Title

Trial history influences the malleability of gender differences in children’s mental rotation performance

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors

Liu, Yaxin

Lourenco, Stella

Publication Date

2023

Peer reviewed

Trial history influences the malleability of gender differences in children's mental rotation performance

Yaxin Liu (yaxin.liu@emory.edu)
Department of Psychology, Emory University
Atlanta, GA, 30322 USA

Stella F. Lourenco (stella.lourenco@emory.edu)
Department of Psychology, Emory University
Atlanta, GA, 30322 USA

Abstract

Despite accumulating evidence of gender differences in mental rotation performance, much remains unknown about the variables that lead to an advantage among boys compared to girls. Here we examined the role of trial history on children's performance. To this end, we manipulated the difficulty of trials and implemented drift diffusion modeling (DDM) to assess how prior exposure to easy versus hard trials affects the parameters of drift rate and decision threshold. In Experiment 1, children were presented with either an easy-to-hard or a hard-to-easy block order. On easy trials, there were no gender differences in accuracy or drift rates, regardless of order. On hard trials, girls matched boys in accuracy and drift rates, but only when easy trials were presented first, suggesting that girls' performance on hard trials benefited from prior exposure to easy trials. In Experiment 2, we ruled out a general practice effect, confirming that improvement in girls' performance is specific to exposure on easy trials prior to hard trials, not just more trials. Additionally, boys, in general, had larger decision thresholds than girls. Taken together, these findings point to gender differences in mental rotation performance that are dependent on trial history and that may reflect differences in affective and/or motivational factors between boys and girls.

Keywords: mental rotation; development; spatial cognition; gender differences; drift diffusion modeling

Introduction

One of the largest gender differences in cognition occurs on mental rotation tasks (Hyde, 2014). These tasks present objects from different orientations, and participants judge whether the objects can be rotated into congruence with one another (Shepard & Metzler, 1971). On average, men and boys outperform women and girls[1] from 6 years of age, with gender differences increasing over development (Lauer et al., 2019; Voyer et al., 1995). However, these differences vary by task characteristics, such as stimulus type (Rahe et al., 2021), stimulus complexity (Heil & Jansen-Osmann, 2008), task format (Voyer et al., 1995), and task difficulty (Collins & Kimura, 1997; Voyer & Hou, 2006).

In the present study, we consider task difficulty in particular because it has been found to affect mental rotation performance and to modulate the gender difference. For example, Collins and Kimura (1997) reported a male advantage among adult participants, but only on difficult, not easy, versions of the mental rotation tasks. More recently, research has found that task difficulty interacts with trial history to impact performance. Adult participants performed better on a hard version of a mental rotation task when the hard version followed an easy version of the task (Rahe et al., 2019). Studies using non-spatial tasks have also demonstrated that a hard-to-easy order of trials reduces participants' optimism about their task performance (Weinstein & Roediger, 2012), suggesting a role for affective and/or motivational factors in explaining effects of trial history, at least in adult participants.

These findings raise a number of intriguing questions. First, is children's performance similarly impacted by trial history? Research on non-spatial tasks suggests that children's decision-making may be influenced by trial history (Odic et al., 2014), but it remains unclear how such decision processes affect mental rotation performance. Second, do boys and girls respond to trial history in a similar manner? If trial history impacts affective and/pr motivational states, then trial history is also likely to impact the performance of boys and girls differently.

On mental rotation tasks, decision-stage processes are closely tied to demands and judgments required by the tasks. Research has shown that both response strategy and effort allocation are involved in mental rotation tasks (Ben-Shachar & Berger, 2023; Campbell et al., 2018; Toth & Campbell, 2019). Response strategy involves setting an internal threshold for making decisions. As such, thresholds may be larger when one gathers more information before committing to a

[1] We acknowledge that gender can be non-binary. However, for consistency with existing literature on mental rotation, and to increase statistical power, we did not assess non-binary gender identity and we use sex/gender terms interchangeably (i.e., men/boys/males and women/girls/females) in this paper.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

decision, suggesting more conservative responding. Thus, this threshold often reflects response caution.

