

Full Length Article

Time pressure amplifies gender differences in mental rotation through shifts in evidence accumulation and response caution

Yaxin Liu^{*}, Stella F. Lourenco^{*}

Emory University, United States of America

ARTICLE INFO

Keywords:

Spatial cognition

Mental rotation

Gender differences

Drift diffusion modeling

ABSTRACT

Meta-analytic data provide evidence for a gender difference favoring men on mental rotation tasks, especially under time pressure. Why might time pressure exaggerate this gender difference? Here we used drift diffusion modeling (DDM) to examine potential mechanisms underlying gender differences in mental rotation. Adult participants ($N = 160$; 80 men, 80 women) performed a mental rotation task with two levels of difficulty under conditions that emphasized either accuracy or speed. DDM parameters—including drift rates, decision thresholds, and non-decision times—were analyzed. We found robust gender differences in decision thresholds when speed was emphasized, with women showing smaller thresholds (less evidence accumulation) than men. Drift rates, an index of processing efficiency, showed weaker and less consistent gender differences; women's drift rates decreased under time pressure but men's remained unchanged. No reliable gender differences emerged in non-decision times. Implications of time pressure on gender differences in mental rotation performance are discussed in relation to affect and strategy use.

1. Introduction

That men and women may perform differentially on tasks in the visuospatial domain has been the focus of much discussion. Most studies point to some differences in performance and/or strategy use across several spatial domains, including spatial navigation (Coluccia & Louse, 2004; Nazareth et al., 2019), object recognition and memory (Voyer et al., 2007), and visuospatial working memory (Voyer et al., 2017). However, the magnitude of the gender differences in small-scale spatial tasks varies considerably: whereas negligible differences or small effects emerge in tasks involving spatial perception and visuospatial working memory (Voyer et al., 2017), medium-to-large differences favoring men are observed in mental rotation tasks (Linn & Petersen, 1985; Voyer et al., 1995), especially under time constraints (Voyer, 2011).

Gender differences in mental rotation—which involves the mental manipulation of images (e.g., visualizing and rotating objects in one's mind) (Shepard & Metzler, 1971)—have been studied extensively, in part because mental rotation predicts consequential outcomes, including math learning in childhood (Atit et al., 2022; Cheung et al., 2020; Verdine et al., 2017), success in STEM fields (Uttal et al., 2013; Wai et al., 2009), and innovation outcomes such as number of patents (Kell et al., 2013). These gender differences emerge as early as 6 years of

age and increase over development (Lauer et al., 2019; Linn & Petersen, 1985; Voyer et al., 1995), underscoring the importance of identifying the mechanisms that drive them. Critically, however, the magnitude of the gender differences is not fixed. These differences reliably interact with stimulus complexity and difficulty (Peters & Battista, 2008), and they also vary as a function of format. In the psychometric format (i.e., paper-and-pencil measures with a global time limit; Vandenberg & Kuse, 1978), the advantage favoring men tends to be relatively large. By contrast, with the chronometric format, in which individual trial-level data are recorded for stimulus pairs, gender differences in accuracy and RT (response times) are minimal or even absent (Jansen-Osmann & Heil, 2007; Voyer et al., 1995). Recent evidence suggests that such format effects stem in part from specific features of trial design and administration (Jost & Jansen, 2024). One reason for the format difference is that, on psychometric tasks, women may attempt fewer items than men under standard limits, and gender differences are smaller when scoring accounts for the number of items attempted (Goldstein et al., 1990; Peters, 2005), suggesting that time limits may yield different response patterns between men and women. Consistent with this account, meta-analytic reviews show that gender differences are most prominent under timed conditions, increasing linearly as the time limit decreases (Voyer, 2011). Under speeded instructions, men

^{*} Corresponding authors at: Emory University, 36 Eagle Row, GA 30322, United States of America.

E-mail addresses: yaxin.liu@emory.edu (Y. Liu), stella.lourenco@emory.edu (S.F. Lourenco).

<https://doi.org/10.1016/j.cognition.2026.106723>

Received 15 February 2026; Received in revised form 3 August 2026; Accepted 10 September 2026

Available online 23 September 2026

0010-0277/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

typically exhibit both higher accuracy and faster responses than women (Linn & Petersen, 1985; Voyer et al., 1995). When tasks are untimed, or loosely timed, women's accuracy more closely approximates men's (Halpern & LaMay, 2000; Heil & Jansen-Osmann, 2008). These findings suggest that time pressure plays a critical role in amplifying gender differences, but the mechanisms through which it does so remain unclear.

Time limits impose speed-accuracy tradeoffs on cognitive tasks broadly, and optimal performance often depends on balancing these competing demands (Bogacz et al., 2010; Pieters & Warlop, 1999). Women appear disproportionately affected by time constraints in mental rotation, exhibiting larger performance declines under speeded conditions (Voyer, 2011). Traditionally, RT differences have been interpreted as differences in processing efficiency, such that longer RTs are assumed to reflect slower or less efficient information processing (Robert & Chevrier, 2003). However, RTs can also reflect response strategies, such as more cautious or conservative responding (Hedge et al., 2018). Indeed, classic psychophysics showed that apparent gender differences in rotation speed largely reflect differential speed-accuracy trade-offs—women's accuracy reaching asymptote sooner than men's on trials requiring more rotation—rather than differences in rotation rate per se (Lohman, 1986). Relatedly, women are less willing to guess on mental rotation tasks (Goldstein et al., 1990; Jost & Jansen, 2024; Kerkman et al., 2000; Ramírez-Uclés & Ramírez-Uclés, 2020) and show greater caution in various non-spatial decision-making domains, including intellectual and physical risk-taking (Byrnes et al., 1999; Croson & Gneezy, 2009). Evidence accumulation models interpret such caution as elevated decision thresholds (Ratcliff & McKoon, 2008). Supporting this view, adolescent girls show larger decision thresholds than boys on a stop-signal task (Epstein et al., 2023; Wiker et al., 2024).

