). Hippocampal neuroanatomy. In P. Andersen, R. Morris, D. Amaral, T. Bliss, & J. O'Keefe (Eds.), *The hippocampus book* (pp. 37-114). New York, NY: Oxford University Press.

Andersen, N. E., Dahmani, L., Konishi, K., & Bohbot, V. D. (2012). Eye tracking, strategies, and sex differences in virtual navigation. *Neurobiology of Learning and Memory, 97*(1), 81-89.

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience, 20*, 303-330. doi:10.1146/annurev.neuro.20.1.303

Andreano, J. M., & Cahill, L. (2009). Sex influences on the neurobiology of learning and memory. *Learning & Memory, 16*(4), 248-266. doi:10.1101/lm.918309 19318467

Antonova, E., Parslow, D., Brammer, M., Simmons, A., Williams, S., Dawson, G. R., & Morris, R. (2011). Scopolamine disrupts hippocampal activity during allocentric spatial memory in humans: An fMRI study using a virtual reality analogue of the Morris Water Maze. *Journal of Psychopharmacology, 25*(9), 1256-1265. doi:10.1177/0269881110379285 20823079

Ariel, R., & Moffat, S. D. (2017). Age-related similarities and differences in monitoring spatial cognition. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*, 1-27.

Arnold, A. P., & Chen, X. (2009). What does the "four core genotypes" mouse model tell us about sex differences in the brain and other tissues? *Front Neuroendocrinol, 30*(1), 1-9.

Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature, 543*(7647), 719-722. doi:10.1038/nature21692

Astur, R. S., Ortiz, M. L., & Sutherland, R. J. (1998). A characterization of performance by men and women in a virtual Morris water task: A large and reliable sex difference. *Behavioral Brain Research, 93*(1-2), 185-190. doi:10.1016/S0166-4328%2898%2900019-9 9659999

Astur, R. S., Purton, A. J., Zaniewski, M. J., Cimadevilla, J., & Markus, E. J. (2016). Human sex differences in solving a virtual navigation problem. *Behavioral Brain Research, 308*, 236-243.

Astur, R. S., Tropp, J., Sava, S., Constable, R. T., & Markus, E. J. (2004). Sex differences and correlations in a virtual Morris water task, a virtual radial arm maze, and mental rotation. *Behavioral Brain Research, 151*(1-2), 103-115. doi:10.1016/j.bbr.2003.08.024

Bachevalier, J., & Hagger, C. (1991). Sex differences in the development of learning abilities in primates. *Psychoneuroendocrinology, 16*(1-3), 177-188. doi:10.1016/0306

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn Sci, 4*(11), 417-423.

Baenninger, M., & Newcombe, N. (1989). The Role of Experience in Spatial Test-Performance - a Meta-Analysis. *Sex Roles, 20*(5-6), 327-344. doi:10.1007/Bf00287729

Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science, 60*(2), 434-448. doi:10.1287/mnsc.2013.1776

Barkley, C. L., & Gabriel, K. I. (2007). Sex differences in cue perception in a visual scene: Investigation of cue type. *Behavioral Neuroscience, 121*(2), 291-300. doi:10.1037/0735

Barrett, R. J., & Ray, O. S. (1970). Behavior in the open field, Lashley III maze, shuttle-box, and Sidman avoidance as a function of strain, sex, and age. *Developmental Psychology, 3*(1), 73-77.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory Predictions Are Based on Ease of Processing. *Journal of Memory and Language, 28*(5), 610-632. doi:10.1016/0749

Berg, J. H., Stephan, W. G., & Dodson, M. (1981). Attributional modesty in women. *Psychology of Women Quarterly, 5*(5S), 711-727. doi:10.1111/j.1471-6402.1981.tb01095.x

Beyer, S., & Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin, 23*(2), 157-172. doi:10.1177/0146167297232005

Binnie-Dawson, J. L., & Cheung, Y. M. (1982). The effects of different types of neonatal feminization and environmental stimulation on changes in sex-associated activity/spatial learning skills. *Biological Psychology, 15*(1-2), 109-140. doi:10.1016/0301-0511%2882%2990035-7 7138997

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 417-444. doi:10.1146

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology, 20*(2), 239-263. doi:10.1002/acp.1182

Blanch, D. C., Hall, J. A., Roter, D. L., & Frankel, R. M. (2008). Medical student gender and issues of confidence. *Patient Educ Couns, 72*(3), 374-381. doi:10.1016/j.pec.2008.05.021 18656322

Blanch, R. J., Brennan, D., Condon, B., Santosh, C., & Hadley, D. (2004). Are There Gender-specific Neural Substrates of Route Learning from Different Perspectives? *Cerebral Cortex, 14*(11), 1207-1213. doi:10.1093/cercor/bhh081 15142964

Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology, 23*(5), 638-663. doi:10.1002/acp.1473

Blumberg, F. C., & Fisch, S. M. (2013). Introduction: Digital games as a context for cognitive development, learning, and developmental research. In F. C. Blumberg & S. M. Fisch

(Eds.), *Digital games: A context for cognitive development* (pp. 1-9). San Francisco, CA: Jossey-Bass.

Bohbot, V. D., Lerch, J., Thorndycraft, B., Iaria, G., & Zijdenbos, A. P. (2007). Gray matter differences correlate with spontaneous strategies in a human virtual navigation task. *J Neurosci, 27*(38), 10078-10083. doi:10.1523/JNEUROSCI.1763-07.2007

Boles, D. B. (1980). X-linkage of spatial ability: A critical review. *Child Development, 51*(3), 625-635. doi:10.2307/1129448 7418502

Bosco, A., Longoni, A. M., & Vecchi, T. (2004). Gender effects in spatial orientation: cognitive profiles and mental strategies. *Appl Cogn Psychol, 18*(5), 519-532. doi:10.1002/acp.1000

Bouchard, T. J., Segal, N. L., & Lykken, D. T. (1990). Genetic and environmental influences on special mental abilities in a sample of twins reared apart. *Acta Geneticae Medicae et Gemellologiae: Twin Research, 39*(2), 193-206.

Boutcher, S. H., Fleischer-Curtian, L. A., & Gines, S. D. (1988). The effects of self-presentation on perceived exertion. *Journal of Sport & Exercise Psychology, 10*(3), 270-280.

Brewer, N., Weber, N., & Wells, G. L. (2004). Does a 10-12 second decision latency optimally discriminate correct from incorrect eyewitness identifications? *Australian Journal of Psychology, 56*, 164-164.

Brown, L. N., Lahar, C. J., & Mosley, J. L. (1998). Age and gender-related differences in strategy use for route information - A "map-present" direction-giving paradigm. *Environment and Behavior, 30*(2), 123-143. doi:10.1177/0013916598302001

Brown, R. E. (2016). Hebb and Cattell: The Genesis of the Theory of Fluid and Crystallized Intelligence. *Frontiers in Human Neuroscience, 10*(606), 1-11. doi:10.3389/fnhum.2016.00606

Bryant, D. J. (1991). Visual-Imagery Versus Visual Experience of Familiar Individuals. *Bulletin of the Psychonomic Society, 29*(1), 41-44.

Buchsbaum, M. S., & Henkin, R. I. (1980). Perceptual abnormalities in patients with chromatin negative gonadal dysgenesis and hypogonadotropic hypogonadism. *Int J Neurosci, 11*(3), 201-209.

Buffington, C., Cerf, B., Jones, C., & Weinberg, B. A. (2016). STEM training and early career outcomes of female and male graduate students: Evidence from UMETRICS data linked to the 2010 census. *The American economic review, 106*(5), 333-338.

Buneo, C. A., & Andersen, R. A. (2006). The posterior parietal cortex: Sensorimotor interface for the planning and online control of visually guided movements. *Neuropsychologia, 44*(13), 2594-2606. doi:10.1016/j.neuropsychologia.2005.10.011

Burgess, N. (2008). Spatial cognition and the brain. *Ann N Y Acad Sci, 1124*, 77-97. doi:10.1196/annals.1440.002

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron, 35*(4), 625-641.

Burkitt, J., Widman, D., & Saucier, D. M. (2007). Evidence for the influence of testosterone in the performance of spatial navigation in a virtual water maze in women but not in men. *Horm Behav, 51*(5), 649-654. doi:10.1016/j.yhbeh.2007.03.007 17462646

Burnett, S. A. (1986). Sex-related differences in spatial ability: Are they trivial? *American Psychologist, 41*(9), 1012-1014.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7*(1), 26-48. doi:10.3758/BF03210724 10780019

Butler, T., Imperato-McGinley, J., Pan, H., Voyer, D., Cordero, J., Zhu, Y. S., . . . Silbersweig, D. (2006). Sex differences in mental rotation: top-down versus bottom-up processing. *Neuroimage, 32*(1), 445-456. doi:10.1016/j.neuroimage.2006.03.030

Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental Aspects of the Feeling of Knowing. *Developmental Psychology, 24*(5), 654-663. doi:10.1037/0012-1649.24.5.654

Byrne, R. W. (1979). Memory for Urban-Geography. *Quarterly Journal of Experimental Psychology, 31*(Feb), 147-154. doi:10.1080/14640747908400714

Caballero-Reinaldo, C., Navarro-Francés, C. I., & Arenas, M. C. (2017). Diferencias sexuales en el aprendizaje del Laberinto Hebb-Williams con ratones OF1 Sex differences in OF1 mice's learning of the Hebb-Williams Maze. *Escritos de Psicología, 10*, 1-12.

Caldera, Y. M., Huston, A. C., & O'Brien, M. (1989). Social interactions and play patterns of parents and toddlers with feminine, masculine, and neutral toys. *Child Development, 60*(1), 70-76. doi:10.2307/1131072 2702876

Calisir, F., & Gurel, Z. (2003). Influence of text structure and prior knowledge of the learner on reading comprehension, browsing and perceived control. *Computers in Human Behavior, 19*(2), 135-145. doi:10.1016/S0747-5632%2802%2900058-4

Calton, J. L., & Taube, J. S. (2009). Where am I and how will I get there from here? A role for posterior parietal cortex in the integration of spatial information and route planning. *Neurobiology of Learning and Memory, 91*(2), 186-196. doi:10.1016/j.nlm.2008.09.015

Canovas, R., Espinola, M., Iribarne, L., & Cimadevilla, J. M. (2008). A new virtual task to evaluate human place learning. *Behavioral Brain Research, 190*(1), 112-118. doi:10.1016/j.bbr.2008.02.024 18374426

Caplan, P. J., Crawford, M., Hyde, J. S., & Richardson, J. T. E. (Eds.). (1997). *Gender Differences in Human Cognition*. New York, NY: Oxford University Press.

Caplan, P. J., Macpherson, G. M., & Tobin, P. (1985). Do Sex-Related Differences in Spatial Abilities Exist - a Multilevel Critique with New Data. *American Psychologist, 40*(7), 786-799.

Carassa, A., Geminiani, G., Morganti, F., & Varotto, D. (2002). Active and passive spatial learning in a complex virtual environment: The effect of efficient exploration. *Cognitive Proc, 3*(4), 65-81.

Cashdan, E., & Gaulin, S. J. (2016). Why Go There? Evolution of Mobility and Spatial Cognition in Women and Men : An Introduction to the Special Issue. *Hum Nat, 27*(1), 1-15. doi:10.1007/s12110-015-9253-4

Castel, A. D., McCabe, D. P., & Roediger, H. L., III. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review, 14*(1), 107-111. doi:10.3758/BF03194036 17546739

Ceci, S. J., & Williams, W. M. (2007). *Why aren't more women in science? Top researchers debate the evidence*. Washington, DC: American Psychological Association.

Chabanne, V., Peruch, P., & Thinus-Blanc, C. (2004). Sex differences and women's hormonal cycle effects on spatial performance in a virtual environment navigation task. *Cahiers de Psychologie Cognitive, 22*(3), 351-375.

Chabris, C. F., & Glickman, M. E. (2006). Sex Differences in Intellectual Performance: Analysis of a Large Cohort of Competitive Chess Players. *Psychological Science, 17*(12), 1040-1046. doi:10.1111/j.1467-9280.2006.01828.x 17201785

Chambers, D. W. (1983). Stereotypic images of the scientist: The Draw-a-Scientist Test. *Science Education, 67*(2), 255-265.

Chance, S. S., Gaunet, F., Beall, A. C., & Loomis, J. M. (1998). Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *Presence-Teleoperators and Virtual Environments, 7*(2), 168-178. doi:10.1162/105474698565659

Chang, K. T., & Antes, J. R. (1987). Sex and Cultural-Differences in Map Reading. *American Cartographer, 14*(1), 29-42.

Charleston, S. (2008). The International Wayfinding Strategy Scale: Evidence for cross-cultural use with a sample from the UK. *Perceptual and Motor Skills, 106*(3), 881-882. doi:10.2466/PMS.106.3.881-882 18712210

Chebat, D. R., Maidenbaum, S., & Amedi, A. (2017). *The Transfer of Non-Visual Spatial Knowledge Between Real and Virtual Mazes via Sensory Substitution*. Paper presented at the 7th International Conference on Virtual Rehabilitation, Montreal, CA.