Guessing behaviors on mental rotation tasks differ between men and women (Cherney & Neff, 2004; Voyer et al., 2004), with greater variability observed in women than men (Heil et al., 2012). Such a gender difference in guessing behaviors may reflect differences in response caution. One piece of evidence supporting such a possibility comes from Liu and Lourenco (2022) who found that women's decision thresholds, as modeled by drift diffusion modeling, were smaller than those of men. In this context, smaller decision thresholds are typically interpreted as less caution.

Other work found that participants higher in motivation had larger decision thresholds, suggesting more response caution (Dix & Li, 2020). Although research on mental rotation has shown that motivation may moderate gender differences in adults' decision thresholds (Liu & Lourenco, 2022), it is unknown how children's decision thresholds may be affected, particularly as a function of trial history.

Sequential sampling models such as drift-diffusion models (DDMs) are especially suitable for disentangling distinct stages of processing (Forstmann et al., 2016; Ratcliff & McKoon, et al., 2008). Specifically, in the context of a mental rotation task, the decision stage processes that involve the amount of evidence accumulation for the judgment (indexed by decision threshold) can be modeled separately from the speed of rotational processing (indexed by drift rate).

Here we examined boys' and girls' performance in response to trial history—specifically, easy versus hard trials on a mental rotation task. In addition to accuracy and RT, we used DDMs to assess two specific parameters—drift rate (v) and decision threshold (a).

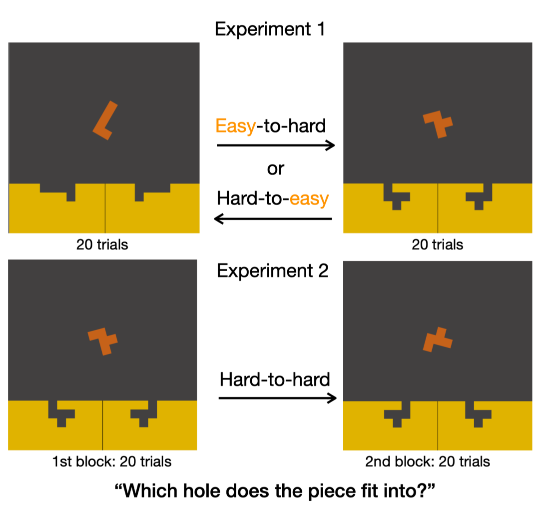

Figure 1: Illustrations of stimulus images and conditions used in Experiment 1 (top) and Experiment 2 (bottom).

Experiment 1

In a first experiment, boys and girls were randomly assigned to either an easy-to-hard (easy-first) or a hard-to-easy (hard-first) condition, such that children were exposed to a different trial history across conditions. We hypothesized that on easy trials (where gender differences are not expected), both boys' and girls' performance would be unaffected by trial history. By contrast, on hard trials (where gender differences are expected), we predicted a gender difference dependent on trial history, with girls showing more sensitivity, than boys, to the order of hard and easy trials.

Method

Participants

An a priori power analysis revealed a minimum sample of 56 participants using a 2-conditions by 2-genders between-subject design. We recruited 60 children ages 6 to 8 years (M_{age} = 7.11 years)—the age at which the male advantage has been found to be reliable (Lauer et al., 2019). Children's legal guardians provided written informed consent on behalf of their children. Families received an Amazon gift card upon completion of the experiment. The IRB approved all procedures.

Children were randomly assigned to either the easy-to-hard condition or hard-to-easy condition. Two participants were excluded because of inattention or technical difficulty. The final sample consisted of 58 participants (29 per condition; 51.7% girls).

Stimuli and Procedure

Children completed a 2AFC mental rotation task (Frick et al., 2013), in which they judged whether an object could be rotated into one of two silhouettes (see Fig. 1). Task stimuli consisted of 2D simple and complex shapes. The simple shape was L-shaped (concave hexagon) and the complex shape was an irregular 10-sided polygon (see Fig. 1). The experiment was programmed using PsychoPy3 and presented on *Pavlovia* (Peirce et al., 2019). All children were tested individually over Zoom.

