



Published in final edited form as:

*Intelligence*. 2018 ; 70: 42–51. doi:10.1016/j.intell.2018.08.001.

## Are there Sex Differences in Confidence and Metacognitive Monitoring Accuracy for Everyday, Academic, and Psychometrically Measured Spatial Ability?

Robert Ariel<sup>1</sup>, Natalie A. Lembeck<sup>2</sup>, Scott Moffat<sup>2</sup>, and Christopher Hertzog<sup>2</sup>

<sup>1</sup>Purdue University

<sup>2</sup>Georgia Institute of Technology

### Abstract

The current study evaluated sex differences in (1) self-perceptions of everyday and academic spatial ability, and (2) metacognitive monitoring accuracy for measures of spatial visualization and spatial orientation. Undergraduate students completed the Paper Folding Test, Spatial Relations Test, and the Revised Purdue Spatial Visualization Test while making confidence judgments (CJs) for each trial. They also made global estimates of performance and rated their ability to perform several everyday and academic spatial scenarios. Across multiple spatial measures, female students displayed lower confidence in their item-level monitoring and global assessments of performance than did male students, even when no actual differences in spatial performance occurred. Women were also less confident in their self-assessments of their visuospatial ability for scientific domains than were men. However, the absolute and relative accuracy of CJs did not differ as a function of sex suggesting that women can monitor their spatial performance as well as men.

### Keywords

Metacognition; confidence; spatial ability; STEM; sex differences

## 1. Introduction

Spatial cognition is a multifaceted construct encompassing the mental operations involved in visualizing, remembering, manipulating, and reasoning about the location and orientation of objects and places (Carroll, 1993; Hegarty & Waller, 2005; Michael, Guilford, Fruchter, & Zimmerman, 1957). It is utilized in many everyday tasks such as remembering the location of your house keys, packing a suit case, assembling objects like furniture, and navigating to

---

Please send correspondence to Robert Ariel, Purdue University, Department of Psychological Sciences, West Lafayette, IN; rariel@purdue.edu.

Declarations of interest: none

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

both familiar and unfamiliar locations. Spatial cognitive processing even facilitates learning and reasoning in science, technology, engineering, and mathematics domains (STEM), presumably because conceptual information in these domains requires one to think and reason spatially about important domain relevant information (Burnett, Lane, & Dratt, 1979; Kozhevnikov, Motes, & Hegarty, 2007; Orion, Ben-Chaim, & Kali, 1997; Newcombe, 2016; Pribyl & Bodner, 1987; Sanchez & Wiley, 2014; Uttal & Cohen, 2012; van Garderen, 2006; Wai, Lubinski, & Benbow, 2009). For example, identifying chirality in stereochemistry involves visualizing the mirror image of a molecule and mentally rotating it until it aligns on itself (see Uttal & Cohen, 2012). Understanding and applying Newton's first law in physics involves knowing how an object's speed and trajectory are impacted by other forces. Even understanding the biological structure of animal cells such as how the sodium potassium pump functions is an inherently spatial concept because it focuses on the movement and location of sodium and potassium ions in relation to the plasma membrane of cells.

Given the critical importance of spatial cognition for both every day and academic domains, people need to be able to accurately monitor their spatial cognitive performance. Inaccurate monitoring of one's spatial abilities could have several negative implications. One implication is that students who are underconfident in their spatial ability may choose not to use spatial strategies during learning. They may even be reluctant to pursue coursework or careers that require routine spatial thinking (e.g., STEM fields). This may be especially true for female students because (a) they believe they have lower spatial ability than male students (for a review, see Syzmanowicz & Furnham, 2011) and they are more likely to experience anxiety when engaging in spatial processing (Maloney, Waechter, Risko, & Fugelsang, 2012). Thus, identifying whether sex differences are present in metacognitive monitoring accuracy and how to improve monitoring accuracy in spatial domains could have important applied implications. The current study examined potential sex differences in metacognitive monitoring of spatial cognition in everyday, academic, and traditional psychometric measures of spatial cognition.

Extensive research has focused on understanding when and why sex differences are observed in spatial cognition (Halpern & Collaer, 2005; Maeda & Yoon, 2013; Voyer, Postma, Brake, & Imperato-McGinley, 2006; Voyer, Voyer, & Bryden, 1995; Voyer, Voyer, & Saint-Aubin, 2017). Substantial sex differences in performance favoring males over females are present for many measures of spatial processing (Halpern & Collaer, 2005). Males typically outperform female students on measures of visual spatial working memory, navigation, spatial orientation (e.g., mental rotation), and spatial visualization (one's ability to mentally transform objects into new forms). However, sex differences are not present for spatial tasks that focus on long-term object location memory. Women often outperform men on many episodic memory tasks; especially verbal memory tasks (Herlitz & Rehnman, 2008; Herlitz, Nilsson, & Backman, 1997).

Despite this large body of research examining sex differences in spatial cognitive performance, few experiments have focused on potential sex differences in metacognitive monitoring accuracy in spatial domains (e.g., Cooke-Simpson & Voyer, 2006; Estes & Felker, 2012). The available evidence suggests that female students may be less accurate at evaluating their spatial performance than male students. However, this evidence is based

primarily on monitoring accuracy of confidence judgments for item responses on the Vandenberg and Kuse (1978) mental rotation test (Cooke-Simpson & Voyer, 2006). Sex differences in item-level monitoring accuracy have not been evaluated for any other spatial task. There is extensive research on sex differences in global self-assessments of spatial ability (for review, see Syzmanowicz & Furnham, 2011) and a few studies have evaluated age differences in monitoring accuracy for visual spatial working memory (Ariel & Moffat, 2018; Thomas, Bonura, Taylor, & Brunyé, 2012) and tasks measuring spatial visualization (e.g. Paper Folding Test), spatial orientation (mental rotation), and spatial navigation (Ariel & Moffat, 2018). The remaining research examining spatial performance monitoring has focused on monitoring accuracy for spatial judgments about length (Schraw, Dunkle, Roedel, & Bendixen, 1995) and spatial reasoning on the Raven's Progressive Matrices Test (Mitchum & Kelly, 2010; Schraw & Nietfeld, 1998) without considering potential individual differences.

Only a few studies have explored whether there are sex differences in metacognitive monitoring accuracy in non-spatial domains (Lichtenstein & Fischhoff, 1981; Lundeberg, Fox, & Punchochar, 1994; Hertzog, Dixon, & Hultsch, 1990). Lichtenstein and Fischhoff (1981) evaluated sex differences in monitoring memory for general knowledge questions and found no evidence for sex differences in monitoring ability. Hertzog, Dixon, and Hultsch (1990) examined sex and age differences in monitoring memory for categorized lists and narrative text recall. Women were more underconfident in their memory for categorized lists than were men but women were more accurate than men at monitoring their narrative text recall. Finally, Lundeberg, Fox, and Punchochar (1994) examined sex differences in memory for content from an undergraduate psychology research methods course. Male students were more overconfident in their memory for incorrect information than female students. Taken together, the limited available evidence suggests that sex differences may be present in some domains (memory for categorical lists, narrative text recall) and not others (general knowledge), and there does not appear to be clear evidence for a general male or female advantage in monitoring ability.

The limited research examining sex differences in monitoring spatial cognition is especially surprising because sex differences in spatial cognitive performance have been indirectly linked to metacognitive variables. For example, female students are more likely than male students to withhold low confidence responses that are accurate on the mental rotation test which contributes to observed sex differences in performance (Cooke-Simpson & Voyer, 2006). They also adopt different strategies than male students to solve spatial problems (Allen & Hogeland, 1978; Goldstein, Haldane, & Mitchell, 1990; Lohman, 1986; Miller & Santoni, 1986; Kail, Carter, & Pellegrino, 1979; Pena, Contreras, Shih, & Santacreu, 2008; Prinzel & Freeman, 1995; Raabe, Hoyer, & Delius, 2006; Tapley & Bryden, 1977). During mental rotation, males are more likely to use a holistic strategy that involves mentally rotating an entire object, whereas female students are more likely to use an analytic strategy that involves mentally rotating smaller pieces of an object and comparing each piece to components of potential response options (Raabe, Hoyer, & Delius, 2006). These differences in strategy preference may be due to differences in the accuracy of monitoring strategy effectiveness.