Chen, S. Y., & Ford, N. J. (1998). Modelling user navigation behaviors in a hypermedia-based learning system: An individual differences approach. *Knowledge Organization, 25*(3), 67-78.

Chrastil, E. R. (2013). Neural evidence supports a novel framework for spatial navigation. *Psychon Bull Rev, 20*(2), 208-227. doi:10.3758/s13423-012-0351-6

Chrisler, J. C., & McCreary, D. R. (2010). *Handbook of gender research in psychology*. New York, NY: Springer.

Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009). Neural correlates of metamemory: a comparison of feeling-of-knowing and retrospective confidence judgments. *Journal of Cognitive Neuroscience, 21*(9), 1751-1765.

Classen, S., Bewernitz, M., & Shechtman, O. (2011). Driving simulator sickness: an evidence-based review of the literature. *Am J Occup Ther, 65*(2), 179-188.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience, 22*, 319-349. doi:10.1146/annurev.neuro.22.1.319

Coluccia, E., Bosco, A., & Brandimonte, M. A. (2007). The role of visuo-spatial working memory in map learning: new findings from a map drawing paradigm. *Psychological Research-Psychologische Forschung, 71*(3), 359-372. doi:10.1007/s00426-006-0090-2

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology, 24*(3), 329-340. doi:10.1016/j.jenvp.2004.08.006

Constantinidis, C., & Procyk, E. (2004). The primate working memory networks. *Cogn Affect Behav Neurosci, 4*(4), 444-465.

Constantinidis, C., & Steinmetz, M. A. (1996). Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *J Neurophysiol, 76*(2), 1352-1355.

Constantinidis, C., & Wang, X. J. (2004). A neural circuit basis for spatial working memory. *Neuroscientist, 10*(6), 553-565. doi:10.1177/1073858404268742

Cooke-Simpson, A., & Voyer, D. (2007). Confidence and gender differences on the mental rotations test. *Learning and Individual Differences, 17*(2), 181-186. doi:10.1016/j.lindif.2007.03.009

Cooper, J. (2006). The digital divide: The special case of gender. *Journal of Computer Assisted Learning, 22*(5), 320-334.

Corley, R. P., Defries, J. C., Kuse, A. R., & Vandenberg, S. G. (1980). Familial Resemblance for the Identical Blocks Test of Spatial Ability - No Evidence for X-Linkage. *Behav Genet, 10*(2), 211-215. doi:10.1007/Bf01066271

Coutinho, S. A., & Neuman, G. (2008). A model of metacognition, achievement goal orientation, learning style and self-efficacy. *Learning Environments Research, 11*(2), 131-151. doi:10.1007/s10984-008-9042-7

Crawford, M., Herrmann, D. J., Holdsworth, M. J., Randall, E. P., & Robbins, D. (1989). Gender and Beliefs About Memory. *British Journal of Psychology, 80*, 391-401.

Cross, C. P., Brown, G. R., Morgan, T. J., & Laland, K. N. (2017). Sex differences in confidence influence patterns of conformity. *British Journal of Psychology, 108*(4), 655-667.

Curtis, C. E. (2006). Prefrontal and parietal contributions to spatial working memory. *Neuroscience, 139*(1), 173-180. doi:10.1016/j.neuroscience.2005.04.070

D'Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., & Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Brain Res Cogn Brain Res, 7*(1), 1-13.

Dabbs, J. M., Chang, E. L., Strong, R. A., & Milun, R. (1998). Spatial ability, navigation strategy, and geographic knowledge among men and women. *Evolution and Human Behavior, 19*(2), 89-98. doi:10.1016/S1090-5138(97)00107-4

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*(1), 10-32. doi:10.1111/j.1467-8535.2009.01038.x

Darken, R. P., & Banker, W. P. (1998). Navigating in natural environments: A virtual environment training transfer study. *Proceedings of the Virtual Reality Annual International Symposium*, 12-19. doi:10.1109/Vrais.1998.658417

de Bruin, A. B., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction, 22*(4), 245-252.

de Carvalho, M. K., & Yuzawa, M. (2001). The effect of social influences and general metacognitive knowledge on metamemory judgments. *Contemporary Educational Psychology, 26*(4), 571-587. doi:10.1006/ceps.2000.1077

De Goede, M., & Postma, A. (2008). Gender differences in memory for objects and their locations: A study on automatic versus controlled encoding and retrieval contexts. *Brain and Cognition, 66*(3), 232-242. doi:10.1016/j.bandc.2007.08.004 17950509

Dede, C., Ketelhut, D., & Ruess, K. (2002). *Motivation, usability, and learning outcomes in a prototype museum-based multi-user virtual environment.* Paper presented at the 5th International Conference of the Learning Sciences, Scattle, Washington.

DeFries, J. C. (1980). Mental Abilities. In J. H. Miekle & M. H. Crawford (Eds.), *Current developments in anthropological genetics* (pp. 397-417). Boston, MA: Springer.

DeFries, J. C., Ashton, G. C., Johnson, R. C., Kuse, A. R., McClearn, G. E., Mi, M. P., . . . Wilson, J. R. (1976). Parent-offspring resemblance for specific cognitive abilities in two ethnic groups. *Nature, 261*(5556), 131-133. doi:10.1038/261131a0 1272380

Denissen, A. M. (2010). The right tools for the job: Constructing gender meanings and identities in the male-dominated building trades. *Human Relations, 63*(7), 1051-1069. doi:10.1177/0018726709349922

Derks, B., Scheepers, D., Van Laar, C., & Ellemers, N. (2011). The threat vs. Challenge of car parking for women: How self- and group affirmation affect cardiovascular responses. *Journal of Experimental Social Psychology, 47*(1), 178-183. doi:10.1016/j.jesp.2010.08.016

Descartes, R. (1637). *Discourse on the method of rightly conducting one's reason and seeking truth in the sciences*. Leiden, Netherlands: Ian Maire Press.

Devlin, A. S., & Bernstein, J. (1997). Interactive way-finding: Map style and effectiveness. *Journal of Environmental Psychology, 17*(2), 99-110. doi:10.1006/jevp.1997.0045

Diana, R. A., & Reder, L. M. (2004). Visual versus Verbal Metacognition: Are They Really Different? In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 187-201). Cambridge, MA: MIT Press.

Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory & Cognition, 33*(6), 1096-1115. doi:10.3758/BF03193216 16496729

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences, 47*, 103-116. doi:10.1016/j.lindif.2015.12.018

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 103-140). San Diego, CA: Elsevier Academic Press.

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychol Aging, 13*(4), 597-607. doi:D10.1037//0882-7974.13.4.597

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology-Learning Memory and Cognition, 27*(5), 1180-1191. doi:10.1037/0278-7393.27.5.1180

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition: A textbook of cognition, educational, life span, and applied psychology*. Thousand Oaks, CA: Sage.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271-280. doi:10.1016/j.learninstruc.2011.08.003

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science, 18*(6), 321-326. doi:10.1111/j.1467-8721.2009.01660.x 20485453

Eccles, J. S., & Harold, R. D. (1991). Gender differences in sport involvement: Applying the Eccles' expectancy-value model. *Journal of Applied Sport Psychology, 3*(1), 7-35. doi:10.1080/10413209108406432

Ecuyer-Dab, I., & Robert, M. (2004). Spatial Ability and Home-Range Size: Examining the Relationship in Western Men and Women (Homo sapiens). *Journal of Comparative Psychology, 118*(2), 217-231. doi:10.1037/0735-7036.118.2.217 15250809

Efklides, A., & Tsiora, A. (2002). Metacognitive experiences, self-concept, and self-regulation. *Psychologia, 45*(4), 222-236. doi:10.2117/psysoc.2002.222

Eichenbaum, H. (2004). Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron, 44*(1), 109-120. doi:10.1016/j.neuron.2004.08.028

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature, 425*(6954), 184-188. doi:10.1038/nature01964

Erkut, S. (1983). Exploring sex differences in expectancy, attribution, and academic achievement. *Sex Roles, 9*(2), 217-231. doi:10.1007/BF00289625

Estes, Z., & Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. *Arch Sex Behav, 41*(3), 550-570. doi:10.1007/s10508-011-9875-5 22130691

Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology, 20*(4), 439-462. doi:10.1016/0010-0285%2888%2990012-6 3191667

Feingold, A. (1998). Gender stereotyping for sociability, dominance, character, and mental health: a meta-analysis of findings from the bogus stranger paradigm. *Genet Soc Gen Psychol Monogr, 124*(3), 253-270.

Fennema, E. (1975). Mathematics Learning: What Research Says about Sex Differences. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Fields, A. W., & Shelton, A. L. (2006). Individual skill differences and large-scale environmental learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 506-515. doi:10.1037/0278-7393.32.3.506 16719662

Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911. doi:10.1037/0003-066x.34.10.906

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53*(1), 25-44.

Foo, P., Duchon, A., Warren, W. H., Jr., & Tarr, M. J. (2007). Humans do not switch between path knowledge and landmarks when learning a new environment. *Psychological Research, 71*(3), 240-251. doi:10.1007/s00426-006-0080-4 16957956

Fralick, B., Kearn, J., Thompson, S., & Lyons, J. (2009). How Middle Schoolers Draw Engineers and Scientists. *Journal of Science Education and Technology, 18*(1), 60-73.

Fralley, J. S., Eliot, J., & Dayton, C. M. (1978). Further study of the X-linked recessive gene hypothesis for inheritance of spatial abilities. *Perceptual and Motor Skills, 47*(2), 1023-1029. doi:10.2466/pms.1978.47.3f.1023

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nat Rev Neurosci, 6*(2), 119-130. doi:10.1038/nrn1607

Friedman, T. (2007). *The world is flat 3.0: A brief history of the twenty-first century* (3rd. ed.). New York, NY: Picador.

Fuhs, M. C., & Touretzky, D. S. (2006). A spin glass model of path integration in rat medial entorhinal cortex. *J Neurosci, 26*(16), 4266-4276. doi:10.1523/JNEUROSCI.4353-05.2006

Furtak, S. C., Wei, S.-M., Agster, K. L., & Burwell, R. D. (2007). Functional neuroanatomy of the parahippocampal region in the rat: The perirhinal and postrhinal cortices. *Hippocampus, 17*(9), 709-722. doi:10.1002/hipo.20314 17604355

Gagnon, K. T., Cashdan, E. A., Stefanucci, J. K., & Creem-Regehr, S. H. (2016). Sex Differences in Exploration Behavior and the Relationship to Harm Avoidance. *Hum Nat, 27*(1), 82-97. doi:10.1007/s12110-015-9248-1

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development, 85*(1), 250-263. doi:10.1111/cdev.12128 23713580

Galea, L. A., & Kimura, D. (1993). Sex differences in route-learning. *Personality and Individual Differences, 14*(1), 53-65. doi:10.1016/0191-8869%2893%2990174-2

Gallistel, C. R. (1989). Animal cognition: the representation of space, time and number. *Annual Review of Psychology, 40*, 155-189. doi:10.1146/annurev.ps.40.020189.001103

Gallistel, C. R. (1990). Representations in Animal Cognition - an Introduction. *Cognition, 37*(1-2), 1-22. doi:10.1016/0010-0277(90)90016-D

Ganley, C. M., & Vasilyeva, M. (2014). The role of anxiety and working memory in gender differences in mathematics. *Journal of Educational Psychology, 106*(1), 105-120. doi:10.1037/a0034099

Gardner, M. R., Sorhus, I., Edmonds, C. J., & Potts, R. (2012). Sex differences in components of imagined perspective transformation. *Acta Psychologica, 140*(1), 1-6. doi:10.1016/j.actpsy.2012.02.002 22426425

Gaulin, S. J. C., & Fitzgerald, R. W. (1986). Sex-Differences in Spatial Ability - an Evolutionary Hypothesis and Test. *American Naturalist, 127*(1), 74-88. doi:10.1086/284468

Gilmartin, P. P. (1986). Maps, Mental-Imagery, and Gender in the Recall of Geographical Information. *American Cartographer, 13*(4), 335-344.

Gobet, F. (1998). Expert memory: A comparison of four theories. *Cognition, 66*(2), 115-152. doi:10.1016/S0010-0277%2898%2900020-1 9677761

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition, 18*(5), 546-550. doi:10.3758/BF03198487 2233267

Golledge, R. G., & Spector, A. N. (1978). Comprehending Urban-Environment - Theory and Practice. *Geographical Analysis, 10*(4), 403-426.

Gron, G., Wunderlich, A. P., Spitzer, M., Tomczak, R., & Riepe, M. W. (2000). Brain activation during human navigation: Gender-different neural networks as substrate of performance. *Nature Neuroscience, 3*(4), 404-408. doi:10.1038/73980 10725932

Gur, R. C., Alsop, D., Glahn, D., Petty, R., Swanson, C. L., Maldjian, J. A., . . . Gur, R. E. (2000). An fMRI study of sex differences in regional activation to a verbal and a spatial task. *Brain and Language, 74*(2), 157-170. doi:10.1006/brln.2000.2325 10950912

Guttman, R., & Shoham, I. (1979). Intrafamilial invariance and parent-offspring resemblance in spatial abilities. *Behav Genet, 9*(5), 367-378. doi:10.1007/BF01066975 539965

Hacker, D. J. (1998). Definitions and empirical foundations. *Metacognition in educational theory and practice, 14*, 1-23.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*(1), 160-170. doi:10.1037/0022-0663.92.1.160

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature, 436*(7052), 801-806. doi:10.1038/nature03721

Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing… or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly, 40*(3), 353-363.