Children were introduced to a game in which the goal was to help a character "fix the road," so that it could cross the road without falling in. Prior to the test trials, children were shown animated videos demonstrating that the target shape fit the hole. Children positioned their fingers on "F" and "J" keys for the left and right holes and were instructed to respond as quickly and as accurately as possible. There were two practice trials, each accompanied by corrective feedback. Each practice trial was repeated up to three times if children answered incorrectly.

There were 40 test trials with two levels of difficulty (20 easy trials, 20 hard trials). Easy trials consisted of the simple shape, with angular differences ranging between 0° and 60° (15° increments). Hard trials consisted of the complex shape, with angular differences ranging between 105° and 165° (15° increments). In the easy-to-hard condition, children completed a block of easy trials first and a block of hard trials second. In the hard-to-easy condition, children were given the reverse order (hard trials first, easy trials second). Trials were separated by an interval of 200 ms. Children were given no feedback on performance during the test trials.

Hierarchical Drift Diffusion Modeling (HDDM)

Task performance was modeled by fitting HDDM (Wiecki et al., 2013) to participant accuracy and RT data. This approach accounts for individual differences in parameter estimates. All models were fitted to individual participants' data using hierarchical Bayesian analysis.

Data Pre-processing. Data were trimmed by removing RTs less than 0.3 s and greater than 10 s (Exp. 1: 4.7% of total trials; Exp. 2: 8.7% of total trials). No data were excluded based on accuracy. Both correct and error trials were included in the analyses, as typically done with DDM. Trial-by-trial accuracy and RTs for each participant were inputted into HDDM.

Model Specifications. Parameter estimates consisted of drift rate (v) and decision threshold (a). HDDM regression analysis was performed to assess main effects and interactions. All models included 5000 sampling iterations and discarded the first 200 iterations of the chain to ensure good posterior estimation. Group mean posteriors were used to perform Bayesian statistical analyses.

Model Fits. Model convergence was checked by inspecting traces of model parameters and Gelman-Rubin convergence diagnostic statistics (Gelman & Rubin, 1992). Posterior predictive checks confirmed adequate model fits. Model parameters were analyzed using Bayesian hypothesis testing (95% Credible Intervals).

Results

Behavioral Performance: Accuracy and RTs[2]

Linear mixed models were fitted separately to accuracy and RTs, with condition (easy-to-hard vs. hard-to-easy), trial type (easy trials vs. hard trials), and gender as fixed factors, as well as interaction terms between condition, trial type, and gender. Participant intercept was included as a random factor. P-values for all pairwise comparisons were Bonferroni corrected.

Accuracy. Analyses revealed a significant condition by gender by trial type interaction (F [1, 520.97] = 9.50, p = .002), suggesting that gender differences in accuracy differed between condition and trial type (see Fig. 2). There was also a main effect of trial type (F [1, 520.97] = 160.07, $p < .001$) and a marginal effect of gender (F [1, 57.93] = 3.97, p = .051). To further understand the interaction effect, pairwise analyses were performed comparing boys and girls within each trial type (see subsequent sections).

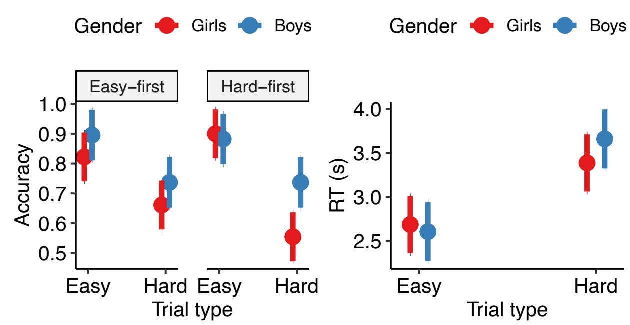

Figure 2: Mean accuracies and RTs by trial type. There was a condition by gender by trial type interaction in accuracy, and a gender by trial type interaction in RTs. Error bars represent 95% CIs.