These observations raise the possibility that response caution contributes to gender disparities under time pressure on mental rotation tasks. If women are generally more cautious, then time limits may disproportionately reduce their performance, induce exaggerated threshold changes, or both. Importantly, greater caution could manifest either as uniformly higher thresholds than men's or as a larger threshold reduction when speed is enforced—a direction that aggregate accuracy and RT cannot dissociate. Yet no study to date has formally modeled gender differences in decision thresholds during mental rotation, despite the successful application of evidence accumulation models to other spatial and cognitive tasks (Feldman & Huang-Pollock, 2021; Larsen, 2014; Provost & Heathcote, 2015).

A key advantage of using evidence accumulation models is that they dissociate processing efficiency (drift rate) from response caution (decision thresholds). These models have been widely used to understand how task demands modulate speed-accuracy tradeoffs (Ratcliff & McKoon, 2008). For example, in motion coherence tasks, emphasizing speed leads to smaller decision thresholds (Herz et al., 2018; Zhang & Rowe, 2014), consistent with reduced evidence accumulation under time pressure. If decision thresholds in mental rotation similarly decrease under speeded conditions—and if this interacts with gender—then diffusion modeling can reveal whether gender differences reflect strategy-level differences in decision-making rather than underlying differences in the efficiency of spatial processing per se.

Drift diffusion modeling (DDM) has been applied to a range of perceptual and cognitive tasks (Forstmann et al., 2016; Wagenmakers et al., 2008; White et al., 2011), including studies examining group differences in development (Feldman & Huang-Pollock, 2021) and age-related changes in drift rates (Theisen et al., 2021). Importantly, changes in drift rates were associated with affective variables (i.e., anxiety) during emotion regulation, with higher anxiety linked to lower drift rates (Warren et al., 2020). As women often report lower confidence and greater anxiety on mental rotation tasks than men (Estes & Felker, 2012; Lyons et al., 2018), their drift rates may be disproportionately affected by anxiety—an effect likely to be exacerbated by time pressure. Likewise, women are more likely to use analytical, piecemeal strategies

rather than holistic rotational strategies (Boone & Hegarty, 2017; Geiser et al., 2006), potentially lowering drift rates under speeded conditions. Diffusion modeling provides a principled way to isolate whether time pressure amplifies gender differences in response caution, processing efficiency, or both.

1.1. Present study

In the present study, we used DDM to test whether emphasizing accuracy versus speed differentially affected men's and women's decision thresholds and drift rates on a chronometric mental rotation task. The task included both rotation (100°) and non-rotation (0°) trials. In DDM, steeper drift rates index faster or more efficient evidence accumulation toward correct answers, whereas larger boundary separation (decision threshold) corresponds to greater caution before committing to a decision (see Methods for parameters). We modeled parameters as a function of difficulty (0° vs. 100°), task condition (accuracy vs. speed), trial type (same vs. mirror), and gender.

Based on prior work, we predicted that decision thresholds would decrease under time pressure, reflecting reduced caution when speed is prioritized. It remains unclear, however, whether gender differences in thresholds would accompany this manipulation. If women are more affected by time pressure than men (Caviola et al., 2017; Voyer, 2011), then gender differences in thresholds should be amplified in the speed-emphasis condition.

We also examined drift rates to test whether time pressure affected processing efficiency. Drift rates might remain stable if time limits do not alter the quality of sensory evidence (Wenzlaff et al., 2011). Alternatively, drift rates might increase when accuracy is emphasized (Ho et al., 2012). Given prior evidence that affective and strategic factors vary by gender, we considered the possibility that gender differences in drift rates could also be amplified under speed emphasis.

2. Method

2.1. Participants

Our sample size was guided by an a priori power analysis based on effect sizes from a previously published study examining gender-by-angle interactions (gender effect: $\eta^2 = 0.06$; Experiment 2; Boone & Hegarty, 2017), and from a pilot study in our lab (gender effect: $\eta^2 = 0.03$). Using a $2 \times 2 \times 2$ mixed design (Accuracy-emphasis vs. Speed-emphasis [between-subjects]; 0° vs. 100° [within-subjects]; Men vs. Women [between-subjects]), we aimed for a minimum of 150 participants to achieve 0.80 power for main effects and interactions (Anderson et al., 2017).

We recruited 180 adults from Prolific, balanced across conditions, to allow for attrition and exclusions. Five participants in the accuracy-emphasis condition and two in the speed-emphasis condition did not complete the study. An additional five (accuracy-emphasis condition) and eight (speed-emphasis condition) were excluded for scoring below chance (0.50) on the lowest difficulty trials (0° same). The final sample comprised 160 participants: 80 (50% men) in the accuracy-emphasis condition and 80 (50% men) in the speed-emphasis condition. Participant age (accuracy-emphasis: $M = 27.4$ years; $SD = 4.9$; speed-emphasis: $M = 27.4$ years; $SD = 4.4$; range = 18–40 years) was similarly distributed across conditions.

During pre-screening, all participants confirmed that their birth sex matched their gender identity, that they had normal or corrected-to-normal vision, and that they were right-handed. All procedures were approved by the local Institutional Review Board (IRB).