Hallam, S. (2001). The development of metacognition in musicians: Implications for education. *British Journal of Music Education, 18*(1), 27-39. doi:10.1017/S0265051701000122

Halpern, D. F. (1986). A different answer to the question," Do sex-related differences in spatial abilities exist?". *American Psychologist, 41*(9), 1014-1015.

Halpern, D. F. (2013). *Sex differences in cognitive abilities* (4th ed.). New York, NY: Psychology Press.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological science in the public interest, 8*(1), 1-51.

Hardies, K., Breesch, D., & Branson, J. (2013). Gender differences in overconfidence and risk taking: Do self-selection and socialization matter? *Economics Letters, 118*(3), 442-444. doi:10.1016/j.econlet.2012.12.004

Harrell, W. A., Bowlby, J. W., & Hall-Hoffarth, D. (2000). Directing wayfinders with maps: The effects of gender, age, route complexity, and familiarity with the environment. *J Soc Psychol, 140*(2), 169-178. doi:10.1080/00224540009600456 10808639

Hartley, T., Trinkler, I., & Burgess, N. (2004). Geometric determinants of human spatial memory. *Cognition, 94*(1), 39-75. doi:10.1016/j.cognition.2003.12.001

Hausmann, M., Schoofs, D., Rosenthal, H. E. S., & Jordan, K. (2009). Interactive effects of sex hormones and gender stereotypes on cognitive sex differences-A psychobiosocial approach. *Psychoneuroendocrinology, 34*(3), 389-401. doi:10.1016/j.psyneuen.2008.09.019

Head, D., & Isom, M. (2010). Age effects on wayfinding and route learning skills. *Behav Brain Res, 209*(1), 49-58. doi:10.1016/j.bbr.2010.01.012

Hebb, D. O., & Williams, K. (1946). A method of rating animal intelligence. *Journal of General Psychology, 34*, 59-65. doi:10.1080/00221309.1946.10544520 21015350

Hedge, C., Weaver, R., & Schnall, S. (2017). Spatial Learning and Wayfinding in an Immersive Environment: The Digital Fulldome. *Cyberpsychology, Behavior, and Social Networking, 20*(5), 327-333.

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence, 34*(2), 151-176. doi:10.1016/j.intell.2005.09.005

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30*(5), 425-447.

Heil, M., & Jansen-Osmann, P. (2008). Gender differences in math and mental rotation accuracy but not in mental rotation speed in 8-years-old children. *European Journal of Developmental Science, 2*(1-2), 190-196.

Heil, M., Jansen, P., Quaiser-Pohl, C., & Neuburger, S. (2012). Gender-specific effects of artificially induced gender beliefs in mental rotation. *Learning and Individual Differences, 22*(3), 350-353. doi:10.1016/j.lindif.2012.01.004

Henderson, N. D. (1982). Human behavior genetics. *Annual Review of Psychology*, 403-440. doi:10.1146/annurev.ps.33.020182.002155 7059138

Henrie, R. L., Aron, R. H., Nelson, B. D., & Poole, D. A. (1997). Gender-related knowledge variations within geography. *Sex Roles, 36*(9-10), 605-623. doi:10.1023/A:1025618024694

Herman, J. F., Heins, J. A., & Cohen, D. S. (1987). Children's spatial knowledge of their neighborhood environment. *Journal of Applied Developmental Psychology, 8*(1), 1-15. doi:10.1016/0193-3973%2887%2990017-7

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Metamemory in adulthood: differentiating knowledge, belief, and behavior. In T. M. Hess (Ed.), *Aging and cognition: knowledge organization and utilization* (pp. 161-212). Oxford, England: Elsevier.

Hill, C., Corbett, C., & St Rose, A. (2010). Why so few? Women in science, technology, engineering, and mathematics (pp. 134). Washington, DC: AAUW.

Hines, M. (2007). Do Sex Differences in Cognition Cause the Shortage of Women in Science? In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science?: Top researchers debate the evidence* (pp. 101-112). Washington, DC: American Psychological Association.

Hirnstein, M., Andrews, L. C., & Hausmann, M. (2014). Gender-stereotyping and cognitive sex differences in mixed- and same-sex groups. *Arch Sex Behav, 43*(8), 1663-1673. doi:10.1007/s10508-014-0311-5 24923876

Hirnstein, M., Freund, N., & Hausmann, M. (2012). Gender stereotyping enhances verbal fluency performance in men (and women). *Zeitschrift fur Psychologie, 220*(2), 70-77. doi:10.1027/2151-2604/a000098

Hiscock, M. (1986). On Sex-Differences in Spatial Abilities. *American Psychologist, 41*(9), 1011-1012.

Hoang, N., Mora, P., Vieites, V., Reeb-Sutherland, B., & Pruden, S. (2017). *The Role of Wayfinding Experience on Spatial Anxiety: Implications for Spatial Aptitude.* Paper presented at the Conference for Undergraduate Research, Florida International University.

Holding, C. S., & Holding, D. H. (1989). Acquisition of Route Network Knowledge by Males and Females. *Journal of General Psychology, 116*(1), 29-41.

Hugdahl, K., Thomsen, T., & Ersland, L. (2006). Sex differences in visuo-spatial processing: An fMRI study of mental rotation. *Neuropsychologia, 44*(9), 1575-1583. doi:10.1016/j.neuropsychologia.2006.01.026 16678867

Hunt, E., & Waller, D. (1999). Orientation and wayfinding: A review. Arlington, VA: ONR Technical Report N00014-96-0380: Office of Naval Research.

Hyde, J. S. (1981). How Large Are Cognitive Gender Differences - a Meta-Analysis Using Omega-2 and D. *American Psychologist, 36*(8), 892-901. doi:10.1037/0003-066x.36.8.892

Hyde, J. S. (2005). The gender similarities hypothesis. *Am Psychol, 60*(6), 581-592. doi:10.1037/0003-066X.60.6.581

Iachini, T., Ruggiero, G., Conson, M., & Trojano, L. (2009). Lateralization of egocentric and allocentric spatial processing after parietal brain lesions. *Brain Cogn, 69*(3), 514-520. doi:10.1016/j.bandc.2008.11.001

Iaria, G., Petrides, M., Dagher, A., Pike, B., & Bohbot, V. D. (2003). Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *J Neurosci, 23*(13), 5945-5952.

Imperato-McGinley, J., Pichardo, M., Gautier, T., Voyer, D., & Bryden, M. P. (1991). Cognitive-Abilities in Androgen-Insensitive Subjects - Comparison with Control Males and Females from the Same Kindred. *Clinical Endocrinology, 34*(5), 341-347. doi:10.1111/j.1365-2265.1991.tb00303.x

Inal, Y., & Cagiltay, K. (2007). Flow experiences of children in an interactive social game environment. *British Journal of Educational Technology, 38*(3), 455-464.

Ishikawa, T., & Montello, D. R. (2006). Spatial knowledge acquisition from direct experience in the environment: individual differences in the development of metric knowledge and the integration of separately learned places. *Cogn Psychol, 52*(2), 93-129. doi:10.1016/j.cogpsych.2005.08.003

Jacobs, J., Kahana, M. J., Ekstrom, A. D., Mollison, M. V., & Fried, I. (2010). A sense of direction in human entorhinal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 107*(14), 6487-6492. doi:10.1073/pnas.0911213107

Jacobs, J., Weidemann, C. T., Miller, J. F., Solway, A., Burke, J. F., Wei, X. X., . . . Kahana, M. J. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. *Nat Neurosci, 16*(9), 1188-1190. doi:10.1038/nn.3466

Jacobs, L. F., & Schenk, F. (2003). Unpacking the cognitive map: the parallel map theory of hippocampal function. *Psychol Rev, 110*(2), 285-315.

Jahoda, G. (1980). Sex and ethnic differences on a spatial-perceptual task: Some hypotheses tested. *British Journal of Psychology, 71*(3), 425-431. doi:10.1111/j.2044-8295.1980.tb01757.x 7407445

Jakobsson, N. (2012). Gender and confidence: are women underconfident? *Applied Economics Letters, 19*(11), 1057-1059. doi:10.1080/13504851.2011.613745

James, W. (1890). The consciousness of self *The principles of psychology*. New York: Henry Holt and Company.

Jansen-Osmann, P., Schmid, J., & Heil, M. (2007). Spatial knowledge of adults and children in a virtual environment: The role of environmental structure. *European Journal of Developmental Psychology, 4*(3), 251-272.

Janzen, G., & Weststeijn, C. G. (2007). Neural representation of object location and route direction: an event-related fMRI study. *Brain Research, 1165*, 116-125.

Jayet Bray, L. C., Quoy, M., Harris, F. C., & Goodman, P. H. (2010). A circuit-level model of hippocampal place field dynamics modulated by entorhinal grid and suppression-generating cells. *Front Neural Circuits, 4*(122), 1-12. doi:10.3389/fncir.2010.00122

Jiang, Y. X., & Kleitman, S. (2015). Metacognition and motivation: Links between confidence, self-protection and self-enhancement. *Learning and Individual Differences, 37*, 222-230. doi:10.1016/j.lindif.2014.11.025

Jones, L., & Jones, L. (1989). Context, confidence and the able girl. *Educational Res, 31*(3), 189-194.

Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., & Mintun, M. A. (1993). Spatial Working-Memory in Humans as Revealed by Pet. *Nature, 363*(6430), 623-625. doi:10.1038/363623a0

Jonsson, F. U., & Olsson, M. J. (2003). Olfactory metacognition. *Chem Senses, 28*(7), 651-658. doi:10.1093/chemse/bjg058

Joseph, R., Hess, S., & Birecree, E. (1978). Effects of Hormone Manipulations and Exploration on Sex-Differences in Maze-Learning. *Behavioral Biology, 24*(3), 364-377. doi:10.1016/S0091-6773(79)90223-2

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review, 107*(2), 384-396. doi:10.1037/0033-295X.107.2.384 10789203

Kaiser, A. (2012). Re-conceptualizing "sex" and "gender" in the human brain. *Zeitschrift fur Psychologie, 220*(2), 130-136. doi:10.1027/2151-2604/a000104

Kaiser, S., Walther, S., Nennig, E., Kronmuller, K., Mundt, C., Weisbrod, M., . . . Vogeley, K. (2008). Gender-specific strategy use and neural correlates in a spatial perspective taking task. *Neuropsychologia, 46*(10), 2524-2531. doi:10.1016/j.neuropsychologia.2008.04.013 18514745

Kelemen, W. L., Frost, P. J., & Weaver, C. A., III. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition, 28*(1), 92-107. doi:10.3758/BF03211579 10714142

Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: spatial ability's unique role. *Psychological Science, 24*(9), 1831-1836.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering Mistaken for Knowing - Ease of Retrieval as a Basis for Confidence in Answers to General Knowledge Questions. *Journal of Memory and Language, 32*(1), 1-24. doi:10.1006/jmla.1993.1001

Kelly, D. M., & Bischof, W. F. (2005). Reorienting in images of a three-dimensional environment. *J Exp Psychol Hum Percept Perform, 31*(6), 1391-1403. doi:10.1037/0096-1523.31.6.1391

Kennedy, R., Lanham, D., Massey, C., Drexler, J., & Lilienthal, M. (1995). Gender differences in simulator sickness incidence: Implications for military virtual reality systems. *Safe Journal, 25*(1), 69-76.

Keren, G. (1991). Calibration and Probability Judgments - Conceptual and Methodological Issues. *Acta Psychologica, 77*(3), 217-273. doi:10.1016/0001-6918(91)90036-Y

Kessels, R. P., d'Alfonso, A. A., Postma, A., & de Haan, E. H. (2000). Spatial working memory performance after high-frequency repetitive transcranial magnetic stimulation of the left and right posterior parietal cortex in humans. *Neurosci Lett, 287*(1), 68-70.

Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., & Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychological Science, 9*(4), 293-298. doi:10.1111/1467-9280.00058

Klatzky, R. L., Loomis, J. M., Golledge, R. G., Cicinelli, J. G., Doherty, S., & Pellegrino, J. W. (1990). Acquisition of route and survey knowledge in the absence of vision. *J Mot Behav, 22*(1), 19-43. doi:10.1080/00222895.1990.10735500 15111279

Kleitman, S., & Gibson, J. (2011). Metacognitive beliefs, self-confidence and primary learning environment of sixth grade students. *Learning and Individual Differences, 21*(6), 728-735. doi:10.1016/j.lindif.2011.08.003

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences, 17*(2), 161-173. doi:10.1016/j.lindif.2007.03.004

Kluwe, R. H. (1982). Cognitive knowledge and executive control: Metacognition. In D. R. Griffin (Ed.), *Animal mind—human mind* (Vol. 21, pp. 201-224). Berlin, Germany: Springer.