Within-gender Comparisons. See Table 1 for statistical details. Within-gender comparisons were performed across conditions (easy-first vs. hard-first) for each difficulty level. In summary, both boys and girls showed better accuracy on easy versus hard trials, regardless of trial history ($ps < .01$). All other comparisons between the easy-to-hard and the hard-to-easy condition were not significant (but, see subsequent sections).

Table: 1: Within-gender Pairwise Comparisons (Easy-first vs. Hard-first)

Condition	Easy-first		Hard-first		Mean Difference	*t*	*p*	*d*
	Mean	*SD*	*Mean*	*SD*				
Boys								
Easy trials	.90	.20	.88	.17	.02	.21	>.999	.05
Hard trials	.74	.28	.74	.27	.00	.00	>.999	.00
Girls								
Easy trials	.82	.22	.90	.17	-.08	- 1.28	.819	.28
Hard trials	.66	.26	.55	.29	.11	1.75	.337	.38

[2] Preliminary analyses revealed that both accuracy and RTs varied according to angle. As angular differences increased, children's accuracy decreased (p = .012) and RTs increased (p = .001), consistent with research in adults (e.g., Shepard & Metzler, 1971) and children (Frick et al., 2013).

Between-gender Comparisons. See Table 2 for statistical details. Pairwise comparisons of boys' and girls' accuracy revealed that, on easy trials, boys' and girls' accuracies did not significantly differ between condition. On hard trials, however, girls' accuracy was significantly lower than boys' accuracy in the hard-to-easy condition. Moreover, and importantly, girls did not show significantly lower accuracy than boys in the easy-to-hard condition. These findings suggest that girls may have benefited from exposure to easy trials prior to hard trials.

Table: 2: Between-gender Pairwise Comparisons

Condition	Boys		Girls		Mean Difference	*t*	*p*	*d*
	Mean	*SD*	*Mean*	*SD*				
Easy-first								
Easy trials	.90	.20	.82	.22	.07	1.18	>.999	.26
Hard trials	.74	.28	.66	.26	.08	1.28	>.999	.28
Hard-first								
Easy trials	.88	.17	.90	.17	-.02	-.29	>.999	.06
Hard trials	.74	.27	.55	.29	.18	2.94	.03*	.64

RTs. Analyses revealed a significant main effect of trial type (F[1, 521] = 98.26, p < .001) and a significant gender by trial type interaction[3] (F[1, 521.06] = 4.01, p = .046), but not an interaction with condition (F[1, 521] < 1, p = .445). Posthoc analyses of the gender by trial type revealed that both boys ($M_{\text{difference}}$ = -1057 ms, t[525] = -.8.26, p < .001, d = .72) and girls ($M_{\text{difference}}$ = -702 ms, t[525] = -5.66, p < .001, d = .49) had faster RTs in the easy trials compared to the hard trials. Between-gender comparisons, however, revealed no differences between boys and girls on either easy ($M_{\text{difference}}$ = -81 ms, t[84] = -.33, p > .999, d = .07) or hard ($M_{\text{difference}}$ = 274 ms, t[84] = 1.10, p > .999, d = .24) trials (see Fig. 2).

HDDM analyses

Drift rate. Analyses revealed a positive regression coefficient of the interaction between condition and gender for the hard trials (100% posterior probability), but not the easy trials (51% posterior probability), indicating a condition by gender by trial type interaction (see Fig. 3). Posthoc within-gender comparisons on the hard trials revealed that, among girls, drift rates were significantly faster in the easy-to-hard than hard-to-easy condition (100% posterior probability); this was not the case for boys (61% posterior probability). Although analyses of accuracy did not reveal a difference between conditions (see Table 1), the increase in drift rate in the easy-to-hard condition compared to the hard-to-easy condition suggests that exposure to easy trials improves processing efficiency on hard trials, especially in girls. Between-gender analyses further revealed that boys had faster drift rates than girls on easy trials, regardless of condition. However, on the hard trials, boys had faster drift rates than girls in the hard-to-easy condition (100% posterior probability) but not the easy-to-hard condition (85% posterior probability). This finding suggests that gender differences in processing efficiency may be reduced by prior exposure to easy trials.