2.2. Stimuli and procedure

Stimuli were objects comprised of individual blocks, adapted from a public stimulus library (Ganis & Kievit, 2015; see Fig. 1). Stimulus sizes

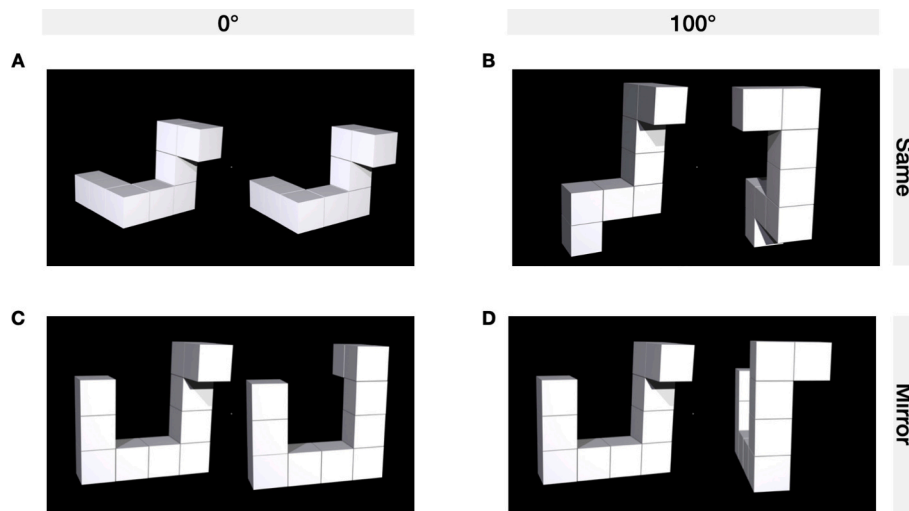


Fig. 1. Schematic representation of the mental rotation task. Examples of 0° “same” (A), 100° “same” (B), 0° “mirror” (C), and 100° “mirror” (D) trials are shown. Stimulus images were obtained from [Ganis and Kievit \(2015\)](#) and adapted for online testing.

were automatically scaled to the width and height of each participant's computer screen. Participants were instructed to keep their eyes level with the center of the screen. The task was programmed using PsychoPy3 software and hosted online on Pavlovia ([Peirce, 2008](#)). On each trial, two objects appeared side by side ([Fig. 1](#)), and participants responded whether one object could be rotated to match the other. If so, they pressed the F key (“same”); if not, they pressed the J key (“mirror”). Non-matching objects were mirror images of one another.

In the *accuracy-emphasis* condition, participants were instructed to “respond as accurately as possible” and were given no explicit time limit. In the *speed-emphasis* condition, participants were instructed to “respond as quickly as possible.” In addition, if they failed to respond within 7500 ms, the objects disappeared and a prompt to respond was presented; the next trial began automatically after a key press. This time restriction followed the procedures associated with the [Ganis and Kievit \(2015\)](#) stimulus set.

Participants first completed eight practice trials with corrective feedback. Following practice, they were reminded of the condition-specific instructions and, for the speed condition, time limits. The test phase consisted of 120 trials without feedback. Object pairs differed by either 0° or 100° (randomized order), with an equal number of same and mirror trials per angle. We included 0° trials as a baseline condition requiring primarily perceptual judgments with minimal rotation. We chose 100° among 50°, 100°, and 150° trials used by others because the 100° disparity level best represents moderate rotational difficulty. This design allowed us to compare performance on rotation versus non-rotation trials and to link our findings to other studies using only perceptual (non-rotational) demands.

2.3. Drift diffusion modeling (DDM)

We estimated model parameters using hierarchical drift diffusion modeling (HDDM; [Wiecki et al., 2013](#))¹ implemented in Python 3.7. In this framework, individual-level parameters (e.g., drift rate, decision threshold) are drawn from group-level distributions. The hierarchical approach jointly estimates fixed and random effects, yielding more robust parameter recovery, particularly when trial numbers per participant are modest ([Wiecki et al., 2013](#)).

2.3.1. RT pre-processing

Because the accuracy-emphasis condition lacked a formal time limit, outlier trimming was required before fitting DDMs. Based on the observed RT distributions, we imposed an upper bound of 10 s in this condition to remove excessively long responses that are unlikely to reflect typical decision processes and are unsuitable for DDM. Trials in the accuracy-emphasis condition with RTs > 10 s were discarded (5.5% of total trials). In the speed-emphasis condition, trials with RTs exceeding the time limit of 7.5 s were excluded (1.2% of total trials). To address the right-skew of RTs, we log-transformed RTs within each participant and removed trials exceeding ± 2.5 SDs from each participant's mean on the transformed scale (accuracy condition: 0.4%; speed condition: 0.9%). Very fast RTs (< 0.3 s) were also excluded (accuracy condition: 0.1%; speed condition: 0.5%) to avoid anticipatory responses and keypress errors. The log-transformed RTs were only used for outlier detection; all DDMs were fit to the raw RTs for the remaining trials.

2.3.2. Model specification

Trial-wise RTs and choice accuracy were inputted into HDDM using the standard HDDM class. We used accuracy coding, with responses coded as correct (1) or incorrect (0) such that the upper boundary represented the correct response and the lower boundary represented the error response. Under this coding scheme, the magnitude of drift rates is directly associated with performance as drift rates reflect stronger evidence accumulating toward the correct boundary ([Wiecki et al., 2013](#)). We used this approach because our primary interest was in performance rather than stimulus-response mapping. It also yields cleaner, more parsimonious interpretations given that modeling of stimulus-response mappings was not our goal.

The main parameters of interest were drift rate (v), decision threshold (a), and non-decision time (t). To capture the effects of experimental conditions and manipulated variables, these parameters were allowed to vary by task condition (accuracy vs. speed), gender (men vs. women), angular disparity (0° vs. 100°) and trial type (same vs. mirror). Inter-trial variability parameters for drift rates (sv) and non-decision times (st) were included and held constant across conditions. The starting point (z) was fixed at 0.50, reflecting the assumption that participants had no a priori bias toward either boundary. We also compared a reduced model without inter-trial variability; results were highly similar, and we report the more flexible specification for completeness (see SI Table 4 for model comparisons).

Models were fit using Bayesian estimation via Markov chain Monte Carlo (MCMC) sampling in HDDM. We included a small outlier component (5%) to capture residual contaminants after RT trimming.

¹ For non-decision time, see online Supporting Information (SI).

Three independent chains of 10,000 samples were generated; the first 2000 samples of each chain were discarded as burn-in, and every second sample was retained (thinning = 2), yielding 4000 posterior samples per chain (12,000 total). Chains were initialized using the HDDM default procedure; no manual starting value was specified.