Koch, G., Oliveri, M., Torriero, S., Carlesimo, G. A., Turriziani, P., & Caltagirone, C. (2005). rTMS evidence of different delay and decision processes in a fronto-parietal neuronal network activated during spatial working memory. *Neuroimage, 24*(1), 34-39. doi:10.1016/j.neuroimage.2004.09.042

Konishi, K., & Bohbot, V. D. (2013). Spatial navigational strategies correlate with gray matter in the hippocampus of healthy older adults tested in a virtual maze. *Front Aging Neurosci, 5*(1), 1-8. doi:10.3389/fnagi.2013.00001

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*(3), 311-333. doi:10.1037/0096-3445.124.3.311

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology-General, 126*(4), 349-370. doi:10.1037//0096-3445.126.4.349

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and cognition, 9*(2), 149-171. doi:10.1006/ccog./2000.0433

Koriat, A. (2006). Are we frightened because we run away? Some evidence from metacognitive feelings. In B. Uttl, N. Ohta, & A. L. Siegenthaler (Eds.), *Memory and emotion: Interdisciplinary perspectives* (pp. 83-103). Oxford, UK: Blackwell Publishing.

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289-325). New York, NY: Cambridge University Press.

Koriat, A. (2012). The relationships between monitoring, regulation and performance. *Learning and Instruction, 22*(4), 296-298. doi:10.1016/j.learninstruc.2012.01.002

Koriat, A., Ackerman, R., Adiv, S., Lockl, K., & Schneider, W. (2014). The effects of goal-driven and data-driven regulation on metacognitive monitoring during learning: A developmental perspective. *Journal of Experimental Psychology: General, 143*(1), 386-403. doi:10.1037/a0031768 23421442

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490-517. doi:10.1037/0033-295x.103.3.490

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 483-502). New York: Guilford Press.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for Confidence. *Journal of Experimental Psychology-Human Learning and Memory, 6*(2), 107-118. doi:10.1037//0278-7393.6.2.107

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*(1), 36-69. doi:10.1037/0096-3445.135.1.36 16478315

Koriat, A., Ma'ayan, H., Sheffer, L., & Bjork, R. A. (2006). Exploring a mnemonic debiasing account of the underconfidence-with-practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 595-608. doi:10.1037/0278-7393.32.3.595 16719669

Koriat, A., Pearlman-Avnion, S., & Ben-Zur, H. (1998). The subjective organization of input and output events in memory. *Psychological Research, 61*(4), 295-307. doi:10.1007/s004260050034 9870297

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal* *of Experimental Psychology-General, 131*(2), 147-162. doi:10.1037//0096-3445.131.2.147

Koscik, T., O'Leary, D., Moser, D. J., Andreasen, N. C., & Nopoulos, P. (2009). Sex differences in parietal lobe morphology: Relationship to mental rotation performance. *Brain and Cognition, 69*(3), 451-459. doi:10.1016/j.bandc.2008.09.004 18980790

Kröner, S., & Biermann, A. (2007). The relationship between confidence and self-concept—Towards a model of response confidence. *Intelligence, 35*(6), 580-590.

Langan, A. M., Shuker, D. M., Cullen, W. R., Penney, D., Preziosi, R. F., & Wheater, C. P. (2008). Relationships between student characteristics and self-,peer and tutor evaluations of oral presentations. *Assessment & Evaluation in Higher Education, 33*(2), 179-190. doi:10.1080/02602930701292498

LaNoue, J. B., & Curtis, R. C. (1985). Improving women's performance in mixed-sex situations by effort attributions. *Psychology of Women Quarterly, 9*(3), 337-356. doi:10.1111/j.1471-6402.1985.tb00885.x

Lauer, J. E., Udelson, H. B., Jeon, S. O., & Lourenco, S. F. (2015). An early sex difference in the relation between mental rotation and object preference. *Front Psychol, 6*(558), 1-8. doi:10.3389/fpsyg.2015.00558

Lavenex, P. B., & Lavenex, P. (2010). Spatial relational learning and memory abilities do not differ between men and women in a real-world, open-field environment. *Behavioral Brain Research, 207*(1), 125-137. doi:10.1016/j.bbr.2009.09.046

Lawton, C. A. (1994). Gender Differences in Way-Finding Strategies - Relationship to Spatial Ability and Spatial Anxiety. *Sex Roles, 30*(11-12), 765-779. doi:10.1007/Bf01544230

Lawton, C. A. (1996). Strategies for indoor wayfinding: The role of orientation. *Journal of Environmental Psychology, 16*(2), 137-145. doi:10.1006/jevp.1996.0011

Lawton, C. A. (2001). Gender and regional differences in spatial referents used in direction giving. *Sex Roles, 44*(5-6), 321-337. doi:10.1023/A:1010981616842

Lawton, C. A., Charleston, S. I., & Zieles, A. S. (1996). Individual- and gender-related differences in indoor wayfinding. *Environment and Behavior, 28*(2), 204-219. doi:10.1177/0013916596282003

Lawton, C. A., & Kallai, J. (2002). Gender differences in wayfinding strategies and anxiety about wayfinding: A cross-cultural comparison. *Sex Roles, 47*(9-10), 389-401. doi:10.1023/A:1021668724970

Lawton, C. A., & Morrin, K. A. (1999). Gender differences in pointing accuracy in computer-simulated 3D mazes. *Sex Roles, 40*(1-2), 73-92. doi:10.1023/A:1018830401088

Leibowitz, H. W., & Post, R. B. (1982). The two modes of processing concept and some implications. In J. Beck (Ed.), *Organization and representation in perception* (pp. 343-363). Mahwah, NJ: Lawrence Erlbaum Associates.

Lejbak, L., Vrbancic, M., & Crossley, M. (2009). The female advantage in object location memory is robust to verbalizability and mode of presentation of test stimuli. *Brain and Cognition, 69*(1), 148-153. doi:10.1016/j.bandc.2008.06.006 18722700

Leonesio, R. J., & Nelson, T. O. (1990). Do Different Metamemory Judgments Tap the Same Underlying Aspects of Memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 16*(3), 464-470. doi:10.1037//0278-7393.16.3.464

Levin, D. T. (2004). *Thinking and seeing: Visual metacognition in adults and children*. Cambridge, MA: MIT Press.

Levine, F. M., & de Simone, L. L. (1991). The effects of experimenter gender on pain report in male and female subjects. *Pain, 44*(1), 69-72. doi:10.1016/0304-3959%2891%2990149-R 2038491

Levine, R., Gillman, M.-J., & Reis, H. (1982). Individual differences for sex differences in achievement attributions? *Sex Roles, 8*(4), 455-466. doi:10.1007/BF00287284

Levine, S. C., Foley, A., Lourenco, S., Ehrlich, S., & Ratliff, K. (2016). Sex differences in spatial cognition: Advancing the conversation. *Wiley Interdisciplinary Reviews: Cognitive Science, 7*(2), 127-155. doi:10.1002/wcs.1380

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: a predictor of preschoolers' spatial transformation skill. *Developmental Psychology, 48*(2), 530-542.

Levy, L. J., Astur, R. S., & Frick, K. M. (2005). Men and Women Differ in Object Memory but Not Performance of a Virtual Radial Maze. *Behavioral Neuroscience, 119*(4), 853-862. doi:10.1037/0735-7044.119.4.853 16187814

Lichtenstein, S., Fischhoff, B., & Phillips, L. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 306-334). Cambridge, UK: Cambridge University Press.

Lieberman, D. A. (2004). *Learning and Memory: An Integrative Approach*. Belmont, CA: Thomson/Wadsworth.

Lin, J. J. W., Duh, H. B. L., Parker, D. E., Abi-Rached, H., & Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings IEEE Virtual Reality*, 164-171. doi:10.1109/Vr.2002.996519

Linden, D. E. (2007). The working memory networks of the human brain. *Neuroscientist, 13*(3), 257-267. doi:10.1177/1073858406298480

Linn, M. C., & Petersen, A. C. (1985). Emergence and Characterization of Sex-Differences in Spatial Ability - a Meta-Analysis. *Child Development, 56*(6), 1479-1498. doi:10.1111/j.1467-8624.1985.tb00213.x

Lloyd, R. (1989). Cognitive Maps - Encoding and Decoding Information. *Annals of the Association of American Geographers, 79*(1), 101-124. doi:10.1111/j.1467-8306.1989.tb00253.x

Lucas, K., & Sherry, J. L. (2004). Sex Differences in Video Game Play: A Communication-Based Explanation. *Communication Research, 31*(5), 499-523. doi:10.1177/0093650204267930

Lundeberg, M. A., Fox, P. W., & Punccohar, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology, 86*(1), 114-121. doi:10.1037/0022-0663.86.1.114

Lynn, R. (1992). Sex differences on the Differential Aptitude Test in British and American adolescents. *Educational Psychology, 12*(2), 101-102. doi:10.1080/0144341920120201

Maccoby, E. E., & Jacklin, C. N. (1974). Myth, Reality and Shades of Gray - What We Know and Dont Know About Sex Differences. *Psychology Today, 8*(7), 109-112.

MacFadden, A., Elias, L., & Saucier, D. (2003). Males and females scan maps similarly, but give directions differently. *Brain and Cognition, 53*(2), 297-300. doi:10.1016/S0278-2626(03)00130-1

MacLin, O. H. (1999). *Metacognitive aspects of recognizing famous faces: the role of partial information and familiarity.* (Doctoral Dissertation), University of Nevada.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America, 97*(8), 4398-4403. doi:10.1073/pnas.070039597

Maguire, E. A., Nannery, R., & Spiers, H. J. (2006). Navigation around London by a taxi driver with bilateral hippocampal lesions. *Brain: A Journal of Neurology, 129*(11), 2894-2907. doi:10.1093/brain/awl286

Maguire, E. A., Woollett, K., & Spiers, H. J. (2006). London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. *Hippocampus, 16*(12), 1091-1101. doi:10.1002/hipo.20233

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of Text Material. *Journal of Experimental Psychology-Learning Memory and Cognition, 10*(4), 663-679. doi:10.1037//0278-7393.10.4.663

Malinowski, J. C., & Gillespie, W. T. (2001). Individual differences in performance on a large-scale, real-world wayfinding task. *Journal of Environmental Psychology, 21*(1), 73-82. doi:10.1006/jevp.2000.0183

Mann, V. A., Sasanuma, S., Sakuma, N., & Masaki, S. (1990). Sex differences in cognitive abilities: A cross-cultural perspective. *Neuropsychologia, 28*(10), 1063-1077. doi:10.1016/0028-3932%2890%2990141-A 2267058

Martens, J., & Antonenko, P. D. (2012). Narrowing gender-based performance gaps in virtual environment navigation. *Computers in Human Behavior, 28*(3), 809-819. doi:10.1016/j.chb.2012.01.008

Martin, S. J., & Clark, R. E. (2007). The rodent hippocampus and spatial memory: from synapses to systems. *Cellular and Molecular Life Sciences, 64*(4), 401-431. doi:10.1007/s00018-007-6336-3

Martini, R., & Shore, B. M. (2008). Pointing to parallels in ability-related differences in the use of metacognition in academic and psychomotor tasks. *Learning and Individual Differences, 18*(2), 237-247. doi:10.1016/j.lindif.2007.08.004

Mason, C. L., Kahle, J. B., & Gardner, A. L. (1991). Draw-a-scientist test: Future implications. *School science and mathematics, 91*(5), 193-198.

Massa, L. J., Mayer, R. E., & Bohon, L. M. (2005). Individual differences in gender role beliefs influence spatial ability test performance. *Learning and Individual Differences, 15*(2), 99-111. doi:10.1016/j.lindif.2004.11.002

Matsumura, N., Nishijo, H., Tamura, R., Eifuku, S., Endo, S., & Ono, T. (1999). Spatial- and task-dependent neuronal responses during real and virtual translocation in the monkey hippocampal formation. *J Neurosci, 19*(6), 2381-2393.

Matthews, M. H. (1986). The influence of gender on the environmental cognition of young boys and girls. *The Journal of Genetic Psychology: Research and Theory on Human Development, 147*(3), 295-302. doi:10.1080/00221325.1986.9914503

Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytic basis for JOLs? *Memory & Cognition, 29*(2), 222-233. doi:10.3758/Bf03194916

Mazzoni, G., & Nelson, T. O. (1995). Judgments of Learning Are Affected by the Kind of Encoding in Ways That Cannot Be Attributed to the Level of Recall. *Journal of*

Experimental Psychology-Learning Memory and Cognition, 21(5), 1263-1274. doi:10.1037//0278-7393.21.5.1263

McCarthy, M. M., Arnold, A. P., Ball, G. F., Blaustein, J. D., & De Vries, G. J. (2012). Sex differences in the brain: the not so inconvenient truth. *J Neurosci, 32*(7), 2241-2247. doi:10.1523/JNEUROSCI.5372-11.2012

McDonald-Miszczak, L., Gould, O. N., & Tychynski, D. (1999). Metamemory predictors of prospective and retrospective memory performance. *Journal of General Psychology, 126*(1), 37-52. doi:10.1080/00221309909595350

McDonough, I. M., & Gallo, D. A. (2012). Illusory expectations can affect retrieval-monitoring accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(2), 391-404. doi:10.1037/a0025548 21942496

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin, 86*(5), 889-918. doi:10.1037/0033-2909.86.5.889 386403

McGee, M. G. (1982). Spatial abilities: The influence of genetic factors. In M. Potegal (Ed.), *Spatial abilities: Development and physiological foundations* (pp. 199-222). New York: Academic Press.