Sex differences in spatial strategy use could also cause sex differences in item-level monitoring accuracy. Metacognitive monitoring is an inferential process that involves evaluating cues (e.g. item characteristics, processing fluency, etc.) that are present at the time of a monitoring judgment and applying beliefs or heuristics to infer the quality of these processes (Dunlosky & Tauber, 2014; Koriat, 1997, Schwartz, Benjamin, & Bjork 1997). Different strategies can afford access to different cues during monitoring that vary in diagnosticity (Mitchum & Kelly, 2010). In spatial domains, holistic spatial strategy use may afford access to cues associated with generating and manipulating spatial representations for items (e.g., processing fluency, vividness of imagery, etc.) that would not be present when people use nonholistic analytical strategies. Thus, one mechanism that could produce sex differences in monitoring accuracy is differences in cue utilization caused by sex differences in strategy preferences.

Metacognitive monitoring accuracy is typically evaluated by comparing performance accuracy on multiple trials of a target task to metacognitive judgments of performance on those trials (e.g, confidence judgments). Absolute accuracy (also referred to as calibration) refers to whether the average magnitude of an individual's judgments corresponds to their overall level of performance. Relative accuracy refers to one's ability to discriminate between correct and incorrect spatial task decisions (i.e., manifest higher confidence for correct than for incorrect item responses). In the current experiment, we compared sex differences for both absolute and relative accuracy on measures of spatial orientation and spatial visualization.

In addition to examining monitoring accuracy in spatial visualization and spatial orientation tasks, we also included measures of visual spatial working memory (Symmetry Span, Oswald et al., 2015) which typically favor male students over female students (for a review, see Voyer, Voyer, & Saint-Aubin, 2017) and general fluid intelligence (Raven's Progressive Matrices; Raven, Raven, & Court, 1998) which sometimes produce small sex differences also favoring males (Lynn & Irwing, 2004; Irwing & Lynn, 2005). Most important, we also examined sex differences in students' subjective assessments of their performance ability and experience in several contexts using a modified version of Salthouse and Mitchell's (1990) Spatial Experience Questionnaire. The Spatial Experience Questionnaire presents participants with spatial scenarios that one might encounter in their daily life (e.g., imagining different arrangements of furniture, visualizing travel directions, considering how a building would look from a different vantage point) and prompts them to rate their general ability, recent experience, and cumulative experience performing the specified spatial task. We modified it by adding four additional items to examine STEM related spatial thinking (e.g., visualizing mathematical relationships, micro-level concepts in biology or chemistry, concepts in physics, and locations/direction in anatomy). These new questions allowed us to contrast potential sex differences in perceptions of everyday spatial ability and academic spatial ability.

Students who are proficient in reasoning spatially during routine daily activities, may not necessarily be proficient in reasoning spatially in academic domains. Reasoning skills for day-to-day tasks do not always transfer to similar academic tasks (Reeve, Palinesar, & Brown, 1987). For example, some students can complete complex mathematical

computations during daily activities like grocery shopping, but they fail to complete similar math problems in the classroom (Carraher, Carraher, & Schliemann, 1985; Lave, 1988). Likewise, a student who is adept at visualizing the spatial orientation of objects when arranging furniture or packing a suitcase, may not be able to apply these same mental rotation skills to solve stereochemistry or other STEM related problems. Thus, we chose to examine students' experience and perceptions of their academic spatial ability separately from their everyday spatial ability.

## 2 Method

### 2.1. Participants.

Two hundred and thirty-three undergraduate students at the Georgia Institute of Technology participated in this experiment for course credit in an introductory psychology course (144 males and 89 females). Our final sample size was not selected a priori. Instead, we recruited as many students as possible across the academic year and terminated data collection when our research pool closed. The majority of students recruited were STEM majors (90% of males and 85% of females). Most of these students were pursuing a degree in engineering (53% of males and 56% of females). The remaining students were pursuing science (biology, chemistry, or physics: 3% of males and 13% of females), technology (computer science or computational media: 33% of males and 15% of females), or mathematics degrees (1% of males and 1% of females).

### 2.2. Materials & Procedure.

All participants were tested individually. Each task except for a demographic survey was administered by computer using a customized program. Tasks were administered in the following order: (1) Demographic Survey, (2) Spatial Experience Questionnaire, (3) Raven's Progressive Matrices, (5) Symmetry Span Task, (6) Paper Folding Test, (7) Spatial Relations Test, and (8) Revised Purdue Spatial Visualization Test. Since we were interested in evaluating individual differences across tasks, we chose to keep the fixed task order identical for all participants (for rationale, see Carlson & Moses, 2001).

The reliability for each spatial performance measure was examined by computing Cronbach's  $\alpha$  reliability coefficient ( $r_{\alpha}$ ). Each task displayed high reliability (Spatial Experience Questionnaire:  $r_{\alpha} = .78$ ; Paper Folding test:  $r_{\alpha} = .76$ ; Spatial Relations test:  $r_{\alpha} = .87$ ; PSVT:R:  $r_{\alpha} = .89$ ). The procedure for each task is described in detail below.

**2.2.1. Spatial Experience Questionnaire.**—Participants first completed an adapted version of Salthouse and Mitchell's (1990) Spatial Experience Questionnaire to measure beliefs and experience performing several different daily spatial tasks. This questionnaire prompted participants to rate their recent experience (average hours per week), cumulative experience (number of years performed), and their beliefs about their ability to perform a number of different spatial tasks (rated on a scale from 1 to 10) including imagining different arrangements of objects like furniture, devising efficient ways for packing or loading a box, and visualizing travel directions from a verbal description (See Table A1 in the Appendix for the complete question set). Four additional items were added to this

questionnaire to assess beliefs and experience using spatial processes in educational domains including imagining mathematical relationships, biological concepts, physics concepts, and visualizing anatomy (see questions 10 to 13 in Table A1). We also added 2 additional ratings to each questionnaire that prompted students to rate their performance on a scale from 1 to 10 relative to other students of the same and opposite gender. However, these ratings were nearly identical to the standard ability ratings students provided, so we do not discuss them further.

**2.2.2. Raven's Progressive Matrices.**—A computerized version of Raven's Progressive Matrices (Raven, Raven, & Court, 1998) was administered to assess non-verbal fluid intelligence. The task consisted of 18 trials ordered ascending in their normative difficulty adapted from Stanovich and Cunningham (1993). On each trial, a display of  $3 \times 3$  array was presented consisting of 8 geometric figures with a missing 9<sup>th</sup> figure presented in the bottom right-hand position of the array. Participants could choose from 8 potential figures positioned below the  $3 \times 3$  array to complete the pattern. Participants had 12 minutes to complete this task.

**2.2.3. Symmetry Span Task.**—A shortened version of the Symmetry Span task was administered to examine sex differences in visual-spatial working memory (Oswald et al., 2015). This complex span task consists of alternating distractor and memory trials. Participants are first asked to judge whether a shape is vertically symmetrical (distractor trial). Next, a to-be remembered red box is presented in one location of a  $4 \times 4$  grid (memory trial). This is followed by another distractor trial and a new memory trial for 2 to 5 additional times. After the last presentation of the set, participants are instructed to click boxes in the grid in the order they appeared during the memory trials. Memory loads range from 2 to 5 and participants are tested on each memory load once.