Standard weakly informative priors provided by the HDDM package were used for all parameters² (Wiecki et al., 2013), ensuring that data, rather than the priors, primarily drove the posterior estimates. Decision threshold and non-decision times were sampled from truncated normal distributions, and drift rates were sampled from a normal distribution (see SI Table 5 for details of each prior).

2.3.3. Model fit

Convergence was first evaluated via visual inspection of trace plots, confirming that the independent chains were well-mixed and stable, and then quantified via Gelman-Rubin R-hat statistics (Gelman & Rubin, 1992). All R-hat values across three chains were below 1.01, indicating successful convergence (Vats & Knudson, 2021). To assess model fit, we performed posterior predictive checks (PPCs), generating 1000 simulated datasets from the posterior and comparing the simulated RT and accuracy distributions to the observed data for each condition (see Results). PPC plots confirmed that the observed data fell within the 95% credible intervals of the posterior predictive distribution. Modest misfit appeared for the slowest RTs. Additional quantile probability plots displaying both accuracy and RT distributions for both correct and error responses (SI Fig. 1) provided converging evidence of adequate fit. The quality of fits of the models to data for each condition, trial type, and gender was plotted separately for 0.1, 0.3, 0.5, 0.7, and 0.9 RT quantiles (see SI Fig. 2). The parameter recovery analyses further confirmed reliable estimation given our trial counts (see SI Fig. 3).

2.4. Data and code availability

Data, analysis code, and stimulus materials are publicly available on OSF (<https://doi.org/10.17605/OSF.IO/ZY6P3>) or Mendeley Data (Liu & Lourenco, 2026).

3. Results

3.1. Behavioral analyses

Mixed ANOVAs examined the effects of task condition (accuracy- vs. speed-emphasis), gender (men vs. women), trial type (same vs. mirror), and angular disparity (0° vs. 100°) on mental rotation accuracy and RT. Given the final sample and design, a post-hoc sensitivity analysis (pwr R package; power = 0.80) indicated the design was sufficiently powered to detect small-to-medium effects ($f = 0.22$, $\eta_p^2 = 0.048$).

3.1.1. Accuracy

There was a significant three-way gender $\times$ trial type $\times$ angular disparity interaction, $F(1, 156) = 4.73$, $p = .031$, $\eta^2 = 0.029$. Given our primary interest in gender differences, we performed planned pairwise comparisons examining both between-gender (i.e., men vs. women in the accuracy condition) and within-gender (i.e., women in the accuracy vs. women in the speed condition) differences across task condition, despite the absence of a significant four-way interaction involving condition. Bonferroni-adjusted comparisons revealed that between-gender differences were largely non-significant across comparisons ($ps > 0.05$). The one (between-gender) comparison that approached significance occurred for 0° mirror trials in which men showed marginally higher accuracy scores than women in the speed-emphasis condition ($t[156] = 2.46$, $p = .061$, $d = 0.39$) (see Fig. 2A).

² Model parameters were comparable regardless of whether weakly informative priors or priors drawn from prior data were used.

Within-gender analyses revealed clear effects of task condition for women. Specifically, women were significantly less accurate in the speed- than accuracy-emphasis condition for mirror trials at both 0° ($t[156] = -2.63$, $p = .038$, $d = 0.42$) and 100° ($t[156] = -2.87$, $p = .019$, $d = 0.46$) (see Fig. 2A). A similar trend appeared on same trials at 100° ($t[156] = -2.51$, $p = .052$, $d = 0.40$). By contrast men's accuracy did not differ significantly between the accuracy- and speed-emphasis conditions for any combination of trial type and angular disparity ($ps > .05$). Together, these results suggest that women's accuracy may be more sensitive to time pressure and task complexity than men's.

As a manipulation check, participants in the speed-emphasis condition ($M = 0.79$) showed significantly lower accuracy scores than those in the accuracy-emphasis condition ($M = 0.87$), $F(1, 156) = 12.25$, $p < .001$, $\eta^2 = 0.073$, confirming that time pressure induced more errors. Additionally, there was a condition $\times$ angular disparity interaction, $F(1, 156) = 9.92$, $p = .002$, $\eta^2 = 0.060$. Post-hoc comparisons revealed that this accuracy reduction under speed-emphasis occurred exclusively at 100° trials, whereas no significant difference appeared on 0° trials (collapsed across gender), suggesting that time pressure selectively impaired performance on more difficult trials.

3.1.2. RT

RT analyses revealed a significant three-way condition $\times$ trial type $\times$ angular disparity interaction, $F(1, 156) = 18.74$, $p < .001$, $\eta^2 = 0.107$. Follow-up comparisons paralleling those for accuracy revealed that men and women responded significantly faster in the speed- versus accuracy-emphasis condition across all trial types and angular disparities ($ps < 0.01$; see Fig. 2B), except for men on 0° same trials ($p > .999$). Between-gender comparisons revealed no reliable RT differences in either condition ($ps > 0.05$; see Fig. 2B).

Together, the behavioral results suggest that speed emphasis reliably reduced RTs for both genders, while selectively reducing accuracy, particularly for women on more complex trials.

3.2. Drift diffusion modeling (DDM)

The behavioral results indicated that time pressure reduced accuracy while accelerating responses, especially for women. These effects could arise from changes in processing efficiency (drift rate), response caution (decision threshold), or both. We therefore modeled drift rate (v), decision threshold (a), and non-decision time (t) as a function of gender, condition, trial type, and angular disparity using Bayesian DDM with weakly informative priors. In this way, we obtained posterior parameter estimates after MCMC sampling using weak priors (see SI Tables 1 & 5 for details). Posterior predictive checks (PPCs) confirmed that the model captured accuracy (see Fig. 3A) and RT (see Fig. 3B) distributions well, as shown by overlapping distributions and 95% credible intervals.