McGuire, M. J., & Maki, R. H. (2001). When knowing more means less: The effects of fan on metamemory judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1172-1179. doi:10.1037/0278-7393.27.5.1172 11550745

McNamara, T. P. (2013). Spatial Memory: Properties and Organization. In D. Waller & L. Nadel (Eds.), *Handbook of Spatial Cognition* (pp. 173-190). Washington, DC: American Psychological Association.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M. B. (2006). Path integration and the neural basis of the 'cognitive map'. *Nat Rev Neurosci, 7*(8), 663-678. doi:10.1038/nrn1932

Meehan, A. M., & Overton, W. F. (1986). Gender differences in expectancies for success and performance on Piagetian spatial tasks. *Merrill-Palmer Quarterly, 32*(4), 427-441.

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94*, 339-354. doi:10.1348/000712603767876262

Merrill, E. C., Yang, Y., Roskos, B., & Steele, S. (2016). Sex differences in using spatial and verbal abilities influence route learning performance in a virtual environment: A comparison of 6- to 12-year old boys and girls. *Front Psychol, 7*(258), 1-17.

Messinger, P. R., Stroulia, E., Lyons, K., Bone, M., Niu, R. H., Smirnov, K., & Perelgut, S. (2009). Virtual worlds-Past, present, and future: New directions in social computing. *Decision Support Systems, 47*(3), 204-228. doi:10.1016/j.dss.2009.02.014

Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Ann Neurol, 10*(4), 309-325. doi:10.1002/ana.410100402

Metcalfe, J., & Kober, H. (2005). Self-reflective consciousness and the projectable self. In H. S. Terrace & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 57-83). New York, NY: Oxford University Press.

Mihalca, L., Mengelkamp, C., & Schnotz, W. (2017). Accuracy of metacognitive judgments as a moderator of learner control effectiveness in problem-solving tasks. *Metacognition and Learning, 12*(3), 357-379. doi:10.1007/s11409-017-9173-2

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology, 107*(3), 631-644. doi:10.1037/edu0000005

Mittelstaedt, M. L., & Mittelstaedt, H. (1980). Homing by Path Integration in a Mammal. *Naturwissenschaften, 67*(11), 566-567. doi:10.1007/Bf00450672

Mizumori, S. J., & Williams, J. D. (1993). Directionally selective mnemonic properties of neurons in the lateral dorsal nucleus of the thalamus of rats. *J Neurosci, 13*(9), 4015-4028.

Modirrousta, M., & Fellows, L. K. (2008a). Dorsal medial prefrontal cortex plays a necessary role in rapid error prediction in humans. *The Journal of Neuroscience, 28*(51), 14000-14005. doi:10.1523/JNEUROSCI.4450-08.2008 19091989

Modirrousta, M., & Fellows, L. K. (2008b). Medial prefrontal cortex plays a critical and selective role in 'feeling of knowing' meta-memory judgments. *Neuropsychologia, 46*(12), 2958-2965. doi:10.1016/j.neuropsychologia.2008.06.011

Moe, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences, 19*(1), 21-27. doi:10.1016/j.lindif.2008.02.002

Moe, A. (2012). Gender difference does not mean genetic difference: Externalizing improves performance in mental rotation. *Learning and Individual Differences, 22*(1), 20-24. doi:10.1016/j.lindif.2011.11.001

Moe, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences, 16*(4), 369-377. doi:10.1016/j.lindif.2007.01.002

Montello, D. R., Waller, D., Hegarty, M., & Richardson, A. E. (2004). Spatial memory of real environments, virtual environments, and maps. In G. L. Allen (Ed.), *Human spatial memory: Remembering where* (pp. 251-285). Mahwah, NJ: Lawrence Erlbaum Associates.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502-517. doi:10.1037/0033-295X.115.2.502 18426301

Moran, T. P. (2016). *The structure of attentional biases in anxiety: A latent variable analysis of anxiety-related modulations of attentional control.* (Doctoral Dissertation), Michigan State University. Retrieved from https://d.lib.msu.edu/etd/3257

Morony, S., Kleitman, S., Lee, Y. P., & Stankov, L. (2013). Predicting achievement: Confidence vs self-efficacy, anxiety, and self-concept in Confucian and European countries. *International Journal of Educational Research, 58*, 79-96. doi:10.1016/j.ijer.2012.11.002

Morris, R. (1984). Developments of a water-maze procedure for studying spatial learning in the rat. *Journal of neuroscience methods, 11*(1), 47-60.

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., . . . Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *J Anat, 207*(1), 35-66. doi:10.1111/j.1469

Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *J Exp Psychol Learn Mem Cogn, 30*(1), 142-157. doi:10.1037/0278

Murias, K., Kwok, K., Castillejo, A. G., Liu, I., Iaria, G., Id, & Murias, K. O. h. o. o. (2016). The effects of video game use on performance in a virtual navigation task. *Computers in Human Behavior*, 398-406. doi:10.1016/j.chb.2016.01.020

Nadel, L., & MacDonald, L. (1980). Hippocampus: Cognitive map or working memory? *Behavioral & Neural Biology, 29*(3), 405-409. doi:10.1016/S0163-1047%2880%2990430-6 7417203

Nardi, D., Newcombe, N. S., & Shipley, T. F. (2010). The role of slope in human reorientation. In C. Hölscher, T. F. Shipley, M. Olivetti Belardinelli, J. A. Bateman, & N. S. Newcombe (Eds.), *Proceedings of the International Conference Spatial Cognition VII* (pp. 32-40). Berlin-Heidelberg: Springer.

Nash, S. C. (1975). Relationship among Sex-Role Stereotyping, Sex-Role Preference, and Sex Difference in Spatial Visualization. *Sex Roles, 1*(1), 15-32. doi:10.1007/Bf00287210

National Science Board. (2010). *Preparing the next generation of STEM innovators: Identifying and developing our nation's human capital*. Retrieved from https://www.nsf.gov/nsb/publications/2010/nsb1033.pdf

Nazareth, A., Herrera, A., & Pruden, S. M. (2013). Explaining sex differences in mental rotation: Role of spatial activity experience. *Cognitive Processing, 14*(2), 201-204. doi:10.1007/s10339-013-0542-8 23381194

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109-133. doi:10.1037/0033-2909.95.1.109 6544431

Nelson, T. O. (1992). *Metacognition: Core readings*. Boston, MA: Allyn & Bacon.

Nelson, T. O., & Dunlosky, J. (1991). When Peoples Judgments of Learning (Jols) Are Extremely Accurate at Predicting Subsequent Recall - the Delayed-Jol Effect. *Psychological Science, 2*(4), 267-270. doi:10.1111/j.1467-9280.1991.tb00147.x

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of Self-Paced Study Time and the Labor-in-Vain Effect. *Journal of Experimental Psychology-Learning Memory and Cognition, 14*(4), 676-686. doi:10.1037//0278-7393.14.4.676

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The psychology of learning and motivation, 26*, 125-141.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1-25). Cambridge, MA: MIT Press.

Neuburger, S., Ruthsatz, V., Jansen, P., & Quaiser-Pohl, C. (2015). Can girls think spatially? Influence of implicit gender stereotype activation and rotational axis on fourth graders' mental-rotation performance. *Learning and Individual Differences*, 169-175. doi:10.1016/j.lindif.2014.09.003

Newcombe, N. S. (2007). Taking Science Seriously: Straight Thinking About Spatial Sex Differences. In S. Ceci & W. Williams (Eds.), *Why aren't more women in science?: Top researchers debate the evidence* (pp. 69-77). Washington, DC: American Psychological Association.

Newcombe, N. S., Bandura, M. M., & Taylor, D. G. (1983). Sex-Differences in Spatial Ability and Spatial Activities. *Sex Roles, 9*(3), 377-386. doi:10.1007/Bf00289672

Newson, J., & Newson, E. (1987). Family and sex roles in middle childhood. In D. J. Hargreaves & A. M. Colley (Eds.), *The psychology of sex roles* (pp. 142-158). New York, NY: Hemisphere Publishing.

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *Journal of General Psychology, 135*(2), 151-165. doi:10.3200/GENP.135.2.151-166 18507315

Nichols, S. (1999). Physical ergonomics of virtual environment use. *Appl Ergon, 30*(1), 79-90. doi:10.1016/S0003-6870%2898%2900045-3 10098819

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education, 74*(1), 7-28.

Nori, R., & Giusberti, F. (2003). Cognitive styles: errors in directional judgments. *Perc, 32*(3), 307-320.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., . . . Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America, 106*(26), 10593-10597. doi:10.1073/pnas.0809921106 19549876

Nowak, N. T., Murali, A., & Driscoll, I. (2015). Factors related to sex differences in navigating a computerized maze. *Journal of Environmental Psychology, 43*, 136-144. doi:10.1016/j.jenvp.2015.06.007

O'Keefe, J., & Burgess, N. (2005). Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus, 15*(7), 853-866. doi:10.1002/hipo.20115

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res, 34*(1), 171-175.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford, England: Clarendon Press

O'Laughlin, E. M., & Brubaker, B. S. (1998). Use of landmarks in cognitive mapping: Gender differences in self report versus performance. *Personality and Individual Differences, 24*(5), 595-601. doi:10.1016/S0191-8869%2897%2900237-7

Olson, J. M., Tongprasearth, K., & Nitz, D. A. (2017). Subiculum neurons map the current axis of travel. *Nat Neurosci, 20*(2), 170-172. doi:10.1038/nn.4464

Oswald, D. L. (2008). Gender stereotypes and women's reports of liking and ability in traditionally masculine and feminine occupations. *Psychology of Women Quarterly, 32*(2), 196-203. doi:10.1111/j.1471-6402.2008.00424.x

Ozkan, T., & Lajunen, T. (2006). What causes the differences in driving between young men and women? The effects of gender roles and sex on young drivers' driving behavior and self-assessment of skills. *Transportation Research Part F: Traffic Psychology and Behavior, 9*(4), 269-277. doi:10.1016/j.trf.2006.01.005

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie, 45*(3), 255-287. doi:10.1037/h0084295

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychol Rev, 15*(3), 105-130. doi:10.1007/s11065-005-7091-6 16328731

Pannu, J. K., Kaszniak, A. W., & Rapcsak, S. Z. (2005). Metamemory for faces following frontal lobe damage. *Journal of the International Neuropsychological Society, 11*(6), 668-676. doi:10.1017/S1355617705050873

Parker, P. D., Van Zanden, B., & Parker, R. B. (in press). Girls get smart, boys get smug: Historical changes in gender differences in math, literacy, and academic social comparison and achievement. *Learning and Instruction*. doi:10.1016/j.learninstruc.2017.09.002

Parron, C., & Save, E. (2004). Evidence for entorhinal and parietal cortices involvement in path integration in the rat. *Experimental Brain Research, 159*(3), 349-359. doi:10.1007/s00221-004-1960-8

Parsons, J. E., Meece, J. L., Adler, T. F., & Kaczala, C. M. (1982). Sex differences in attributions and learned helplessness. *Sex Roles, 8*(4), 421-432. doi:10.1007/BF00287281

Perfect, T. J., & Hollins, T. S. (1996). Predictive feeling of knowing judgments and postdictive confidence judgments in eyewitness memory and general knowledge. *Applied Cognitive Psychology, 10*(5), 371-382. doi:10.1002/(Sici)1099-0720

Perfect, T. J., & Hollins, T. S. (1999). Feeling-of-knowing judgments do not predict subsequent recognition performance for eyewitness memory. *Journal of Experimental Psychology-Applied, 5*(3), 250-264. doi:10.1037//1076-898x.5.3.250

Pezaris, E., & Casey, M. B. (1991). Girls who use "masculine" problem-solving strategies on a spatial task: Proposed genetic and environmental factors. *Brain and Cognition, 17*(1), 1-22. doi:10.1016/0278-2626%2891%2990062-D 1781977

Philbin, D. A., Ribarsky, W., Walker, N., & Hubbard, C. E. (1998). Training in virtual environments: Analysis of task appropriateness. *Proceedings of the Virtual Reality Annual International Symposium*, 210-210. doi:10.1109/Vrais.1998.658492

Picucci, L., Caffo, A. O., & Bosco, A. (2011). Besides navigation accuracy: Gender differences in strategy selection and level of spatial confidence. *Journal of Environmental Psychology, 31*(4), 430-438. doi:10.1016/j.jenvp.2011.01.005

Plancher, G., Tirard, A., Gyselinck, V., Nicolas, S., & Piolino, P. (2012). Using virtual reality to characterize episodic memory profiles in amnestic mild cognitive impairment and Alzheimer's disease: Influence of active and passive encoding. *Neuropsychologia, 50*(5), 592-602. doi:10.1016/j.neuropsychologia.2011.12.013 22261400

Plomin, R., Pedersen, N. L., Lichtenstein, P., McClearn, G. E., Id, & Plomin, R. O. h. o. o. (1994). Variability and stability in cognitive abilities are largely genetic later in life. *Behav Genet, 24*(3), 207-215. doi:10.1007/BF01067188 7945151

Porteus, S. D. (1965). *Porteus Maze Tests: Fifty years' application*. Palo Alto, CA: Pacific Book Publishers.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *J Neurosci, 4*(7), 1863-1874.