**2.2.4. Paper Folding.**—The VZ-2 Paper Folding test was modified to examine sex differences in monitoring and performance for spatial visualization (French, Ekstrom, & Price, 1985). The task consisted of 20 trials presented in a randomized order where on each trial participants viewed a drawing of a paper folded one to three times. Participants were instructed to visualize that a circular hole was punched into the folded paper and then to select among 5 response options the piece of paper that contains the hole located in the correct locations when unfolded. They chose a response option by clicking a button below it. After selecting their response, they rated their confidence in the accuracy of their selection by moving a slider to any value between 0 (not at all confident) and 100 (extremely confident). Participants then viewed a screen for 2 seconds instructing them to get ready for the next trial. Before beginning the task, participants were asked to make a global prediction about the percentage of trials they believed they could answer correctly. After completing the task, they also made global postdictions about the percentage of trials they believed they performed correctly. All predictions were made by moving a slider between values of 0 and 100.

**2.2.5. Mental Rotation.**—Monitoring differences in spatial orientation were examined using a modified version of Thurstone and Thurstone's (1947) Spatial Relations test which



involved 2-dimensional (2-D) mental rotation. The task consisted of 30 trials in which participants viewed a 2-D drawing and then 5 response options containing either identical drawings rotated into a different orientation or similar but different drawings presented in various orientations. The goal is to select any option that is the same as the target drawing and multiple options can be correct on any given trial. Participants were instructed to select the response options that were identical but rotated differently on each trial by clicking a check box below each one. After selecting response options, participants clicked a button to indicate they were finished. They then were instructed to make a CJ using the following prompt: “How confident are you that you selected the correct figures?”. After moving the slider to make their CJ, they were instructed to get ready for the next trial and following a 1 s delay the procedure above repeated until all trials were completed. Participants made global CJs before and after completing the Spatial Relations test in the same manner as the previous task.

**2.2.6. Revised Purdue Spatial Visualization Test.**—An altered version of the Revised Purdue Spatial Visualization Test (PSVT:R) was administered to examine monitoring spatial orientation for complex 3-dimensional (3-D) objects (Guay, 1977; Branoff, 2000). The original task consists of 30 trials where participants view a 3-D object oriented in an initial position and then rotated on its x, y, or z plane into a different orientation. A third object is presented and participants must choose among 5 response options, the option that presents the new object rotated in an identical manner to the reference object. Before the response options were presented, participants rated how confident they were that they could identify the correct figure using the same scale and slider method used to rate confidence in the Spatial Relations test and Paper Folding test. After making this CJ, the 5 response options were presented and participant selected the option that they believe was the target figure rotated in the position specified by the reference object. After completing all 30 trials, participants, made a global CJ in the same manner as the previous task. Participants made global CJs before and after completing the PSVT:R in the same manner as the previous task.

### 3. Results

All analyses were preplanned except for analyses of the Spatial Experience Questionnaire. In all cases, analyses were exploratory because specific predictions were not made about the nature of sex differences that might be encountered for each measure. Unless noted otherwise, conclusions about sex differences were determined by examining confidence intervals of effects sizes (Cohen’s *d*).

#### 3.1. Spatial Reasoning and Visual Spatial Working Memory Measures.

Table 1 shows the mean performance for males and females on the Raven’s Progressive Matrices and Symmetry Span Task, and corresponding t-tests. Consistent with previous findings, male students performed significantly better on the symmetry span task than female students. There were no sex differences in performance on the Raven’s Progressive Matrices task.

### 3.2. Spatial Orientation and Spatial Visualization Measures.

Table 1 also shows male and female students' mean performance measures, confidence judgments, global predictions, and global postdictions for the Paper Folding test, Spatial Relations test, and the PSVT:R. There were no significant sex differences in performance on the Paper Folding test or the Spatial Relations test, but male students did outperform female students on the PSVT:R. Although females performed as well as male students on several spatial tasks, female students consistently predicted lower performance than male students on *all* spatial tasks. They generated lower mean CJs, global predictions, and global postdictions for each test, producing significant sex differences in every task and for each prediction type except for global predictions for the Paper Folding test.

### 3.3. Metacognitive Monitoring Accuracy.

Table 2 presents the mean absolute and relative accuracy of item-level CJs for male and female students for the spatial orientation and spatial visualization measures. Absolute accuracy of CJs were examined by computing the calibration component of Murphy's (1973) decomposition of a Brier score for each participant.<sup>1</sup> Calibration scores near zero reflect perfect absolute accuracy with increasing values reflecting deviations from perfect accuracy. Relative accuracy was examined by computing GoodmanKruskal gamma correlations between CJs and performance for each individual on each task (for rationale, see Gonzalez & Nelson, 1996; Nelson, 1984). Table 2 shows that male and female students were equally accurate in terms of both their absolute and relative accuracy for all spatial measures except for the relative accuracy of performance on the PSVT:R. Surprisingly, female students displayed higher relative accuracy on the PSVT:R than did male students. Apparently they were better able to discriminate correct from incorrect responses, even though males performed better on the test. These findings are inconsistent with previous research that indicates male students are more accurate at monitoring their performance in mental rotation tasks than female students (Cooke-Simpson & Voyer, 2006).

Absolute accuracy for students' global self-assessments of their performance on the same spatial visualization and spatial orientation measures was computed by subtracting each student's global predictions and global postdictions for each task from their actual performance (See Table 3). Values of zero reflect perfect absolute accuracy with positive deviations from zero reflecting overconfidence and negative values reflecting underconfidence. Table 3 shows that both male and female students were underconfident in their global predictions and postdictions for their performance on the Spatial Relations Test and PSVT:R but their estimates were well calibrated for the Paper Folding Test. There were no sex differences in the degree of underconfidence students displayed for either predictions or postdictions on the PSVT:R or for predictions of performance on the Spatial Relations Test. Absolute accuracy for global postdictions of performance for the Spatial Relations Test were significantly lower for females than male students which indicates that female students

---

<sup>1</sup>Group differences in absolute accuracy can be difficult to interpret when the magnitude of each group's metacognitive judgments are similar but performance differs or groups differ in performance but produce similar magnitudes of judgments (for rationale, see Connor, Dunlosky, & Hertzog, 1997). In such scenarios, group differences in measures of absolute accuracy may not reflect true differences in under or over confidence. In the current study, interpretation of absolute accuracy is not likely to be ambiguous because the above conditions did not occur for any of our spatial measures.



were more underconfident in their performance than male students after completing the Spatial Relations Test. There were no sex differences for either global predictions or postdictions for the Paper Folding test.

Relative accuracy for global predictions and postdictions were examined by correlating students' estimates of performance with their actual performance. Fisher r-to-z tests indicated that there were no sex differences in the relative accuracy for global prediction of performance for the Spatial Relations Test (Male:  $r = .22$ ,  $p = .01$ ; Female:  $r = .37$ ,  $p = .001$ ),  $Z = 1.2$ ,  $p = .23$ , Paper Folding Test (Male:  $r = .40$ ,  $p = .001$ ; Female:  $r = .29$ ,  $p = .01$ ),  $Z = .91$ ,  $p = .36$ , or PSVT:R (Male:  $r = .48$ ,  $p = .001$ ; Female:  $r = .56$ ,  $p = .001$ ),  $Z = 0.90$ ,  $p = .37$ . There were also no sex differences in relative accuracy for global postdictions of performance on the Spatial Relations Test (Male:  $r = .44$ ,  $p = .001$ ; Female:  $r = .60$ ,  $p = .001$ ),  $Z = 1.61$ ,  $p = .11$ , Paper Folding Test (Male:  $r = .71$ ,  $p = .001$ ; Female:  $r = .68$ ,  $p = .001$ ),  $Z = .42$ ,  $p = .68$ , or PSVT:R (Male:  $r = .67$ ,  $p = .001$ ; Female:  $r = .59$ ,  $p = .001$ ),  $Z = .97$ ,  $p = .32$ .