Decision threshold. Analyses revealed a marginally significant condition by gender interaction on the easy trials (93% posterior probability) and a significant interaction on the hard trials (95% posterior probability). Posthoc within-gender comparisons revealed that girls' decision thresholds on easy trials were larger in the hard-to-easy than easy-to-hard condition (see Fig. 3; 100% posterior probability). Boys showed a marginally significant difference (94% posterior probability). These findings suggest greater response caution on easy trials following exposure to hard trials.

Between-gender analyses revealed larger decision thresholds in boys than girls across all comparisons (> 95% posterior probabilities), consistent with adult findings showing larger decision thresholds in men than women (Liu & Lourenco, 2022). One exception occurred with easy trials in the hard-to-easy condition, which showed no gender difference (85% posterior probability). The lack of a gender difference in the hard-to-easy condition reflects an increase in decision threshold (compared to the easy-to-hard condition) by girls, but not boys, and suggests that gender differences in response caution may be altered by trial history (see General Discussion).

Together, the findings from this experiment suggest that girls benefited from prior exposure to easy trials, particularly with respect to drift rates, such that, on hard trials, drift rates increased following easy trials. One possible explanation for this effect is that girls benefit from a trial history that scaffolds them for hard trials, possibly by increasing motivation and/or reducing anxiety. However, another possibility is that the hard trials presented before easy trials may have led to a practice effect, independent of motivation or anxiety, especially for girls. Although a similar effect was not present on easy trials when hard trials were presented

[3] An analysis conducted on RTs for just the correct trials similarly showed a significant gender by trial type interaction (F[1, 511.49] = 4.73, p = .030).

first, which would seem to argue against this possibility, we conducted Experiment 2 with two blocks of hard trials to directly rule out a practice effect.

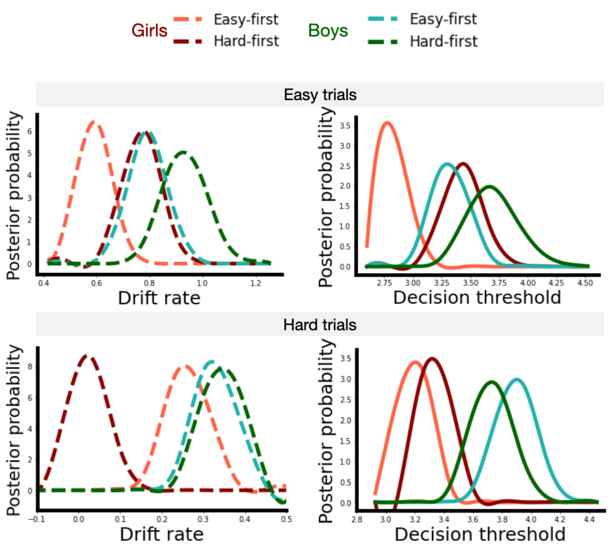

Figure 3: The gender by condition by trial type interaction for drift rate and decision threshold in Experiment 1 on easy and hard trials.

Experiment 2

As already noted, a concern about the previous findings is that, rather than a benefit of easy trials presented before hard trials, the order effect reflected general practice. To address this concern, we tested children in the same paradigm, but with two blocks of hard trials. If the effects in Experiment 1 were simply due to practice, then girls should show similar 'improvement' when given twice as many hard trials. However, if, instead, the effects were specifically related to prior exposure to easier trials, then prior exposure to hard trials should not lead to the same improvement. Moreover, this experiment allowed us to test whether repetitive exposure to hard trials might actually have adverse effects.

Method

Participants

We recruited 37 children between 6 and 8 years (M_{age} = 7.03 years) through the *Lookit* online platform (Scott & Schulz, 2017). As in Experiment 1, written informed consent was obtained by legal guardians and all procedures were IRB approved. Three children were excluded (1: failure to complete experiment; 2: failure to follow instructions). The final sample consisted of 34 children (17 boys and 17 girls).