Next, we compared within- and between-gender differences at each level of task condition, trial type, and angular disparity. Following the standard approach to hypothesis testing with hierarchical drift diffusion models (Wiecki et al., 2013), inference was based on the posterior distributions of the parameters. For each comparison, we report the posterior mean difference, its 95% highest density intervals (HDI), and the posterior probability of specified direction (Makowski et al., 2019). We interpret strength of evidence based on posterior probabilities: a probability of 95% or greater indicates strong evidence supporting a difference, a probability between 90% and 95% indicates moderate evidence, and a probability below 90% indicates weak or null evidence (Kruschke, 2013; McElreath, 2020). The 95% HDI served to note the precision of each effect rather than as a separate hypothesis test. Descriptive statistics (i.e., Means, SDs, differences, 95% HDIs, R-hat) of all model parameters are reported in SI Table 1.

3.2.1. Decision thresholds

Analyses of posterior parameter values revealed that decision thresholds were smaller in the speed-emphasis condition than the

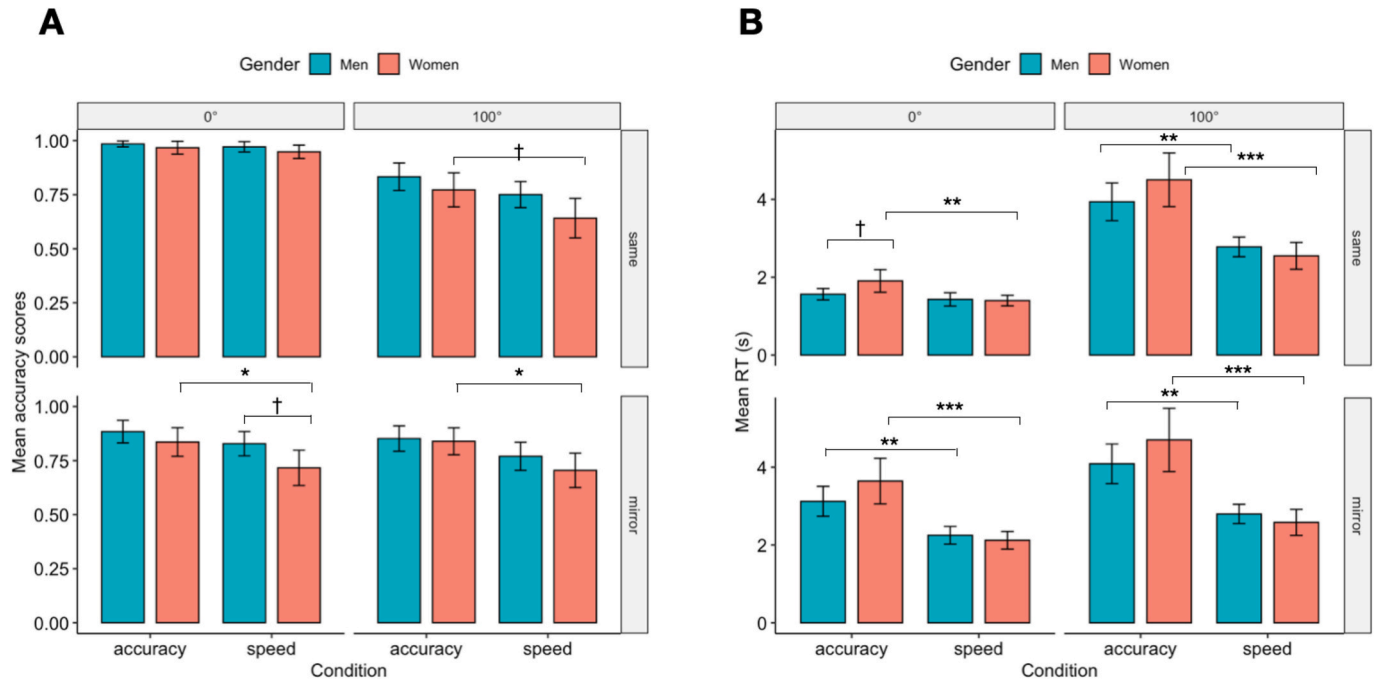


Fig. 2. Mean accuracies (proportion correct; Panel A) and RTs (Panel B) in accuracy-emphasis and speed-emphasis conditions by trial type and angular disparity. Error bars represent 95% CIs. * $p < .05$, ** $p < .01$, *** $p < .001$, † marginal significance. Non-significant (between-gender or within-gender) markers are not shown.

accuracy-emphasis condition for both genders (all $P[\text{speed} < \text{accuracy}] > 95\%$), as expected if caution decreases when speed is prioritized over accuracy. There was strong evidence for this reduction across trials and disparities in both genders (all $P[\text{speed} < \text{accuracy}] \geq 97\%$; see posterior differences and HDIs in SI Table 2) (see Fig. 4A). The 95% HDIs excluded zero in all within-gender comparisons except men's 0° same trials ($\Delta = -0.59 [-1.18, 0.04]$; $P[\text{speed}_{\text{men}} < \text{accuracy}_{\text{men}}] = 97\%$). Thus, across participants, time pressure reduced response caution. Under greater time pressure, participants required less evidence before responding, consistent with a shift toward less cautious decisions.