### 3.4. Correlations between Spatial Performance Measures.

The correlations between spatial performance measures for male (lower triangle) and female students (upper triangle with bolded values) are presented in Table 4. It shows that both female and male student performed consistently across each spatial measure, with positive moderate to high correlations between most tests. Fisher r-to-z tests testing sex differences in the magnitude of each correlation found no reliable effects.

### 3.5. Correlations between Measures of Monitoring Accuracy.

Table 5 shows the correlations between measures of absolute and relative accuracy of metacognitive judgments for male (lower triangles) and female (upper triangles with bolded values) students. Absolute accuracy measures for each task were significantly positively correlated for male students. The same pattern was present for female students with the exception that their absolute accuracy for the Paper Folding test and PSVT:R were not significantly correlated. However, Fisher r-to-z tests indicated that there were no significant differences between the magnitudes of the absolute accuracy correlations for male or female students. Relative accuracy measures for the Spatial Relations test and the Paper Folding test were also significantly positively correlated. No other correlations between relative accuracy measures were significant and no sex differences were present for these correlations. Correlations between the absolute and relative accuracy measures were not significant with the exception that female student's relative and absolute accuracy for the PSVT:R was significantly negatively correlated. A Fisher r-to-z test indicated that this negative correlation for female students was significantly different from the correlation for male students,  $Z = 2.31$ ,  $p < .05$ . The discrepancies between absolute and relative accuracy measures confirm that these different types of accuracy reflect unique aspects of monitoring ability.

### 3.7. Self-Perceptions of Everyday and Academic Spatial Ability.

Male and female students' mean self-reported recent experience, cumulative experience, and ability ratings for each of the specified spatial scenarios in the Spatial Experience Questionnaire are presented in Table A1 of the Appendix.

Most important for the current purposes are students' self-ratings of their abilities for everyday and academic spatial tasks. A confirmatory factor analysis on students' self-ratings was conducted using Mplus (Version 8) with full-information maximum likelihood (FIML) estimation under the assumption of multivariate normality of item responses. For ease of interpretation, only the standardized factor loadings and factor correlations are reported below. A two-factor model representing separate factors for perceived everyday spatial ability and perceived academic spatial ability provided a good fit to the data,  $\chi^2(61) = 107.54$ , RMSEA = .06, CFI = .95. Estimated factor correlation of the two factors was .46. Estimated factor loadings are presented in Table 6. The final model included residual covariances suggested by LaGrangian Multipliers between questions 7 and 8 and between questions 11 and 13. We also chose to load Question 6 from the original Spatial Experience Questionnaire on our academic spatial factor because conceptually producing and interpreting technical drawings or blueprints of 3-D objects is a skill utilized in engineering fields. An exploratory factor analysis also revealed that Question 13 fit better with the everyday spatial factor than the new academic spatial factor so we chose not to load this measure on the academic spatial factor.

Hierarchical multiple regression analyses were conducted to examine the relationship between students' self-perceptions of their everyday and academic spatial ability and their sex, and their actual spatial performance. To avoid multicollinearity between our performance measures which were highly correlated (Table 4), we created a spatial ability composite variable by standardizing and averaging across students' performance for Raven's Progressive Matrices, the Spatial Span task, the Paper Folding Test, the Spatial Relations Test, and the PSVT:R. First, we created a model with students' self-perceptions of their everyday spatial ability as the dependent variable where we entered sex and the spatial ability composite variable into the first step of the equation. During the second step we entered the Sex  $\times$  Spatial Ability interaction term.

The initial model predicted 8% of the variability in students self-perceptions of their everyday spatial ability,  $R^2 = .08$ , *adjusted*  $R^2 = .08$ ,  $F(3, 227) = 10.30$ ,  $p < .001$ ). Male and female students had similar self-perceptions of their everyday spatial ability ( $\beta = -.01$ ,  $p = .77$ ). Actual spatial ability was the only variable that significantly predicted self-perceptions of everyday spatial ability ( $\beta = .28$ ,  $p < .001$ ). The addition of the Sex  $\times$  Spatial Ability interaction ( $\beta = .08$ ,  $p = .37$ ) at the second step did not improve the model fit,  $F$ change (1, 226) = 0.82,  $p = .37$ .

Next, we computed another hierarchical regression analysis with the same predictors above entered in the same order but with students' self-perceptions of their academic spatial ability as the dependent variable. The initial model predicted 13% of the variability in students self-perceptions of their academic spatial ability,  $R^2 = .13$ , *adjusted*  $R^2 = .12$ ,  $F(3, 227) = 16.17$ ,  $p < .001$ . Students' sex ( $\beta = -.23$ ,  $p < .001$ ) and their actual spatial ability both significantly predicted their self-perceptions of academic spatial ability ( $\beta = .23$ ,  $p < .01$ ). Female students believed their academic spatial ability was lower than male students, and students' with higher spatial ability believed they were better at reasoning spatially in academic domains than student's lower in spatial ability. The addition of the Sex  $\times$  Spatial

Ability interaction ( $\beta = -.07, p = .40$ ) did not improve the model,  $F_{\text{change}}(1, 226) = 0.71, p = .40$ .

#### 4. Discussion

The current study examined sex differences in self-perceptions of everyday and academic spatial ability, and sex differences in metacognitive monitoring accuracy for measures of spatial visualization (Paper Folding Test) and spatial orientation (Spatial Relations Test and PSVT:R). Across multiple spatial measures, female students displayed lower confidence in their item-level monitoring and global assessments of performance than did male students, even when no actual differences in spatial performance were present (e.g., Paper Folding Test and Spatial Relations Test). These findings are consistent with previous research indicating that females are less confident in their spatial ability than are male students (Cooke-Simpson & Voyer, 2007; Estes & Felker, 2012; Syzmanowicz & Furnham, 2011). Furthermore, female STEM majors had lower self-evaluations than their male counterparts of visuospatial abilities needed for scientific reasoning.

However, in contrast to previous research focused exclusively on mental rotation (Cooke-Simpson & Voyer, 2007; Estes & Felker, 2012), we found no evidence that female students' item-level assessments of their spatial performance were less accurate than their male counterparts (see Relative and Absolute accuracy in Table 2). Indeed, female students' CJs were even better at discriminating between correct and incorrect trials on the PSVT:R – a measure of 3-D mental rotation ability – than were male students' CJs (see relative accuracy in Table 2). These results contradict prior conclusions about impaired spatial monitoring accuracy for female students. It is possible that the sex differences in monitoring accuracy previously identified are unique to the Vandenberg and Kuse (1978) Mental Rotation Test (Cooke-Simpson & Voyer, 2007; Estes & Felker, 2012). The quality of the cues that people attend to when monitoring their performance can differ across tasks and sometimes within tasks as a function of strategy utilization (Mitchum & Kelly, 2010).

Since female students sometimes adopt different strategies than male students when performing the Mental Rotation Test (Allen & Hogeland, 1978; Miller & Santoni, 1986; Kail, Carter, & Pellegrino, 1979; Pena, Contreras, Shih, & Santacreu, 2008; Raabe, Hoger, & Delius, 2006), they probably also sample different cues when monitoring their performance on this task. Mitchum and Kelly (2010) reported that people who use a constructive matching strategy to solve Raven's Progressive Matrices problems are more accurate at monitoring their performance than students who use a response elimination strategy. Interestingly, male students are more likely to use constructive matching to solve problems on the Mental Rotation Test, whereas female students are more likely to use response elimination (Raabe, Hoger, & Delius, 2006). Thus, one might expect female students to have impaired monitoring accuracy on the Mental Rotation Test because of their preferred solution strategy.