Stimuli and Procedure

Children completed two blocks of hard trials (20 trials each). All other procedures were identical to Experiment 1.

Results

Behavioral Performance: Accuracy and RTs

Linear mixed models, with trial block (first vs. second block), gender, and an interaction of block by gender as fixed factors, were fitted separately to accuracy and RTs. Participant intercept was entered as a random factor.

Accuracy. Analyses revealed a significant effect of gender ($F[1, 33.88] = 4.54$, $p = .040$), such that boys scored higher than girls ($M_{difference} = .15$, $t[36] = 2.07$, $p = .046$, $d = .69$). However, there was no effect of block ($F[1, 304] < 1$, $p = .549$); nor was there a block by gender interaction ($F[1, 304] < 1$, $p = .604$), as might be expected if girls, in particular, benefited from practice.

RTs. Analyses revealed a significant effect of block ($F[1, 303.94] = 4.41$, $p = .037$), such that RTs were faster in the second block compared to the first block ($M_{difference}$ = -236 ms, $t[306] = -2.09$, $p = .037$, $d = .24$). But there was no effect of gender ($F[1, 33.89] < 1$, $p = .514$) nor block by gender interaction ($F[1, 304] = 2.31$, $p = .130$).

HDDM analyses

Drift rate. Analyses revealed an effect of gender (> 99% posterior probability), such that boys had faster drift rates than girls (see Fig. 4), but no effect of block (29% posterior probability) or block by gender interaction (46% posterior probability).

Decision threshold. Analyses revealed effects of gender (97% posterior probability) and block (99% posterior probability), as well as a block by gender interaction (> 99% posterior probability). Gender differences in decision thresholds were significant in both blocks (first: 98% posterior probability; second: > 99% posterior probability), such that boys had larger decision thresholds than girls. Posthoc within-gender comparisons revealed that whereas girls' decision thresholds were smaller in the second compared to the first block (> 99% posterior probability), boys showed no difference between blocks (80% posterior probability; see Fig. 4). Note that although girls showed a difference in decision thresholds across blocks, this effect contrasts with the hard-to-easy condition in Experiment 1 where girls increased their decision thresholds in the easy trials following the hard trials (see General Discussion).

Altogether, the findings from Experiment 2 suggest that gender differences in Experiment 1 were specific to trial history: namely, a benefit of an easy-to-hard block order for girls. When exposed to additional hard trials, an enhancement in girls' performance was not observed, contra an account based strictly on practice. Actually, prolonged exposure to hard trials may have had a negative effect on response caution, at least in the decision thresholds for girls. Decreased decision thresholds suggest reduced response caution with continued hard trials, possibly through decreased motivation from the trial history.

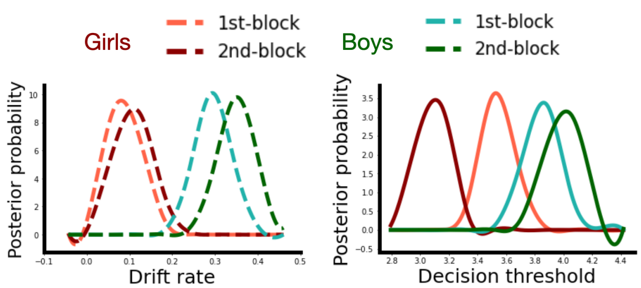

Figure 4: The gender by block interactions in drift rates and decision thresholds in Experiment 2.

General Discussion

This study examined how trial history affects boys' and girls' mental rotation task performance. Using DDM, our study sheds new light on the gender differences in children's rotational and decision-stage processes. Across two experiments, we found that trial history differentially impacted task performance among boys and girls. These findings suggest that gender differences in mental rotation are malleable to situational contexts, with a specific benefit, for girls, on hard trials following exposure to easy trials. The absence of a similar benefit when the number of hard trials was doubled suggests specific roles for motivational and/or affective factors linked to trial history, as modulated by difficulty.