Do decision thresholds differ between men and women? Between-gender comparisons revealed that in the accuracy-emphasis condition, there was no evidence of a gender difference across same (0°: $\Delta = 0.09 [-0.61, 0.78]$; $P[\text{accuracy}_{\text{women}} < \text{accuracy}_{\text{men}}] = 41\%$; 100°: $\Delta = -0.07 [-0.82, 0.66]$; $P[\text{accuracy}_{\text{women}} < \text{accuracy}_{\text{men}}] = 58\%$) and mirror (0°: $\Delta = -0.02 [-0.67, 0.65]$; $P[\text{accuracy}_{\text{women}} < \text{accuracy}_{\text{men}}] = 53\%$; 100°: $\Delta = -0.08 [-0.80, 0.66]$; $P[\text{accuracy}_{\text{women}} < \text{accuracy}_{\text{men}}] = 58\%$) trial types. In the speed-emphasis condition, by contrast, women exhibited smaller decision thresholds than men on all but the 0° same trials. Evidence was strong at 100° for both trial types (same: $\Delta = -0.42 [-0.77, 0.03]$; $P[\text{speed}_{\text{women}} < \text{speed}_{\text{men}}] = 98\%$; mirror: $\Delta = -0.44 [-0.84, 0.02]$; $P = 98\%$) and at 0° mirror trials ($\Delta = -0.35 [-0.73, 0.06]$; $P = 96\%$), but weak at 0° same trials ($\Delta = -0.24 [-0.70, 0.18]$; $P[\text{speed}_{\text{women}} < \text{speed}_{\text{men}}] = 84\%$). This finding suggests that women were more affected by time pressure, accumulating less evidence before responding.

Altogether, decision thresholds decreased under time pressure for both men and women, consistent with previous studies showing that decision thresholds are smaller under time constraints (Herz et al., 2018; Zhang & Rowe, 2014). Yet gender differences in response caution emerged specifically under time pressure. Women showed smaller decision thresholds than men, though the difference was weakest on the easiest (0° same) trials.

3.2.2. Drift rates

Analyses of posterior parameters revealed that drift rates were consistently higher at 0° than 100° for same trials in both men and

women regardless of condition (all $P[v_{0^\circ} > v_{100^\circ}] > 95\%$; all 95% HDIs excluding zero; see SI Table 2), consistent with increased difficulty reducing evidence accumulation. By contrast, on mirror trials, drift rates did not reliably differ between 0° and 100° (all $P[v_{0^\circ} > v_{100^\circ}] < 95\%$; all 95% HDIs including zero; see SI Table 1). These findings suggest an interaction between trial type (same vs. mirror) and angular disparity in processing efficiency, which may indicate divergent strategies.

To examine the effect of task condition on drift rates, within-gender comparisons of posterior drift rates indicated that both men and women remained largely unaffected by speed-emphasis for easier (0° same) trials (men: $\Delta = 0.13 [-0.16, 0.40]$; $P[\text{speed}_{\text{men}} < \text{accuracy}_{\text{men}}] = 18\%$; women: $\Delta = 0.21 [-0.07, 0.51]$; $P[\text{speed}_{\text{women}} < \text{accuracy}_{\text{women}}] = 8\%$ [or 92% speed > accuracy]) (see Fig. 4B). However, for more difficult trials (100° same trials), women's drift rates were lower in the speed-emphasis condition than the accuracy-emphasis condition ($\Delta = -0.23 [-0.46, -0.03]$; $P[\text{speed}_{\text{women}} < \text{accuracy}_{\text{women}}] = 97\%$), whereas men's were not ($\Delta = -0.09 [-0.34, 0.15]$; $P = 76\%$). Similar, though weaker, patterns appeared for mirror trials, both at 100° (men: $\Delta = -0.08 [-0.31, 0.16]$, $P = 71\%$; women: $\Delta = -0.18 [-0.42, 0.06]$; $P = 93\%$; and 0° (men: $\Delta = -0.03 [-0.25, 0.22]$; $P = 56\%$; women: $\Delta = -0.15 [-0.39, 0.09]$; $P = 89\%$).

Between-gender comparisons further revealed that men's drift rates were higher than women's in the speed-emphasis condition at 100° same trials ($\Delta = -0.32 [-0.59, -0.06]$; $P[\text{speed}_{\text{women}} < \text{speed}_{\text{men}}] = 99\%$), whereas men and women did not differ at 0° same trials ($\Delta = -0.12 [-0.43, 0.20]$; $P[\text{speed}_{\text{women}} < \text{speed}_{\text{men}}] = 80\%$; see SI Table 3). On mirror trials, however, gender differences appeared at 0° ($\Delta = -0.30 [-0.56, -0.05]$; $P[\text{speed}_{\text{women}} < \text{speed}_{\text{men}}] = 99\%$) but not 100° ($\Delta = -0.17 [-0.42, 0.10]$; $P[\text{speed}_{\text{women}} < \text{speed}_{\text{men}}] = 90\%$) trials.

Interestingly, the accuracy-emphasis condition also showed moderate-to-strong evidence of gender differences across most trial types and angular disparities (0° same: $\Delta = -0.20 [-0.46, 0.06]$, $P[\text{accuracy}_{\text{women}} < \text{accuracy}_{\text{men}}] = 94\%$; 0° mirror: $\Delta = -0.18 [-0.40, 0.02]$, $P = 96\%$; 100° same: $\Delta = -0.18 [-0.37, 0.03]$, $P = 95\%$), with the exception of 100° mirror trials ($\Delta = -0.06 [-0.26, 0.15]$, $P[\text{accuracy}_{\text{women}} < \text{accuracy}_{\text{men}}] = 73\%$). Although none of these HDIs excluded zero—unlike the two differences under the speed-emphasis