The cues people attend to can also vary when monitoring is prospective vs. retrospective (Nelson & Narens, 1990). A prospective confidence procedure like the procedure we used to examine performance monitoring on the PSTV:R involves making a confidence judgment

without viewing potential response options. This procedure may have limited the cues available to students to item specific information regarding the complexity of the probe item and the cues associated with generating and manipulating one's spatial representation for that item (e.g, processing fluency, vividness, etc.). In contrast, retrospective confidence procedures which were used for the Spatial Relations test and Paper Folding test involve making CJs after selecting a response option. This procedure affords access to cues relevant to the decision process involved in selecting a response option. For example, students can base their CJs on whether the spatial representation they generate is present among potential response options or on the ease of excluding unlikely response options.

The current study was not designed to evaluate which cues students attended to when making CJs. However, the sex differences in relative accuracy we observed for the PSTV:R, suggest that the quality of the cues students used to monitor their performance may have differed for men and women. It is unclear whether these differences are task-specific or reflect sex differences in cue utilization during prospective monitoring that are less prevalent in retrospective monitoring tasks. Since previous research has focused exclusively on retrospective monitoring of spatial performance and prospective monitoring was examined with one task in the current study, more research is necessary to competitively evaluate these hypotheses.

An alternative explanation for why the current results diverged from previous findings involves differences in the characteristics of our sample compared to previous work. Our sample consisted primarily of STEM majors from a selective STEM-focused university. Both male and female students in this sample probably have above-average spatial ability and more spatial reasoning experience than samples examined in previous work. Males and females in this study also had similar amounts of spatial experience as reflected by their self-reported cumulative everyday (Females:  $M = 3.91$ ,  $SE = .35$ ; Males:  $M = 3.91$ ,  $SE = .31$ ) and academic spatial reasoning experience (Females:  $M = 3.69$ ,  $SE = .34$ ; Males:  $M = 3.48$ ,  $SE = .22$ ) on the Spatial Experience Questionnaire,  $t_s < 1$ . Sex differences in spatial experience can contribute to sex differences in spatial performance (Baenninger & Newcombe, 1995; Casey, 1996), perhaps because increased familiarity and practice with spatial processing improves the calibration of one's confidence in their spatial thinking (Estes and Felker, 2012). If so, the sex differences in spatial monitoring accuracy observed in previous research could be due to sex differences in experience performing spatial oriented tasks not characteristic of our sample.

It is also possible that the students in our sample were less susceptible to sex-related stereotypes that could adversely affect their metacognitive monitoring accuracy than students in other samples. Stereotype threat has been proposed as a mechanism to account for sex differences in science, mathematics, and spatial skills (McGlone & Aronson, 2006; Nguyen & Ryan, 2008; Ortner & Sieverding, 2008; Spencer, Steele, & Quinn, 1999). According to stereotype threat theory, engaging in spatially oriented task can activate negative sex-related stereotypes which lower female students' confidence causing increased stress, anxiety and ultimately performance decrements. These effects may be contingent on how much one selfidentifies with the target domain that is being threatened (Nguyen & Ryan, 2008; but see Zigerell, 2017). For example, highly math-identified women are less

susceptible to stereotype threat regarding math ability than women who are moderately math-identified (Nguyen & Ryan, 2008). Students in our sample – who are STEM majors and likely higher than average for both male and females on spatial cognitive ability – may also be more likely to self-identify highly with spatially oriented tasks. If so, they may not experience sex related stereotype threat that might impair monitoring when performing these tasks. Although this hypothesis is intriguing, it should be interpreted cautiously due to concerns about the robustness of stereotype threat effects and the recent evidence of publication bias in this literature (Flore & Wicherts, 2015; Zigerell, 2017).

Although we observed limited sex-related differences in monitoring accuracy in the current study, more substantial sex-related differences could be present in qualitatively different tasks that require dynamic spatial processing. Dynamic spatial tasks require people to continuously monitor performance and attend to multiple objects across time and space whereas static spatial tasks like the ones examined in the current study involve mentally rotating or spatially transforming a single object at a fixed time point (Hunt et al., 1988). The cognitive load required to both maintain and manipulate visual spatial information in working memory while simultaneously monitoring the quality of one's spatial processing may be too great for some participants in dynamic spatial domains. For example, older adults (age 60 to 82) have impaired spatial performance monitoring compared to younger adults (college aged) in dynamic spatial tasks like navigation but no age related differences are present for monitoring accuracy in static spatial tasks including the Mental Rotation Test, Spatial Relations Test, and Paper Folding Test (Ariel & Moffat, 2018). Given that female students have lower visual-spatial working memory spans than male students (Voyer, Voyer, & Saint-Aubin, 2017), they may be more susceptible to monitoring errors in dynamic spatial domains that they are not susceptible to in static spatial tasks.

Regardless, the current results show that even female students pursuing STEM degrees are less confident than male students in their ability to reason spatially about some STEM related content. It is unclear if these differences reflect true underconfidence or if they are due to actual differences in academic spatial ability. Research examining medical students' self-perceptions of their abilities indicates that female students have higher rates of neuroticism, general anxiety, and test related anxiety than male students which can cause them to doubt their abilities (Hojat et al, 1999; Hojat et al., 2003). Cooper, Krieg, and Brownell (2018) recently reported that men in an undergraduate physiology course were more likely than women to believe they were smarter than their classmates even when controlling for GPA. Underconfidence in females seems to persist even for medical school students; women who are nominally equally competent as men consistently rate their abilities lower than male medical students (Coutts & Rogers; 1999; Minter, Gruppen, Napolitano, & Gauger, 2003; Rees, 2003). These data suggest that underconfidence in STEM domains, and perhaps for spatial reasoning in these domains, may be pervasive even in high ability female students. Consistent with this underconfidence interpretation, female students were more underconfident than male students in their ability to evaluate their global performance for each of the spatial orientation tasks we examined. Specifically, the absolute accuracy for global postdictions of performance for both the spatial orientation tasks was worse for female than for male students although these sex differences were only statistically significant for the Spatial Relations Test (see Table 3).

In summary, the current study indicates that sex differences in global self-assessments of performance do not always coincide with sex differences in moment-to-moment spatial performance monitoring. Even though female students were in most cases less confident than male students in their general spatial ability, their trial-by-trial metacognitive monitoring accuracy was not impaired in either an absolute or relative sense. Thus, female students appear to have relatively accurate perceptions of their spatial performance for spatial orientation and spatial visualization tasks. Future research should evaluate the potential effects of spatial experience and general spatial ability on spatial monitoring accuracy to determine whether sex differences previously observed in the literature are due to these factors. Future research should also examine whether sex differences are present in other spatial domains; especially dynamic spatial tasks which may be more predictive of STEM related success than the static spatial tasks examined in the current study (Hegarty, 1992; Sanchez & Wiley, 2014).

## Acknowledgments

This research was supported in part by a Ruth L. Kirschstein National Research Service Award (NRSA) Institutional Research Training Grant from the National Institutes of Health (National Institute on Aging), Grant #5T32AG000175, and Georgia Tech's President's Undergraduate Research Award. A subset of these data were collected to fulfill an undergraduate senior thesis for Natalie Lembeck at Georgia Institute of Technology. We thank Alyssa Candelmo, Haley Landis, and Lee Martin Frazer for their assistance with data collection.

## Appendix

**Table A1.**

Mean response to each question on the spatial experience questionnaire as a function of sex.