Postma, A., Izendoorn, R., & De Haan, E. H. (1998). Sex differences in object location memory. *Brain Cogn, 36*(3), 334-345. doi:10.1006/brcg.1997.0974

Prestopnik, J. L., & Roskos-Ewoldsen, B. (2000). The relations among wayfinding strategy use, sense of direction, sex, familiarity, and wayfinding ability. *Journal of Environmental Psychology, 20*(2), 177-191. doi:10.1006/jevp.1999.0160

Pruden, S. M., & Levine, S. C. (2017). Parents' Spatial Language Mediates a Sex Difference in Preschoolers' Spatial Language Use. *Psychological Science, 28*(11), 1583-1596.

Pruden, S. M., Levine, S. C., & Huttenlocher, J. (2011). Children's spatial thinking: Does talk about the spatial world matter? *Developmental Science, 14*(6), 1417-1430. doi:10.1111/j.1467-7687.2011.01088.x 22010900

Quirk, G. J., Muller, R. U., Kubie, J. L., & Ranck, J. B., Jr. (1992). The positional firing properties of medial entorhinal neurons: description and comparison with hippocampal place cells. *J Neurosci, 12*(5), 1945-1963.

Rabinovitch, M. S., & Rosvold, H. E. (1951). A closed-field intelligence test for rats. *Canadian Journal of Psychology/Revue canadienne de psychologie, 5*(3), 122-128. doi:10.1037/h0083542 14870071

Rahman, Q., Abrahams, S., & Jussab, F. (2005). Sex differences in a human analogue of the radial arm maze: The "17-Box Maze Test". *Brain and Cognition, 58*(3), 312-317. doi:10.1016/j.bandc.2005.03.001

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2012). Spatial anxiety, spatial ability, and working memory in early elementary school. *Cognitive Processing, 13*(1S), 49-49.

Ramos-Loyo, J., & Sanchez-Loyo, L. M. (2011). Gender differences in EEG coherent activity before and after training navigation skills in virtual environments. *Human Physiology, 37*(6), 68-75. doi:10.1134/S0362119711040165

Ranck Jr, J. (1984). Head direction cells in the deep cell layer of dorsal presubiculum in freely moving rats. *Society for Neuroscience Abstracts, 10*(176.12).

Reed-Jones, R. J., Reed-Jones, J. G., Trick, L. M., & Vallis, L. A. (2007). Can galvanic vestibular stimulation reduce simulator adaptation syndrome. *Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 534-540.

Reep, R. L., & Corwin, J. V. (2009). Posterior parietal cortex as part of a neural network for directed attention in rats. *Neurobiology of Learning and Memory, 91*(2), 104-113. doi:10.1016/j.nlm.2008.08.010 18824116

Resnick, S. M., Berenbaum, S. A., Gottesman, I. I., & Bouchard, T. J. (1986). Early hormonal influences on cognitive functioning in congenital adrenal hyperplasia. *Developmental Psychology, 22*(2), 191-198. doi:10.1037/0012-1649.22.2.191

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review, 16*(3), 550-554. doi:10.3758/PBR.16.3.550 19451383

Richardson, A. E., Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & Cognition, 27*(4), 741-750. doi:10.3758/BF03211566

Ring, P., Neyse, L., David-Barett, T., & Schmidt, U. (2016). Gender differences in performance predictions: Evidence from the cognitive reflection test. *Front Psychol, 7*(1680), 1-7.

Rizzo, M., Sheffield, R. A., Stierman, L., & Dawson, J. (2003). Demographic and driving performance factors in simulator adaptation syndrome. *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design"*, 201-208.

Rosenthal, H. E. S., Norman, L., Smith, S. P., & McGregor, A. (2012). Gender-Based Navigation Stereotype Improves Men's Search for a Hidden Goal. *Sex Roles, 67*(11-12), 682-695. doi:10.1007/s11199-012-0205-8

Roskos-Ewoldsen, B., McNamara, T. P., Shelton, A. L., & Carr, W. (1998). Mental representations of large and small spatial layouts are orientation dependent. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(1), 215-226. doi:10.1037/0278-7393.24.1.215 9438960

Rossano, M. J., & Moak, J. (1998). Spatial representations acquired from computer models: Cognitive load, orientation specificity and the acquisition of survey knowledge. *British Journal of Psychology, 89*(3), 481-497. doi:10.1111/j.2044-8295.1998.tb02698.x

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1997). Navigating buildings in "desk-top" virtual environments: Experimental investigations using extended navigational experience. *Journal of Experimental Psychology-Applied, 3*(2), 143-159. doi:10.1037/1076-898x.3.2.143

Rushworth, M. F. S., Behrens, T. E. J., & Johansen-Berg, H. (2006). Connection Patterns Distinguish 3 Regions of Human Parietal Cortex. *Cerebral Cortex, 16*(10), 1418-1430. doi:10.1093/cercor/bhj079 16306320

Ruthsatz, V., Neuburger, S., Jansen, P., & Quaiser-Pohl, C. (2015). Cars or dolls? Influence of the stereotyped nature of the items on children's mental-rotation performance. *Learning and Individual Differences*, 75-82. doi:10.1016/j.lindif.2015.08.016

Salyer, D. L., Lund, T. D., Fleming, D. E., Lephart, E. D., & Horvath, T. L. (2001). Sexual dimorphism and aromatase in the rat retina. *Brain Res Dev Brain Res, 126*(1), 131-136.

Sanders, B., Cohen, M. R., & Soares, M. P. (1986). The Sex Difference in Spatial Ability - a Rejoinder. *American Psychologist, 41*(9), 1015-1016.

Sandstrom, N. J., Kaufman, J., & Huettel, S. A. (1998). Males and females use different distal cues in a virtual environment navigation task. *Brain Res Cogn Brain Res, 6*(4), 351-360.

Saracho, O. N. (1995). Relationship between young children's cognitive style and their play. *Early Child Development and Care*, 77-84. doi:10.1080/0300443951130107

Save, E., Guazzelli, A., & Poucet, B. (2001). Dissociation of the effects of bilateral lesions of the dorsal hippocampus and parietal cortex on path integration in the rat. *Behav Neurosci, 115*(6), 1212-1223.

Save, E., & Moghaddam, M. (1996). Effects of lesions of the associative parietal cortex on the acquisition and use of spatial memory in egocentric and allocentric navigation tasks in the rat. *Behav Neurosci, 110*(1), 74-85.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336-356. doi:10.1037/0033-295X.115.2.336

Schmidt, T., & Lee, E. Y. (2006). Spatial memory organized by environmental geometry. *Spatial Cognition and Computation, 6*(4), 347-369.

Schneider, G. E. (1967). Contrasting visuomotor functions of tectum and cortex in the golden hamster. *Psychologische Forschung, 31*(1), 52-62.

Schneider, G. E. (1969). Two visual systems. *Science, 163*(3870), 895-902.

Schooler, J. W. (2002). Re-representing consciousness: dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences, 6*(8), 339-344. doi:10.1016/S1364-6613(02)01949

Schreiber, T. A., & Nelson, D. L. (1998). The relation between feelings of knowing and the number of neighboring concepts linked to the test cue. *Memory & Cognition, 26*(5), 869-883. doi:10.3758/BF03201170 9796223

Schug, M. G. (2016). Geographical cues and developmental exposure: Navigational style, wayfinding anxiety, and childhood experience in the faroe islands. *Human Nature, 27*(1), 68-81. doi:10.1007/s12110-015-9245-4 26577342

Schwartz, B. L. (2006). The feeling of going: Judgments of learning (JOLs) for maps and directions. *The International Journal of Creativity & Problem Solving, 16*(2), 5-16.

Schwartz, B. L., & Efklides, A. (2012). Metamemory and memory efficiency: Implications for student learning. *Journal of Applied Research in Memory and Cognition, 1*(3), 145-151. doi:10.1016/j.jarmac.2012.06.002

Serbin, L. A., & Connor, J. M. (1979). Sex-Typing of Childrens Play Preferences and Patterns of Cognitive Performance. *Journal of Genetic Psychology, 134*(2), 315-316.

Shackman, A. J., Sarinopoulos, I., Maxwell, J. S., Pizzagalli, D. A., Lavric, A., & Davidson, R. J. (2006). Anxiety selectively disrupts visuospatial working memory. *Emotion, 6*(1), 40-61. doi:10.1037/1528-3542.6.1.40 16637749

Sharps, M. J., Price, J. L., & Williams, J. K. (1994). Spatial Cognition and Gender - Instructional and Stimulus Influences on Mental Image Rotation Performance. *Psychology of Women Quarterly, 18*(3), 413-425. doi:10.1111/j.1471-6402.1994.tb00464.x

Sharps, M. J., Welton, A. L., & Price, J. L. (1993). Gender and Task in the Determination of Spatial Cognitive Performance. *Psychology of Women Quarterly, 17*(1), 71-83. doi:10.1111/j.1471

Shaughnessy, J. J., & Zechmeister, E. B. (1992). Memory-monitoring accuracy as influenced by the distribution of retrieval practice. *Bulletin of the Psychonomic Society, 30*(2), 125-128.

Shelton, A. L., & Gabrieli, J. (2002). Neural correlates of encoding space from route and survey perspectives. *The Journal of Neuroscience, 22*(7), 2711-2717.

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cogn Psychol, 43*(4), 274-310. doi:10.1006/cogp.2001.0758

Sherman, J. A. (1967). Problem of Sex Differences in Space Perception and Aspects of Intellectual Functioning. *Psychological Review, 74*(4), 290-299. doi:10.1037/h0024723

Shore, D. I., Stanford, L., MacInnes, W. J., Klein, R. M., & Brown, R. E. (2001). Of mice and men: Virtual Hebb-Williams mazes permit comparison of spatial learning across species. *Cognitive Affective & Behavioral Neuroscience, 1*(1), 83-89. doi:10.3758/Cabn.1.1.83

Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. *Adv Child Dev Behav, 10*, 9-55.

Signorella, M. L., & Jamison, W. (1986). Masculinity, Femininity, Androgyny, and Cognitive Performance - a Metaanalysis. *Psychological Bulletin, 100*(2), 207-228. doi:10.1037//0033

Silverman, I., Choi, J., Mackewn, A., Fisher, M., Moro, J., & Olshansky, E. (2000). Evolved mechanisms underlying wayfinding: further studies on the hunter-gatherer theory of spatial sex differences. *Evolution and Human Behavior, 21*(3), 201-213. doi:10.1016/S1090-5138(00)00036-2

Silverman, I., Choi, J., & Peters, M. (2007). The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Arch Sex Behav, 36*(2), 261-268. doi:10.1007/s10508-006

Silverman, I., & Eals, M. (1995). Sex differences in spatial abilities: Evolutionary theory and data. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 533-549). New York, NY: Oxford University Press.

Silverman, I., Phillips, K., & Silverman, L. K. (1996). Homogeneity of effect sizes for sex across spatial tests and cultures: Implications for hormonal theories. *Brain and Cognition, 31*(1), 90-94. doi:10.1006/brcg.1996.0027 8790937

Sivak, M. (2013). Female drivers in the United States, 1963–2010: from a minority to a majority? *Traffic Inj Prev, 14*(3), 259-260.

Slabbekoorn, D., van Goozen, S. H., Megens, J., Gooren, L. J., & Cohen-Kettenis, P. T. (1999). Activating effects of cross-sex hormones on cognitive functioning: a study of short-term and long-term hormone effects in transsexuals. *Psychoneuroendocrinology, 24*(4), 423-447.

Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences of the United States of America, 95*(20), 12061-12068.

Snyder, L. H., Grieve, K. L., Brotchie, P., & Andersen, R. A. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. *Nature, 394*(6696), 887-891.

Sommer, W., Heinz, A., Leuthold, H., Matt, J., & Schweinberger, S. R. (1995). Metamemory, Distinctiveness, and Event-Related Potentials in Recognition Memory for Faces. *Memory & Cognition, 23*(1), 1-11. doi:10.3758/Bf03210552

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology-Learning Memory and Cognition, 26*(1), 204-221. doi:10.1037//0278-7393.26.1.204

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, 415-437. doi:10.1146/annurev-psych-073115-103235 26361054

Spiers, M. V., Sakamoto, M., Elliott, R. J., & Baumann, S. (2008). Sex differences in spatial object-location memory in a virtual grocery store. *CyberPsychology & Behavior, 11*(4), 471-473. doi:10.1089/cpb.2007.0058 18721096

Stankov, L. (1999). Mining on the "no man's land" between intelligence and personality. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and Individual Differences: Process, Trait, and Content Determinants.* (pp. 315-337). Washington, DC: American Psychological Association.