Processing efficiency. Boys outperforming girls in behavioral accuracy has been well documented (Lauer et al., 2019). What is unknown is whether there are gender differences in processing efficiency, at least as indexed by drift rate in DDM. The present study provides novel modeling results showing that girls do not always have worse processing efficiency than boys. Instead, girls matched boys in drift rates (and accuracy) on hard trials when easy trials were presented first, possibly because easy trials enhanced performance by increasing motivation, previously found to increase drift rates in adults (Liu & Lourenco, 2022). In addition, easy trials may have decreased spatial anxiety, which has been found to be negatively associated with accuracy and processing efficiency. Prior work on spatial anxiety in mental rotation has also found that spatial anxiety may negatively affect performance in girls, but not boys (Ramirez et al., 2012). It is possible that girls, who exhibit higher levels of anxiety (on average) than boys, benefit particularly from exposure to easy trials, which may decrease anxiety. Conversely, hard trials may have increased anxiety in girls, and consequently, increased gender differences in processing efficiency. Yet neither motivation nor anxiety were directly tested in the present study. Thus, we urge caution in accepting this interpretation fully prior to additional research. Indeed, future research should address the specific influence(s) of motivation and/or anxiety on children's processing efficiency on mental rotation tasks.

Response caution. Another novel contribution of the present study is the finding that gender differences in response caution, at least as indexed by decision thresholds in DDM, varied as a function of trial history. Interestingly, whereas boys' decision thresholds were relatively unaffected by condition, girls showed more variability, with a decrease in decision thresholds following repeated exposure to hard trials. One explanation for this difference is that girls' decreased motivation from the tasks may have triggered an escape strategy, with accompanying smaller decision thresholds. It has been shown that anxiety can induce task avoidance (Choe et al., 2019); accordingly, repeated exposure to hard trials may be associated with increased task anxiety, which, in turn, may be linked to task avoidance. Notably, one exception of girls' smaller decision thresholds occurred when girls were exposed to hard trials before easy trials. If girls had lower response caution on the hard trials, then it is surprising that they regained response caution when trials switched to easy. However, a potential explanation for this is that the switch to easy trials may have made girls re-engage in the task, resulting in larger decision thresholds. The precise role of trial history on response caution remains to be tested. Future work may further dissociate the decision-stage processes to understand the role of affective and/or motivational factors on children's decision thresholds.

The present study demonstrates how trial history can affect spatial performance. Prior work has suggested that optimally challenging spatial toys are best for children's learning (Levine et al., 2012). Our own work further implies that tailored spatial experiences, even within a task, are likely to impact children's task performance, perhaps especially in girls.

Acknowledgements

This research was supported by Laney Graduate School Professional Development Research Funds. We thank Argie Dabrowski and Mrudhula Nithiyakumar for their assistance with data collection.

References

Ben-Shachar, M. S., & Berger, A. (2023). Visual-Spatial Abilities Are NOT Related to the Speed of Mental Rotation. https://psyarxiv.com/gkzxf/download?format=pdf

Campbell, M. J., Toth, A. J., & Brady, N. (2018). Illuminating sex differences in mental rotation using pupillometry. Biol Psychol, 138, 19-26. https://doi.org/10.1016/j.biopsycho.2018.08.003

Cherney, I. D., & Neff, N. L. (2004). Role of strategies and prior exposure in mental rotation. Perceptual and motor skills, 98(3_suppl), 1269-1282. https://journals.sagepub.com/doi/pdf/10.2466/pms.98.3c.1269-1282

Choe, K. W., Jenifer, J. B., Rozek, C. S., Berman, M. G., & Beilock, S. L. (2019). Calculated avoidance: Math anxiety predicts math avoidance in effort-based decision-making. Sci Adv, 5(11), eaay1062. https://doi.org/10.1126/sciadv.aay1062

Collins, D. W., & Kimura, D. (1997). A large sex difference on a two-dimensional mental rotation task. Behav Neurosci, 111(4), 845-849. https://doi.org/10.1037//0735-7044.111.4.845