	Females			Males		
	Recent Experience	Cumulative Experience	Ability Rating	Recent Experience	Cumulative Experience	Ability Rating
Q1: Imagining different arrangements of furniture or other objects	2.51 (.46)	3.58 (.44)	6.20 (.17)	2.70 (.32)	3.73 (.39)	6.21 (.13)
Q2: Considering how an object or building would look from a different viewing position	2.59 (.40)	3.75 (.47)	5.85 (.18)	2.70 (.35)	4.00 (.41)	6.33 (.14)
Q3: Devising efficient ways of packing or loading a box or car trunk	3.19 (.46)	4.65 (.48)	6.99 (.15)	2.49 (.30)	4.44 (.41)	6.95 (.13)
Q4: Following instructions for the assembly of furniture, toys, models, and so on	2.41 (.37)	5.22 (.56)	6.96 (.21)	2.12 (.32)	6.07 (.48)	7.18 (.15)
Q5: Visualizing travel directions from a verbal description	3.30 (.40)	3.92 (.38)	5.70 (.19)	3.66 (.61)	4.00 (.35)	5.79 (.16)
Q6: Producing or interpreting technical drawings (e.g., blueprints) of threedimensional objects	3.11 (.64)	2.17 (.40)	5.14 (.23)	2.43 (.39)	2.11 (.24)	5.55 (.18)
Q7: Performing paper-folding activities such as origami	.69 (.15)	3.15 (.46)	5.23 (.24)	.40 (.08)	2.80 (.37)	4.53 (.19)
Q8: Solving piece-assembly games such as jigsaw puzzles	1.66 (.30)	6.74 (.63)	6.61 (.20)	1.11 (.18)	5.10 (.48)	6.20 (.14)



	Females			Males		
	Recent Experience	Cumulative Experience	Ability Rating	Recent Experience	Cumulative Experience	Ability Rating
Q9: Working on spatialmanipulation puzzles like Rubik's Cube	.54 (.11)	2.02 (.32)	3.67 (.21)	1.16 (.18)	2.96 (.36)	4.96 (.18)
*Q10: Imagining mathematical relationships (e.g., 3-D objects in calculus or otherwise)	7.85 (1.50)	4.90 (.52)	6.20 (.20)	7.07 (.90)	5.04 (.35)	7.17 (.15)
*Q11: Imagining micro-level concepts in biology or chemistry (the process of transcription, organic molecules, etc.)	5.69 (1.14)	3.99 (.52)	5.83 (.22)	3.81 (.69)	3.05 (.26)	5.69 (.18)
*Q12: Imagining concepts in physics (momentum, force, electrical current, etc.)	8.41 (1.58)	3.68 (.36)	6.11 (.20)	9.83 (3.34)	3.63 (.25)	6.85 (.14)
*Q13: Visualizing location and direction in anatomy or physiology (parts of the brain and body)	3.63 (.68)	2.20 (.27)	5.72 (.19)	2.74 (.46)	2.19 (.26)	5.02 (.17)

Note. Standard error of the means are in parenthesis.

\* denotes new STEM related spatial questions added to questionnaire.

## References

- Ariel R, & Moffat S (2018). Age-related similarities and differences in monitoring spatial cognition. *Aging, Neuropsychology, & Cognition*, 25, 351–377.
- Allen MJ, & Hogeland R (1978). Spatial problem-solving strategies as functions of sex. *Perceptual and Motor Skills*, 47, 348–350. [PubMed: 724374]
- Baenninger M, & Newcombe N (1995). Environmental input to the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences*, 7(4), 363–379.
- Branoff TJ (2000). Spatial Visualization Measurement: A Modification of the Purdue Spatial Visualization Test-Visualization of Rotations. *Engineering Design Graphics Journal*, 64, 14–22.
- Burnett SA, Lane DM, & Dratt LM (1979). Spatial visualization and sex differences in quantitative ability. *Intelligence*, 3, 345–354.
- Carraher TN, Carraher DW, & Schliemann AD (1985). Mathematics in the streets and in schools. *British Journal of Developmental Psychology*, 3, 21–29.
- Carroll JB (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Casey MB (1996). Understanding individual differences in spatial ability within females: A nature/nurture interactionist framework. *Developmental Review*, 16, 241–260.
- Connor LT, Dunlosky J, & Hertzog C (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*, 12, 50–71. [PubMed: 9100268]
- Cooke-Simpson A, & Voyer D (2007). Confidence and gender differences on the mental rotation test. *Learning and Individual Differences*, 17, 181–186.
- Cooper KM, Krieg A, & Brownell SE (2018). Who perceives they are smarter? Exploring the influence of student characteristics on student academic self-concept in physiology. *Advances in Physiology Education*, 42, 200–208. [PubMed: 29616569]
- Coutts L, & Rogers J (1999). Predictors of student self-assessment accuracy during a clinical performance exam: comparisons between over-estimators and under-estimators of SPEvaluated performance. *Academic Medicine*, 74, S128–30. [PubMed: 10536616]

- Dunlosky J, & Tauber SK (2014). Understanding people's metacognitive judgments: An isomechanism framework and its implications for applied and theoretical research In Perfect T & Lindsay S (Eds). *Handbook of Applied Memory* (pp. 444–464). Sage: Thousand Oaks, CA.
- Estes Z, & Felker S (2012). Confidence mediates the sex difference in mental rotation performance. *Archives of Sexual Behavior*, 41, 557–570. [PubMed: 22130691]
- Flore PC, & Wicherts JM (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, 53(1), 25–44. [PubMed: 25636259]
- French JW, Ekstrom RB, & Price LA (1963). *Manual for a kit of factor referenced tests for cognitive factors*, 1–122.
- Goldstein D, Haldane D, & Mitchell C (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition*, 18, 546–550. [PubMed: 2233267]
- Gonzalez R, & Nelson TO (1996). Measuring ordinal association in situations that contain tied scores. *Psychological Bulletin*, 119, 159–165. [PubMed: 8559859]
- Guay RB (1977). *Purdue spatial visualization test – Visualization of rotations*. West Lafayette, Indiana: Purdue Research Foundation.
- Halpern DF, & Collaer ML (2005). *More Than Meets the Eye The Cambridge handbook of visuospatial thinking* (pp. 170–112), New York, NY: Cambridge University Press.
- Hegarty M (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1084.
- Hegarty M, & Waller D (2005). Individual differences in spatial abilities In Shah P (Ed.) & Miyake A, *The Cambridge Handbook of Visuospatial Thinking* (pp. 121–169). New York, NY, US: Cambridge University Press.
- Herlitz A, Nilsson LG, & Bäckman L (1997). Gender differences in episodic memory. *Memory & Cognition*, 25, 801–811. [PubMed: 9421566]
- Herlitz A, & Rehnman J (2008). Sex differences in episodic memory. *Current Directions in Psychological Science*, 17, 52–56.
- Hertzog C, Dixon RA, & Hultsch DF (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging*, 5, 215–227. [PubMed: 2378687]
- Hojat M, Glaser K, Xu G, Veloski JJ, & Christian EB (1999). Gender comparisons of medical students' psychosocial profiles. *Medical Education*, 33, 342–349. [PubMed: 10336769]
- Hojat M, Gonnella JS, Erdmann JB, & Vogel WH (2003). Medical students' cognitive appraisal of stressful life events as related to personality, physical well-being, and academic performance: A longitudinal study. *Personality and Individual Differences*, 35, 219–235.
- Hunt E, Pellegrino JW, Frick RW, Farr SA, & Alderton D (1988). The ability to reason about movement in the visual field. *Intelligence*, 12(1), 77–100.
- Irwing P, & Lynn R (2005). Sex differences in means and variability on the progressive matrices in university students: A meta analysis. *British Journal of Psychology*, 96, 505–524. [PubMed: 16248939]
- Kail R, Carter P, & Pellegrino J (1979). The locus of sex differences in spatial ability. *Perception & Psychophysics*, 26, 182–186.
- Koriat A (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370.
- Kozhevnikov M, Motes MA, & Hegarty M (2007). Spatial visualization in physics problem solving. *Cognitive Science*, 31, 549–579. [PubMed: 21635308]
- Lave J (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. New York: Cambridge University Press.
- Lichtenstein S, & Fischhoff B (1981). The effects of gender and instructions on calibration Decision Research Report, 81–5. Eugene, OR: Decision Research.
- Lundeberg MA, Fox PW, & Pun coha J (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, 86, 114–121.
- Lynn R, & Irwing P (2004). Sex differences on the progressive matrices: A metaanalysis. *Intelligence*, 32, 481–498.