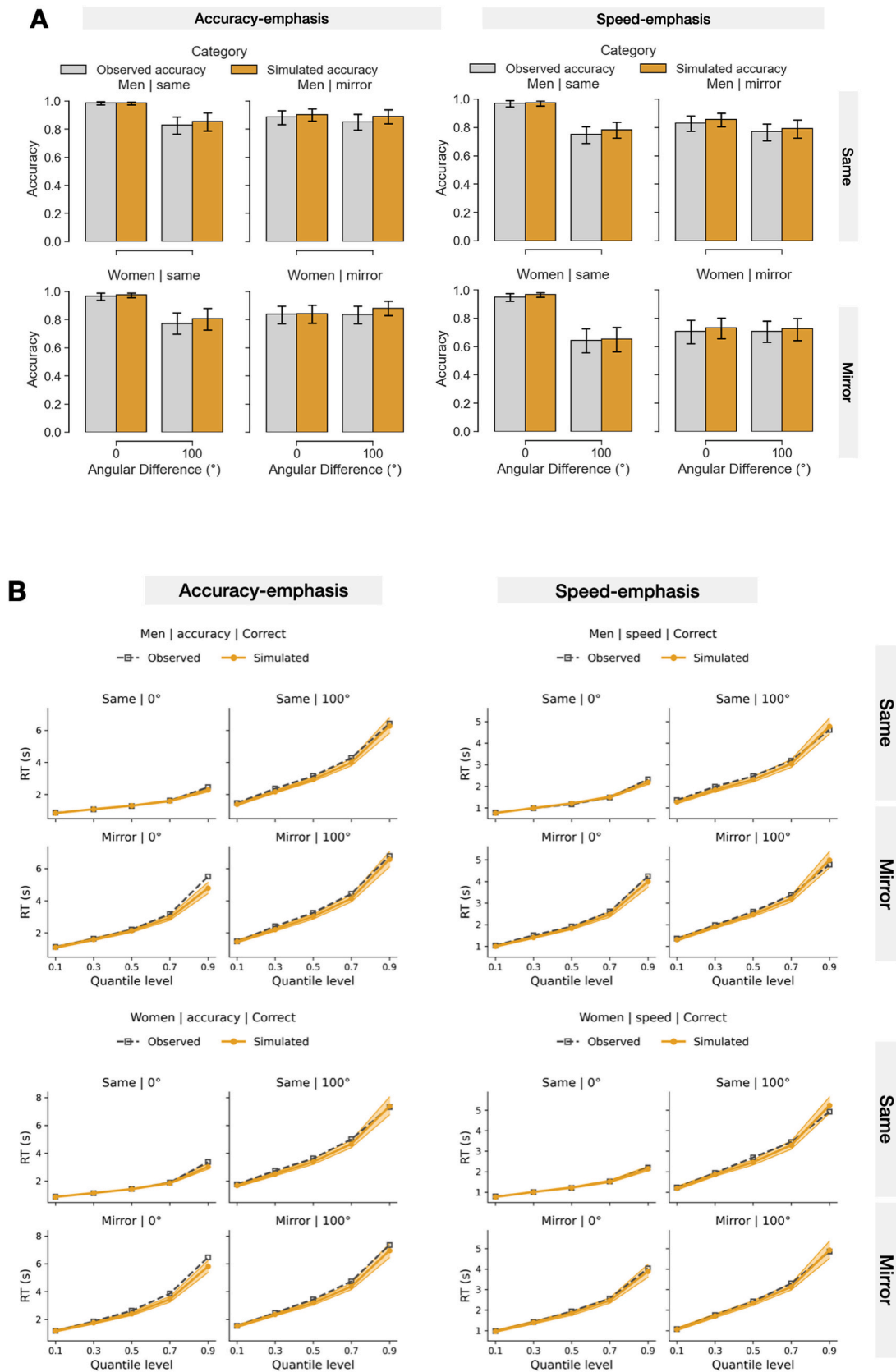


Fig. 3. Posterior predictive check (PPC) plots comparing empirical versus model-simulated accuracy rates (A), and empirical versus model-simulated RTs (B) were plotted against the 0.1, 0.3, 0.5, 0.7, 0.9 quantiles. The overlapping error bars suggest that the observed response patterns are well captured by the model. Shaded regions represent the 95% credible interval of the posterior predictive distribution. The model fits the data well across the 0° trials, though it slightly underestimates the 0.9 quantile RTs in the 100° trials. Panels show separate fits for same vs. mirror trials at the two levels of angular disparity.