Dix, A., & Li, S. C. (2020). Incentive motivation improves numerosity discrimination: Insights from pupillometry combined with drift-diffusion modelling. Sci Rep, 10(1), 2608. https://doi.org/10.1038/s41598-020-59415-3

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. Annual Review of Psychology, 67(1), 641-666. https://doi.org/10.1146/annurev-psych-122414-033645

Frick, A., Ferrara, K., & Newcombe, N. S. (2013). Using a touch screen paradigm to assess the development of mental rotation between 3½ and 5½ years of age. Cogn Process, 14(2), 117-127. https://doi.org/10.1007/s10339-012-0534-0

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. International Journal of Developmental Science, 2(1-2), 190-196. https://doi.org/10.3233/dev-2008-21212

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. https://doi.org/10.1016/j.lindif.2012.01.004

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65,* 373– 398. https://doi.org/10.1146/annurev-psych-010213-115057

Lauer, J. E., Yhang, E., & Lourenco, S. F. (2019). The Development of Gender Differences in Spatial Reasoning: A Meta-Analytic Review. Psychological Bulletin, 145(6), 537-565. https://doi.org/10.1037/bul0000191

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. https://doi.org/10.1037/a0025913

Liu, Y., & Lourenco, S. (2022). Affective Factors Affect Visuospatial Decision-making: A Drift Diffusion Modeling Approach. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44. Retrieved from https://escholarship.org/uc/item/9kd457qg

Lourenco, S. F., & Liu, Y. (2023). The Impacts of Anxiety and Motivation on Spatial Performance: Implications for Gender Differences in Mental Rotation and Navigation. Current Directions in Psychological Science, 0(0). https://doi.org/10.1177/09637214231153072

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. J Exp Psychol Gen, 143(1), 255-265. https://doi.org/10.1037/a0030825

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. Behavior Research Methods, 51(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Rahe, M., Ruthsatz, V., & Quaiser-Pohl, C. (2021). Influence of the stimulus material on gender differences in a mental-rotation test. Psychological Research, 85(8), 2892-2899. https://doi.org/10.1007/s00426-020-01450-w

Rahe, M., Ruthsatz, V., Schürmann, L., & Quaiser-Pohl, C. (2019). The effects of feedback on the gender differences in the performance in a chronometric mental-rotation test. Journal of Cognitive Psychology, 31(4), 1-9. https://doi.org/10.1080/20445911.2019.1621872

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2012). Spatial anxiety relates to spatial abilities as a function of working memory in children. Q J Exp Psychol (Hove), 65(3), 474-487. https://doi.org/10.1080/17470218.2011.616214

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. 20(4), 873-922. https://doi.org/10.1162/neco.2008.12-06-420

Scott, K., & Schulz, L. (2017). Lookit (Part 1): A New Online Platform for Developmental Research. Open Mind, 1(1), 4-14. https://doi.org/10.1162/opmi_a_00002

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. Science, 171(3972), 701-703. https://doi.org/10.2307/1731476

Toth, A. J., & Campbell, M. J. (2019). Investigating sex differences, cognitive effort, strategy, and performance on a computerised version of the mental rotations test via eye tracking. Scientific Reports, 9(1), 19430. https://doi.org/10.1038/s41598-019-56041-6

Voyer, D., & Hou, J. (2006). Type of items and the magnitude of gender differences on the Mental Rotations Test. Can J Exp Psychol, 60(2), 91-100. https://doi.org/10.1037/cjep2006010

Voyer, D., Rodgers, M. A., & McCormick, P. A. (2004). Timing conditions and the magnitude of gender differences on the Mental Rotations Test. Mem Cognit, 32(1), 72-82. https://doi.org/10.3758/bf03195821

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

Weinstein, Y., & Roediger, H. L., 3rd (2012). The effect of question order on evaluations of test performance: how does the bias evolve?. *Memory & cognition*, *40*(5), 727–735. https://doi.org/10.3758/s13421-012-0187-3

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. Frontiers in Neuroinformatics, 7, 14. https://doi.org/10.3389/fninf.2013.00014