- Lohman DF (1986). The effect of speed-accuracy tradeoff on sex differences in mental rotation. *Perception & Psychophysics*, 39, 427–436. [PubMed: 3748756]
- Maeda Y & Yoon SY (2013). A meta-analysis on gender differences in mental rotation ability measured by the purdue spatial visualization tests: visualization of rotations (PSVT:R). *Educational Psychology Review*, 25, 69–94.
- Maloney EA, Waechter S, Risko EF, & Fugelsang JA (2012). Reducing the sex difference in math anxiety: The role of spatial processing ability. *Learning and Individual Differences*, 22, 380–384.
- McGlone MS, & Aronson J (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, 27(5), 486–493
- Michael WB, Guilford JP, Fruchter B, & Zimmerman WS (1957). The description of spatial-visualization abilities. *Educational and Psychological Measurement*, 17, 185–199.
- Mitchum AL, & Kelley CM (2010). Solve the problem first: Constructive solution strategies can influence the accuracy of retrospective confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 699–710.
- Miller LK, & Santoni V (1986). Sex differences in spatial abilities: Strategic and experiential correlates. *Acta Psychologica*, 62, 225–235. [PubMed: 3766198]
- Minter RM, Gruppen LD, Napolitano KS, & Gauger PG (2005). Gender differences in the self-assessment of surgical residents. *The American Journal of Surgery*, 189, 647–650. [PubMed: 15910713]
- Murphy AH (1973). A new vector partition of the probability score. *Journal of Applied Meteorology*, 12, 595–600.
- Nelson TO (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109–133. [PubMed: 6544431]
- Nelson TO (1990). Metamemory: A theoretical framework and new findings In Bower G (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 26, pp. 125–173). New York: Academic Press.
- Newcombe NS (2016). Thinking spatially in the science classroom. *Current Opinion in Behavioral Sciences*, 10, 1–6.
- Nguyen HHD, & Ryan AM (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of applied psychology*, 93(6), 1314. [PubMed: 19025250]
- Ortner TM, & Sieverding M (2008). Where are the gender differences? Male priming boosts spatial skills in women. *Sex Roles*, 29, 274–281.
- Orion N, Ben-Chaim D, & Kali Y (1997). Relationship between earth-science education and spatial visualization. *Journal of Geoscience Education*, 45, 129–132.
- Ortner TM, & Sieverding M (2008). Where are the gender differences? Male priming boosts spatial skills in women. *Sex Roles*, 59(3–4), 274–281.
- Oswald FL, McAbee ST, Redick TS, & Hambrick DZ (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, 47(4), 1343–1355. [PubMed: 25479734]
- Peña D, Contreras MJ, Shih PC, & Santacreu J (2008). Solution strategies as possible explanations of individual and sex differences in a dynamic spatial task. *Acta Psychologica*, 128(1), 1–14. [PubMed: 17977506]
- Prinzel LJ, & Freeman FG (1995). Sex differences in visuo-spatial ability: Task difficulty, speed-accuracy tradeoff, and other performance factors. *Canadian Journal of Experimental Psychology*, 49(4), 530. [PubMed: 9183990]
- Pribyl JR, & Bodner GM (1987). Spatial ability and its role in organic chemistry: A study of four organic courses. *Journal of Research in Science Teaching*, 24(3), 229–240.
- Raabe S, Höger R, & Delius JD (2006). Sex differences in mental rotation strategy. *Perceptual and Motor Skills*, 103, 917–930. [PubMed: 17326523]
- Raven JC, Raven JE, & Court JH (1998). *Progressive matrices*. Oxford England: Oxford University Press.
- Rees C (2003). Self assessment scores and gender. *Medical Education*, 37(6), 572–573.

- Reeve RA, Palincsar AS, & Brown AL (1987). Everyday and academic thinking: implications for learning and problem solving. *Journal of Curriculum Studies*, 19, 123–133.
- Salthouse TA & Mitchell DR (1990). Effects of age and naturally occurring experience on spatial visualization performance. *Developmental Psychology*, 26, 845–854.
- Sanchez CA & Wiley J (2014). The role of dynamic spatial ability in geoscience text comprehension. *Learning and Instruction*, 31, 33–45.
- Schraw G, & Nietfeld J (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology*, 90, 236–248.
- Schraw G, Dunkle ME, Bendixen LD, & Roedel TD (1995). Does a general monitoring skill exist? *Journal of Educational Psychology*, 87, 433–444.
- Schwartz BL, Benjamin AS, & Bjork RA (1997). The inferential and experiential basis of metamemory. *Current Directions in Psychological Science*, 6, 132–137.
- Spencer SJ, Steele CM, & Quinn DM (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28.
- Stanovich KE, & Cunningham AE (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology*, 85, 211–229.
- Syzmanowicz A, & Furnham A (2011). Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta analyses. *Learning and Individual Differences*, 21, 493–504.
- Tapley SM, & Bryden MP (1977). An investigation of sex differences in spatial ability: Mental rotation of three-dimensional objects. *Canadian Journal of Psychology*, 31, 122–130. [PubMed: 922592]
- Thomas AK, Bonura BM, Taylor HA, & Brunyé TT (2012). Metacognitive monitoring in visuospatial working memory. *Psychology and Aging*, 27(4), 1099–1110. [PubMed: 22663158]
- Thurstone LL, & Thurstone TL (1947). *Primary Mental Abilities*. Chicago, IL: Science Research Associates.
- Uttal DG & Cohen CA (2012). Spatial thinking and STEM education: when, why, and how? *Psychology of Learning and Motivation*, 57, 147–178.
- van Garderen D (2006). Spatial visualization, visual imagery, and mathematical problem solving of students with varying abilities. *Journal of learning disabilities*, 39, 496–506. [PubMed: 17165617]
- Vandenberg SG, & Kuse AR (1978). Mental rotations, a group test of three-dimensional spatial visualization, *Perceptual and Motor Skills*, 47, 599–604. [PubMed: 724398]
- 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. [PubMed: 17546728]
- Voyer D, Voyer S, & Bryden MP (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117, 250270.
- Voyer D, Voyer SD, & Saint-Aubin J (2017). Sex differences in visual-spatial working memory: a meta-analysis. *Psychonomic Bulletin & Review*, 24, 307–334. [PubMed: 27357955]
- Wai J, Lubinski D, & Benbow CP (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101, 817–835.
- Zigerell LJ (2017). Potential publication bias in the stereotype threat literature: Comment on Nguyen and Ryan (2008). *Journal of Applied Psychology*, 102, 1159–1168. [PubMed: 28795832]

### Highlights

- Female students are less confident than male students in their spatial performance.
- Women can monitor the accuracy of their spatial performance as well as men.
- Lower spatial confidence for women persists in academic domains for STEM majors.

**Table 1.**

Sex Differences in mean performance accuracy, confidence judgments (CJs), global predictions, global postdictions, with corresponding independent sample t-tests.