Stankov, L., Lee, J., Luo, W. S., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences, 22*(6), 747-758. doi:10.1016/j.lindif.2012.05.013

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol, 69*(5), 797-811.

Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology, 102*(4), 947-963. doi:10.1037/a0019920

Stevens, A., & Coupe, P. (1978). Distortions in Judged Spatial Relations. *Cognitive Psychology, 10*(4), 422-437. doi:10.1016/0010-0285(78)90006-3

Stevens, C. A., & Carlson, R. A. (2016). Are you sure the library is that way? Metacognitive monitoring of spatial judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(7), 1034-1049. doi:10.1037/xlm0000229 26866660

Stinchcombe, A., Kadulina, Y., Lemieux, C. L., Aljied, R., & Gagnon, S. (2017). Driving is not a game: Video game experience is associated with risk-taking behaviors in the driving simulator. *Computers in Human Behavior, 69*, 415-420. doi:10.1016/j.chb.2016.12.006

Tambs, K., Sundet, J. M., & Magnus, P. (1984). Heritability analysis of the WAIS subtests: A study of twins. *Intelligence, 8*(4), 283-293. doi:10.1016/0160-2896%2884%2990012-6

Tascon, L., Leon, I., & Cimadevilla, J. M. (2016). Viewpoint-related gender differences in a spatial recognition task. *Learning and Individual Differences*, 270-274. doi:10.1016/j.lindif.2016.08.007

Taube, J. S. (1995). Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. *J Neurosci, 15*(1), 70-86.

Taube, J. S. (2007). The head direction signal: origins and sensory-motor integration. *Annual Review of Neuroscience, 30*, 181-207. doi:10.1146/annurev.neuro.29.051605.112854

Taube, J. S., Muller, R. U., & Ranck, J. B., Jr. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *J Neurosci, 10*(2), 436-447.

Tauber, S. K., & Rhodes, M. G. (2012). Measuring memory monitoring with judgments of retention (JORs). *The Quarterly Journal of Experimental Psychology, 65*(7), 1376-1396. doi:10.1080/17470218.2012.656665 22524495

Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles, 53*(5-6), 433-441. doi:10.1007/s11199-005-6765-0

Therrien, M. E., & Collin, C. A. (2010). Spatial vision meets spatial cognition: Examining the effect of visual blur on human visually guided route learning. *Perception, 39*(8), 1043-1064. doi:10.1068/p5991 20942357

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66-73. doi:10.1037/0022-0663.95.1.66

Thiede, K. W., & Dunlosky, J. (1994). Delaying Students Metacognitive Monitoring Improves Their Accuracy in Predicting Their Recognition Performance. *Journal of Educational Psychology, 86*(2), 290-302. doi:10.1037/0022-0663.86.2.290

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology-Learning Memory and Cognition, 25*(4), 1024-1037. doi:10.1037/0278-7393.25.4.1024

Thinus-Blanc, C. (1988). Animal Spatial Cognition. In L. Weiskrantz (Ed.), *Though without language* (pp. 371-395). Oxford, England: Clarendon Press.

Tippett, C. D., & Milford, T. (2013). Using visual data to investigate influences on preservice teachers' attitudes about scientists. In K. D. Finson & J. Pedersen (Eds.), *Visual data and their use in science education* (pp. 127-148). Charlotte, NC: IAP Information Age Publishing.

Titze, C., Jansen, P., & Heil, M. (2011). Single-sex school girls outperform girls attending a co-educative school in mental rotation accuracy. *Sex Roles, 65*(9-10), 704-711. doi:10.1007/s11199-011-9947

Tlauka, M., Keage, H., & Clark, C. R. (2005). Viewing a map versus reading a description of a map: Modality-specific encoding of spatial information. *Cognitive Science, 29*(5), 807-818. doi:10.1207/s15516709cog0000_26

Tobias, S., & Everson, H. T. (1997). Studying the relationship between affective and metacognitive variables. *Anxiety, Stress & Coping: An International Journal, 10*(1), 59-81. doi:10.1080/10615809708249295

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychol Rev, 55*(4), 189-208.

Townsend, C. L., & Heit, E. (2011). Judgments of learning and improvement. *Memory & Cognition, 39*(2), 204-216. doi:10.3758/s13421-010-0019-2 21264622

Tracy, D. M. (1987). Toys, spatial ability, and science and mathematics achievement: Are they related? *Sex Roles, 17*(3-4), 115-138. doi:10.1007/BF00287620

Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forsch, 31*(4), 299-348.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247*(4940), 301-306. doi:10.1126/science.2296719 2296719

Tversky, B. (1992). Distortions in Cognitive Maps. *Geoforum, 23*(2), 131-138. doi:10.1016/0016-7185(92)90011-R

U.S. Department of Commerce. (2012). *The competitiveness and innovative capacity of the Unites States*. Retrieved from http://www.esa.doc.gov/sites/default/files/thecompetitivenessandinnovative capacityoftheunitedstates.pdf

Underwood, B. J. (1966). Individual and Group Predictions of Item Difficulty for Free Learning. *Journal of Experimental Psychology, 71*(5), 673-679. doi:10.1037/H0023107

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The Malleability of Spatial Skills: A Meta-Analysis of Training Studies. *Psychological Bulletin, 139*(2), 352-402. doi:10.1037/a0028446

Van Asselen, M., Kessels, R. P., Neggers, S. F., Kappelle, L. J., Frijns, C. J., & Postma, A. (2006). Brain areas involved in spatial working memory. *Neuropsychologia, 44*(7), 1185-1194.

van Gerven, D. J., Schneider, A. N., Wuitchik, D. M., & Skelton, R. W. (2012). Direct measurement of spontaneous strategy selection in a virtual Morris water maze shows females choose an allocentric strategy at least as often as males do. *Behavioral Neuroscience, 126*(3), 465-478.

van Groen, T., & Wyss, J. M. (1992). Projections from the Laterodorsal Nucleus of the Thalamus to the Limbic and Visual Cortices in the Rat. *Journal of Comparative Neurology, 324*(3), 427-448. doi:10.1002/cne.903240310

van Groen, T., & Wyss, J. M. (1995). Projections from the Anterodorsal and Anteroventral Nucleus of the Thalamus to the Limbic Cortex in the Rat. *Journal of Comparative Neurology, 358*(4), 584-604.

van Strien, N. M., Cappaert, N. L., & Witter, M. P. (2009). The anatomy of memory: an interactive overview of the parahippocampal-hippocampal network. *Nat Rev Neurosci, 10*(4), 272-282. doi:10.1038/nrn2614 van Veen, H. A., Distler, H. K., Braun, S. J., & Bülthoff, H. H. (1998). Navigating through a virtual city: Using virtual reality technology to study human action and perception. *Future Generation Computer Systems, 14*(3), 231-242.

Vanderberg, S. G. (1969). Human behavior genetics: Present status and suggestions for future research. *Merrill-Palmer Quarterly, 15*(1), 121-154.

Vecchi, T., Albertin, L., & Cornoldi, C. (1999). Self-assessment of everyday spatial memory and performance on memory tasks in old age. *Clinical gerontologist, 20*(3), 57-66.

Veenman, M. V. J., Kerseboom, L., & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: Availability versus production deficiencies. *Anxiety, Stress & Coping: An International Journal, 13*(4), 391-412. doi:10.1080/10615800008248343

Vogt, B. A., & Miller, M. W. (1983). Cortical connections between rat cingulate cortex and visual, motor, and postsubicular cortices. *J Comp Neurol, 216*(2), 192-210. doi:10.1002/cne.902160207

von Stulpnagel, R., & Steffens, M. C. (2012). Can active navigation be as good as driving? A comparison of spatial memory in drivers and backseat drivers. *Journal of Experimental Psychology: Applied, 18*(2), 162-177. doi:10.1037/a0027133 22329733

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. *Psychonomic Bulletin & Review, 18*(2), 267-277.

Voyer, D., & Doyle, R. A. (2010). Item type and gender differences on the Mental Rotations Test. *Learning and Individual Differences, 20*(5), 469-472. doi:10.1016/j.lindif.2010.04.010

Voyer, D., & Jansen, P. (2017). Motor expertise and performance in spatial tasks: A meta-analysis. *Human Movement Science, 54*, 110-124.

Voyer, D., Postma, A., Brake, B., & Imperato-McGinley, J. (2007). Gender differences in object location memory: A meta-analysis. *Psychonomic Bulletin & Review, 14*(1), 23-38. doi:10.3758/BF03194024 17546728

Voyer, D., & Sullivan, A. M. (2003). The relation between spatial and mathematical abilities: Potential factors underlying suppression. *International Journal of Psychology, 38*(1), 11-23. doi:10.1080/00207590244000241

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of Sex-Differences in Spatial Abilities - a Metaanalysis and Consideration of Critical Variables. *Psychological Bulletin, 117*(2), 250-270. doi:10.1037/0033-2909.117.2.250

Voyer, D., Voyer, S., & Saint-Aubin, J. (2017). Sex differences in visual-spatial working memory: A meta-analysis. *Psychonomic Bulletin & Review, 24*(2), 307-334. doi:10.3758/s13423-016-1085-7 27357955

Vytal, K. E., Cornwell, B. R., Letkiewicz, A. M., Arkin, N. E., & Grillon, C. (2013). The complex interaction between anxiety and cognition: Insight from spatial and verbal working memory. *Frontiers in Human Neuroscience, 7*(93), 1-11. doi:10.3389/fnhum.2013.00093 23542914

Waller, D. (2000). Individual differences in spatial learning from computer-simulated environments. *Journal of Experimental Psychology: Applied, 6*(4), 307-321.

Waller, D., Knapp, D., & Hunt, E. (2001). Spatial representations of virtual mazes: The role of visual fidelity and individual differences. *Human Factors, 43*(1), 147-158. doi:10.1518/001872001775992561

Waller, D., & Nadel, L. (2013). *Handbook of Spatial Cognition*. Washington, D.C.: American Psychological Association.

Walton, G. M., & Cohen, G. L. (2003). Stereotype Lift. *Journal of Experimental Social Psychology, 39*(5), 456-467. doi:10.1016/S0022-1031(03)00019-2

Wang, L., & Carr, M. (2014). Working memory and strategy use contribute to gender differences in spatial ability. *Educational Psychologist, 49*(4), 261-282. doi:10.1080/00461520.2014.960568

Wang, R., & Spelke, E. (2002). Human spatial representation: insights from animals. *Trends in Cognitive Sciences, 6*(9), 376-382.

Ward, S. L., Newcombe, N., & Overton, W. F. (1986). Turn Left at the Church, or 3 Miles North - a Study of Direction Giving and Sex-Differences. *Environment and Behavior, 18*(2), 192-213. doi:10.1177/0013916586182003

Watier, N., & Collin, C. (2010). Metamemory for faces, names, and words. *Canadian Journal of Experimental Psychology, 64*(4), 324-324.

Watier, N., & Collin, C. (2011). Metamemory for faces, names, and common nouns. *Acta Psychologica, 138*(1), 143-154. doi:10.1016/j.actpsy.2011.05.017

Watier, N., & Collin, C. (2012). The effects of distinctiveness on memory and metamemory for face-name associations. *Memory, 20*(1), 73-88. doi:10.1080/09658211.2011.637935

Wehner, R., Michel, B., & Antonsen, P. (1996). Visual navigation in insects: Coupling of egocentric and geocentric information. *Journal of Experimental Biology, 199*(1), 129-140.

Weniger, G., Ruhleder, M., Wolf, S., Lange, C., & Irle, E. (2009). Egocentric memory impaired and allocentric memory intact as assessed by virtual reality in subjects with unilateral parietal cortex lesions. *Neuropsychologia, 47*(1), 59-69. doi:10.1016/j.neuropsychologia.2008.08.018

West, R. L., Boatwright, L. K., & Schleser, R. (1984). The link between memory performance, self-assessment, and affective status. *Exp Aging Res, 10*(4), 197-200. doi:10.1080/03610738408258464 6535732

Wiener, J. M., Buchner, S. J., & Holscher, C. (2009). Taxonomy of human wayfinding tasks: A knowledge-based approach. *Spatial Cognition and Computation, 9*(2), 152-165. doi:10.1080/13875860902906496

Williams, C. L., Barnett, A. M., & Meck, W. H. (1990). Organizational effects of early gonadal secretions on sexual differentiation in spatial memory. *Behavioral Neuroscience, 104*(1), 84-97. doi:10.1037/0735-7044.104.1.84 2317288

Williams, C. L., & Meck, W. H. (1991). The Organizational Effects of Gonadal-Steroids on Sexually Dimorphic Spatial Ability. *Psychoneuroendocrinology, 16*(1-3), 155-176. doi:10.1016/0306

Witmer, B. G., Bailey, J. H., Knerr, B. W., & Parsons, K. C. (1996). Virtual spaces and real world places: Transfer of route knowledge. *International Journal of Human-Computer Studies, 45*(4), 413-428. doi:10.1006/ijhc.1996.0060

Wolbers, T., & Hegarty, M. (2010). What determines our navigational abilities? *Trends Cogn Sci, 14*(3), 138-146. doi:10.1016/j.tics.2010.01.001

Wolbers, T., Hegarty, M., Buchel, C., & Loomis, J. M. (2008). Spatial updating: how the brain keeps track of changing object locations during observer motion. *Nat Neurosci, 11*(10), 1223-1230. doi:10.1038/nn.2189

Wolbers, T., Weiller, C., & Buchel, C. (2004). Neural foundations of emerging route knowledge in complex spatial environments. *Brain Res Cogn Brain Res, 21*(3), 401-411. doi:10.1016/j.cogbrainres.2004.06.013

Wraga, M., Duncan, L., Jacobs, E. C., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review, 13*(5), 813-819. doi:10.3758/Bf03194002

Wraga, M., Helt, M., Jacobs, E., & Sullivan, K. (2007). Neural basis of stereotype-induced shifts in women's mental rotation performance. *Social Cognitive and Affective Neuroscience, 2*(1), 12-19. doi:10.1093/scan/nsl041 18985116

Wyss, J. M., & van Groen, T. (1992). Connections between the Retrosplenial Cortex and the Hippocampal-Formation in the Rat - a Review. *Hippocampus, 2*(1), 1-12.