	Females	Males	<i>t</i>	df	<i>p</i>	<i>d</i>	95% CI	
							Lower	Upper
<i>Raven's Matrices</i>								
Performance	8.70 (.35)	8.99 (.27)	0.67	230	.50	0.09	-0.17	0.36
<i>Symmetry Span</i>								
Performance	15.06 (.55)	18.10 (.39)	4.62	230	<.001	0.62	0.35	0.89
<i>Paper Folding</i>								
Performance	71.69 (1.58)	75.38 (1.31)	1.87	230	.06	0.25	-0.01	0.52
CJs	73.58 (1.56)	78.49 (1.31)	2.38	230	.02	0.32	0.05	0.59
Global Prediction	71.24 (1.66)	75.96 (1.49)	2.05	230	.04	0.28	0.01	0.54
Global Postdiction	68.33 (2.10)	74.78 (1.73)	2.35	230	.02	0.32	0.05	0.58
<i>Spatial Relations</i>								
Performance	94.00 (.91)	95.90 (.49)	1.99	230	.05	0.27	0.003	0.53
CJs	75.13 (1.70)	81.79 (1.32)	3.11	230	.002	0.42	0.15	0.69
Global Prediction	72.44 (1.71)	77.83 (1.30)	2.53	230	.01	0.34	0.07	0.61
Global Postdiction	71.01 (1.95)	80.95 (1.39)	4.73	230	<.001	0.64	0.37	0.91
<i>PSVT:R</i>								
Performance	59.78 (2.40)	71.71 (1.78)	4.04	230	<.001	0.55	0.28	0.81
CJs*	54.79 (2.10)	65.31 (1.67)	3.91	230	<.001	0.53	0.26	0.80
Global Prediction	51.45 (2.22)	62.15 (1.92)	3.57	230	<.001	0.48	0.21	0.75
Global Postdiction	43.49 (2.48)	58.19 (1.95)	4.67	230	<.001	0.63	0.36	0.90

*Note.* Standard errors of the means are in parenthesis.

\* denotes that confidences judgment was made in a prospective fashion. 95% CIs reflect confidence intervals for Cohen's *d*.



**Table 2.**

Mean absolute accuracy of confidence judgments, relative accuracy of confidence judgments as corresponding independent sample t-tests.

	Females	Males	<i>t</i>	df	<i>p</i>	<i>d</i>	95% CI	
							Lower	Upper
<i>Absolute Accuracy</i>								
Spatial Relations	.07 (.01)	.06 (.01)	0.91	175	.37	0.06	-0.24	0.35
Paper Folding	.07 (.01)	.07 (.01)	0.64	230	.52	0.08	-0.19	0.34
PSVT:R *	.08 (.01)	.08 (.01)	0.24	230	.81	0.03	-0.23	0.30
<i>Relative Accuracy</i>								
Spatial Relations	.37 (.06)	.34 (.06)	0.36	183	.72	0.05	-0.24	0.35
Paper Folding	.41 (.05)	.38 (.04)	0.56	213	.58	0.08	-0.20	0.35
PSVT:R *	.44 (.04)	.23 (.04)	3.41	227	.001	0.46	0.19	0.73

Note. Standard errors of the means are in parenthesis. Absolute accuracy = calibration component (also known as reliability) of Murphy's decomposition of the brier score computed for each student. Relative accuracy reflects gamma correlation computed for each student.

\* denotes that confidences judgment was made in a prospective fashion. 95% CIs reflect confidence intervals for Cohen's *d*.

**Table 3.**

Mean absolute accuracy of global self-assessments of spatial performance and corresponding independent sample t-tests.

	Females	Males	t	df	p	d	95% CI Lower Upper
<i>Spatial Relations</i>							
Global Prediction	-21.56 (1.65)	-18.06 (129)	1.69	230	.09	0.23	-0.04 0.49
Global Postdiction	-22.99 (1.58)	-14.94 (125)	3.99	230	<.001	0.54	0.27 0.81
<i>Paper Folding</i>							
Global Prediction	-0.45 (1.93)	0.57 (1.49)	-0.42	230	.68	0.06	-0.17 0.36
Global Postdiction	-3.36 (1.55)	-.61 (1.20)	1.40	230	.16	0.19	-0.08 0.45
<i>PSVT:R</i>							
Global Prediction	-8.33 (2.22)	-9.55 (1.90)	0.42	230	.68	-0.06	-0.32 0.21
Global Postdiction	-16.28 (2.22)	-13.52 (152)	1.06	230	.29	0.14	-0.12 0.41

Note. Standard errors of the means are in parenthesis. Absolute accuracy = difference between estimates and performance. Scores of zero indicate perfect absolute accuracy. Negative values reflect underconfidence and positive value reflect overconfidence. 95% CIs reflect confidence intervals for Cohen's d.

**Table 4.**

Correlations between performance measures for females (upper triangle bolded values) and males (lower triangle).

	<b>Raven's Progressive Matrices</b>	<b>Symmetry Span</b>	<b>Paper Folding</b>	<b>Spatial Relations</b>	<b>PSVT:R</b>
Raven's Progressive Matrices	—	<b>43***</b>	<b>47***</b>	<b>42***</b>	<b>44*</b>
Symmetry Span	35***	—	<b>43***</b>	<b>39***</b>	<b>.26***</b>
Paper Folding	.55***	.43***	—	<b>59***</b>	<b>52***</b>
Spatial Relations	41***	37***	.47***	—	<b>53***</b>
PSVT:R	52***	.37***	59***	67***	—

Note.

\* denotes correlation is significant at  $p < .05$ .

\*\* denotes  $p < .01$ .

\*\*\* denotes  $p < .001$ .

**Table 5.**

Correlations between item level monitoring accuracy measures for females (upper triangle and bolded text) and males (lower triangle).

	Absolute Accuracy			Relative Accuracy		
	Paper Folding	Spatial Relations	PSVT:R	Paper Folding	Spatial Relations	PSVT:R
<i>Absolute Accuracy</i>						
Paper Folding	—	<b>.34**</b>	<b>.16</b>	<b>.13</b>	<b>.10</b>	<b>.01</b>
Spatial Relations	.49***	—	<b>.25*</b>	<b>.04</b>	<b>-.08</b>	<b>.11</b>
PSVT:R	.35***	.42***	—	<b>-.01</b>	<b>-.02</b>	<b>-.23*</b>
<i>Relative Accuracy</i>						
Paper Folding	-.06	-.13	-.02	—	<b>.35***</b>	<b>.21</b>
Spatial Relations	-.03	-.01	.08	.48***	—	<b>.11</b>
PSVT:R	.02	-.09	.09	.01	.03	—

Note.

\* denotes correlation is significant at  $p < .05$ .

\*\* denotes  $p < .01$ .

\*\*\* denotes  $p < .001$ .

**Table 6.**

## Factor Loadings for the Two Factor Model for Perceived Spatial Ability

Item	Description	Everyday Spatial Ability	Academic Spatial Ability	R <sup>2</sup>
Q1	Imagining different arrangements of furniture or other objects	.77		.59
Q2	Considering how an object or building would look from a different viewing position	.67		.45
Q3	Devising efficient ways of packing or loading a box or car trunk	.63		.40
Q4	Following instructions for the assembly of furniture, toys, models, and so on	.67		.45
Q5	Visualizing travel directions from a verbal description	.40		.16
Q6	Producing or interpreting technical drawings (e.g., blueprints) of three-dimensional objects	.28	.40	.34
Q7	Performing paper-folding activities such as origami	.34		.12
Q8	Solving piece-assembly games such as jigsaw puzzles	.59		.34
Q9	Working on spatial-manipulation puzzles like Rubik's Cube		.48	.23
Q10	Imagining mathematical relationships (e.g., 3-D objects in calculus or otherwise)		.81	.62
Q11	Imagining micro-level concepts in biology or chemistry (the process of transcription, organic molecules, etc.)		.30	.09
Q12	Imagining concepts in physics (momentum, force, electrical current, etc.)		.76	.57
Q13	Visualizing location and direction in anatomy or physiology (parts of the brain and body)	.38		.14