Xiao, C., Mou, W., & McNamara, T. P. (2009). Use of self-to-object and object-to-object spatial relations in locomotion. *J Exp Psychol Learn Mem Cogn, 35*(5), 1137-1147. doi:10.1037/a0016273

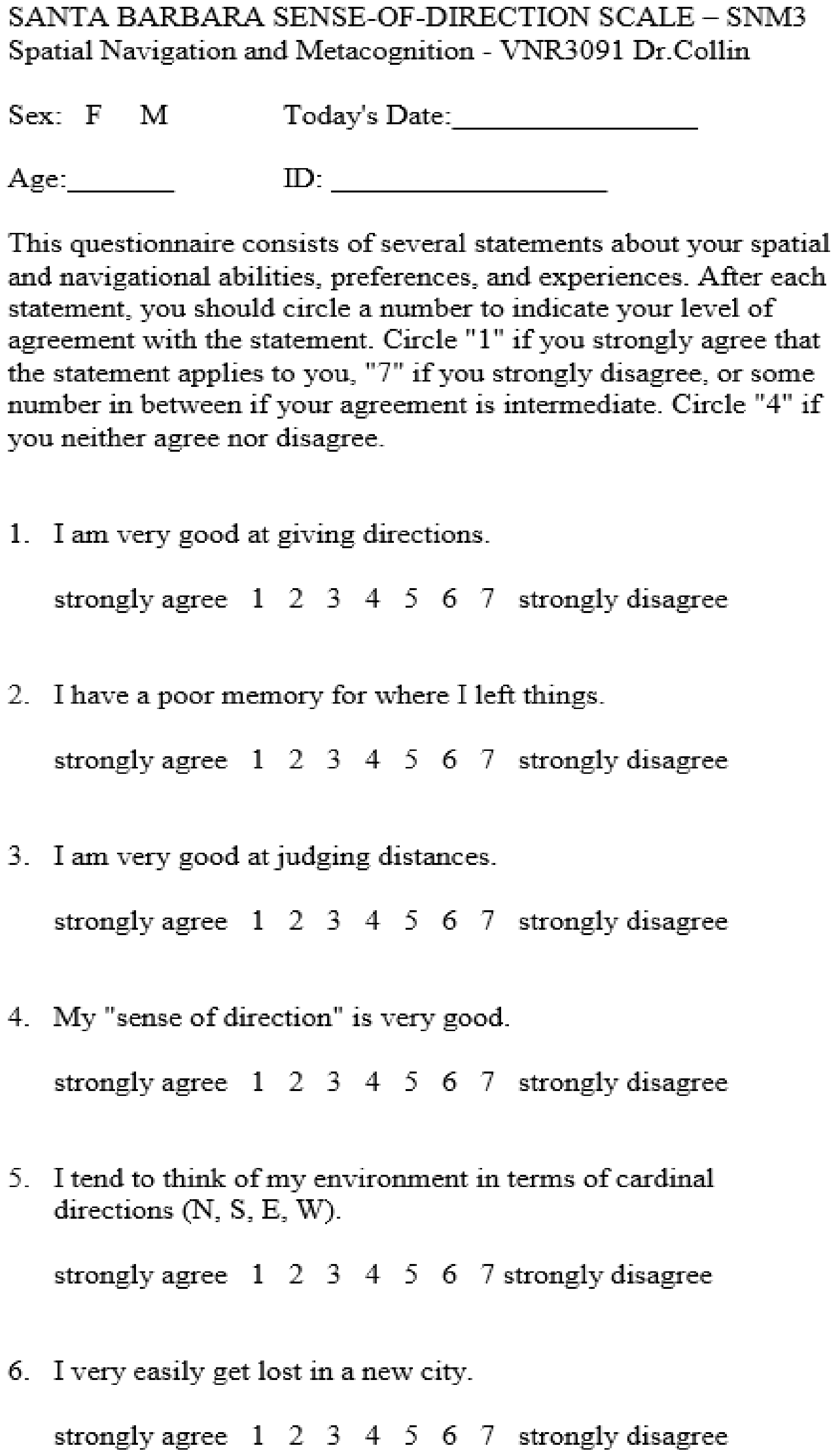
Yeung, N. C. J., & von Hippel, C. (2008). Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers. *Accident Analysis and Prevention, 40*(2), 667-674. doi:10.1016/j.aap.2007.09.003 18329419

Yilmaz, R. M., Baydas, O., Karakus, T., & Goktas, Y. (2015). An examination of interactions in a three-dimensional virtual world. *Computers & Education, 88*, 256-267. doi:10.1016/j.compedu.2015.06.002

Yoon, S.-Y., Choi, Y. J., & Oh, H. (2015). User attributes in processing 3D VR-enabled showroom: Gender, visual cognitive styles, and the sense of presence. *International Journal of Human-Computer Studies, 82*, 1-10.

Young, D. M., Rudman, L. A., Buettner, H. M., & McLean, M. C. (2013). The influence of female role models on women's implicit science cognitions. *Psychology of Women Quarterly, 37*(3), 283-292. doi:10.1177/036168431348210

APPENDIX A: SANTA BARBARA SENSE-OF-DIRECTION SCALE (SBSOD)



SANTA BARBARA SENSE-OF-DIRECTION SCALE – SNM3
Spatial Navigation and Metacognition - VNR3091 Dr.Collin

Sex: F M Today's Date:________________

Age:_______ ID: __________________

This questionnaire consists of several statements about your spatial and navigational abilities, preferences, and experiences. After each statement, you should circle a number to indicate your level of agreement with the statement. Circle "1" if you strongly agree that the statement applies to you, "7" if you strongly disagree, or some number in between if your agreement is intermediate. Circle "4" if you neither agree nor disagree.

1. I am very good at giving directions.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

2. I have a poor memory for where I left things.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

3. I am very good at judging distances.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

4. My "sense of direction" is very good.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

5. I tend to think of my environment in terms of cardinal directions (N, S, E, W).

   strongly agree 1 2 3 4 5 6 7 strongly disagree

6. I very easily get lost in a new city.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

7. I enjoy reading maps.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

8. I have trouble understanding directions.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

9. I am very good at reading maps.

   strongly agree 1 2 3 4 5 6 7 strongly disagree

10. I don't remember routes very well while riding as a passenger in a car.

    strongly agree 1 2 3 4 5 6 7 strongly disagree

11. I don't enjoy giving directions.

    strongly agree 1 2 3 4 5 6 7 strongly disagree

12. It's not important to me to know where I am.

    strongly agree 1 2 3 4 5 6 7 strongly disagree

13. I usually let someone else do the navigational planning for long trips.

    strongly agree 1 2 3 4 5 6 7 strongly disagree

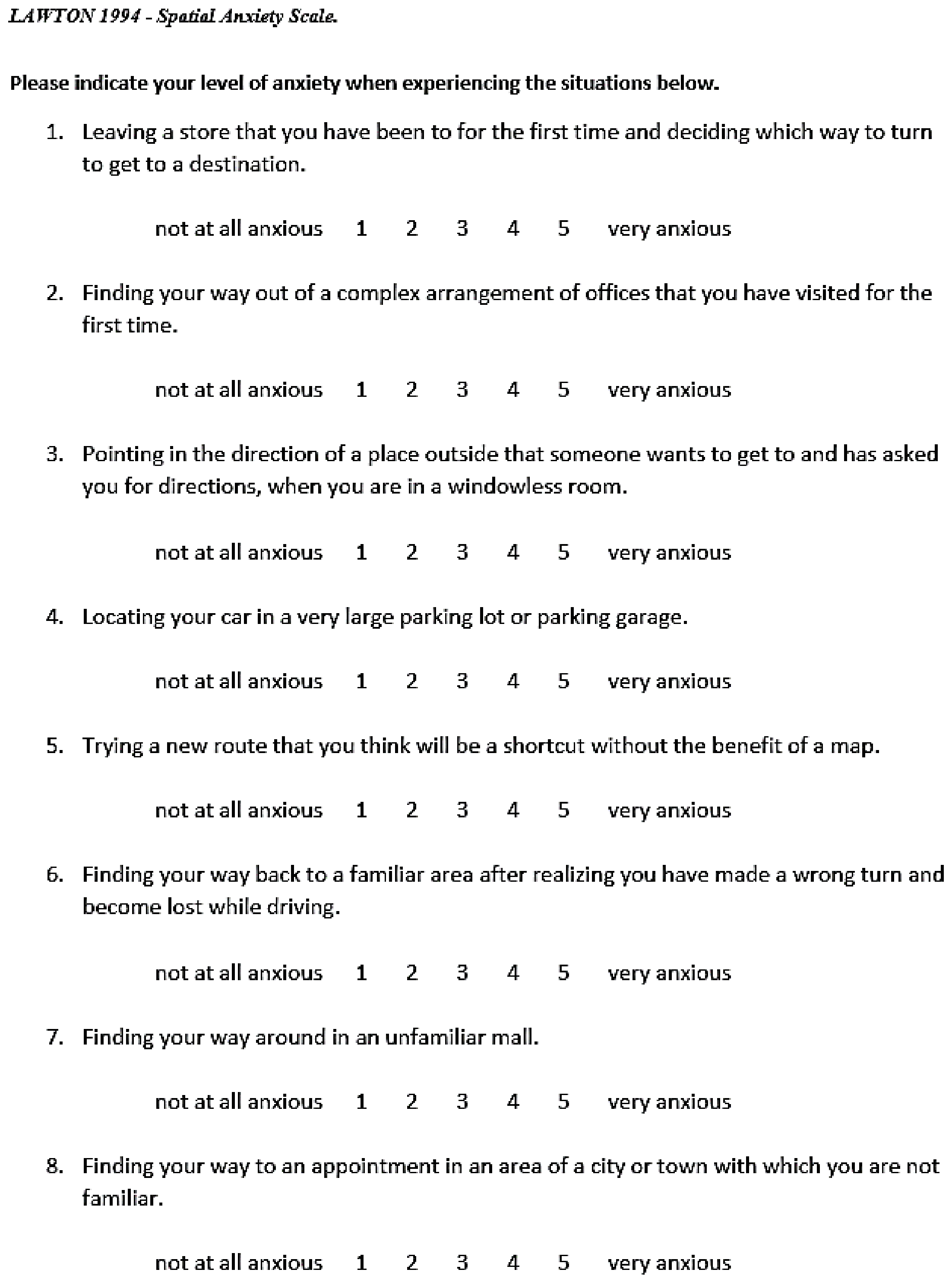
14. I can usually remember a new route after I have traveled it only once.

    strongly agree 1 2 3 4 5 6 7 strongly disagree

15. I don't have a very good "mental map" of my environment.

    strongly agree 1 2 3 4 5 6 7 strongly disagree

APPENDIX B: SPATIAL ANXIETY SCALE (SAS)



LAWTON 1994 - Spatial Anxiety Scale.

Please indicate your level of anxiety when experiencing the situations below.

1. Leaving a store that you have been to for the first time and deciding which way to turn to get to a destination.

   not at all anxious 1 2 3 4 5 very anxious

2. Finding your way out of a complex arrangement of offices that you have visited for the first time.

   not at all anxious 1 2 3 4 5 very anxious

3. Pointing in the direction of a place outside that someone wants to get to and has asked you for directions, when you are in a windowless room.

   not at all anxious 1 2 3 4 5 very anxious

4. Locating your car in a very large parking lot or parking garage.

   not at all anxious 1 2 3 4 5 very anxious

5. Trying a new route that you think will be a shortcut without the benefit of a map.

   not at all anxious 1 2 3 4 5 very anxious

6. Finding your way back to a familiar area after realizing you have made a wrong turn and become lost while driving.

   not at all anxious 1 2 3 4 5 very anxious

7. Finding your way around in an unfamiliar mall.

   not at all anxious 1 2 3 4 5 very anxious

8. Finding your way to an appointment in an area of a city or town with which you are not familiar.

   not at all anxious 1 2 3 4 5 very anxious