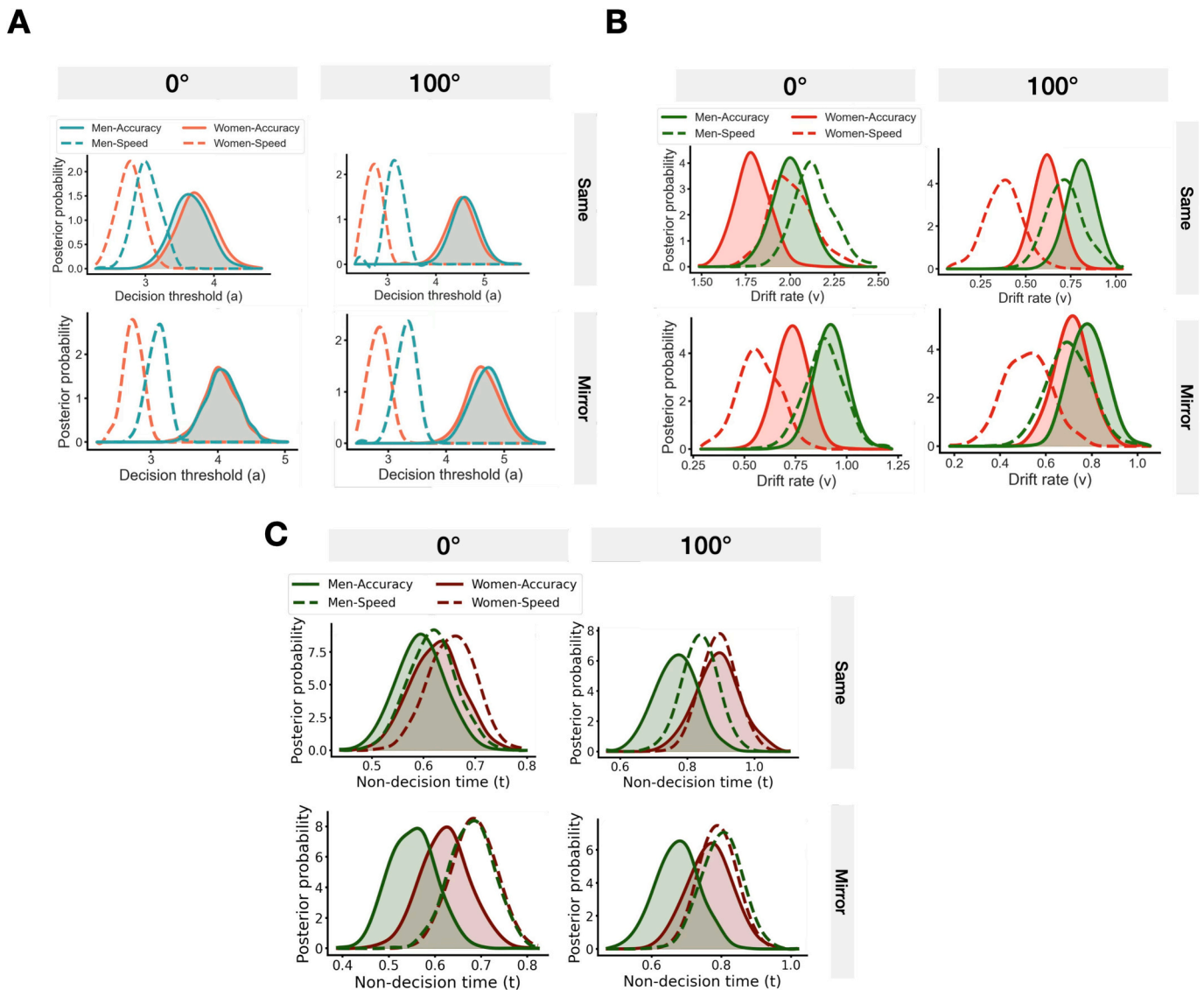


Fig. 4. (A) Decision thresholds, (B) drift rates, and (C) non-decision times are plotted separately for each condition by trial type and angular disparity.

condition—this consistent pattern warrants further discussion (see Discussion).

Taken together, these findings suggest that both speed- and accuracy-emphasis conditions showed gender differences in drift rates at various trial type and difficulty levels. In particular, women's drift rates were disproportionately impaired by time pressure with 100° trials compared to 0° trials. Men's drift rates were comparatively unaffected by time pressure regardless of trial type and angular disparity.

3.2.3. Non-decision times

We next examined the posterior parameter estimates for non-decision time (t), which captures processes outside of evidence accumulation (e.g., stimulus encoding and motor execution).

Overall, within-gender comparisons revealed that women showed no evidence of differences in non-decision times between speed- and accuracy-emphasis conditions across trial type and angular disparity ($P[\text{accuracy}_{\text{women}} < \text{speed}_{\text{women}}] < 90\%$; see SI Table 2 for full statistics) (see Fig. 4C). Men, however, showed longer non-decision times under speed emphasis on 0° mirror trials ($\Delta = 0.12 [-0.01, 0.24]$, $P[\text{accuracy}_{\text{men}} < \text{speed}_{\text{men}}] = 96\%$), with a similar but weaker trend on 100° mirror trials ($\Delta = 0.11 [-0.04, 0.28]$, $P = 92\%$). No evidence of differences was observed for same trials at either 0° ($\Delta = 0.02 [-0.10,$

$0.15]$, $P = 60\%$) or 100° ($\Delta = 0.00 [-0.15, 0.16]$, $P = 52\%$).

Between-gender comparisons revealed weak gender differences in both conditions (below 95% threshold), with one relatively stronger effect in which women showed longer non-decision times than men on 100° same trials in the accuracy-emphasis condition ($\Delta = 0.11 [-0.03, 0.24]$, $P = 94\%$; see SI Table 3 for full statistics). Thus, gender differences were not systematic in perceptual encoding and/or motor execution of responses.

4. Discussion

Despite longstanding interest in gender differences on mental rotation tasks, the mechanisms underlying these differences—particularly under time pressure—remain poorly understood. In the present study, we manipulated task demands to examine how time pressure affects performance and its underlying decision processes. Consistent with prior meta-analytic evidence that time pressure amplifies gender differences in mental rotation (Voyer, 2011; Voyer et al., 2004), we found that time pressure and speeded instructions reduced accuracy on difficult trials, particularly among women, and shortened response times for both genders. Importantly, DDM provided insight into why these effects emerged. Whereas both men and women showed the expected reduction

in decision thresholds under speed emphasis, women's thresholds were disproportionately reduced relative to men's. This finding suggests that, under time pressure, women adopted a less cautious response policy compared to men. Drift rates showed weaker and less consistent gender differences, though women tended to exhibit lower drift rates under speed emphasis on mental rotation trials. Men's drift rates, by contrast, were comparatively stable.

In what follows, we consider possible mechanisms underlying these gender-specific effects, focusing on response criteria, affective processes (including stereotype threat), and rotational strategies.

4.1. Response criteria

In line with extensive literature on speed-accuracy trade-offs, both men and women exhibited reduced decision thresholds under time pressure—a pattern widely observed across perceptual and cognitive tasks (Starns & Ratcliff, 2010; Zhang & Rowe, 2014). The reduction in thresholds likely reflects the adaptive shift toward less cautious responding when speed is prioritized over accuracy (Ivanoff et al., 2008; Palmer et al., 2005; Ratcliff & McKoon, 2008).

The novel theoretical contribution of the present work lies in the gender-specific pattern in this adaptation. Contrary to common assumptions that women are generally more cautious responders, women's decision thresholds were smaller than men's under time pressure (i.e., speed-emphasis condition). Thus, rather than adopting a stricter criterion, women accumulated less evidence before responding. This pattern contrasts with prior research in other domains, such as the stop-signal task, where women typically show greater caution (Wiker et al., 2024) and with risk-taking tasks where women adopt more conservative policies under stress (i.e., using cold pressor; Lighthall et al., 2012). Our results, however, challenge the notion of trait-level gender differences in caution and, instead, highlight the importance of task demands, affective context, and time pressure in modulating response criteria.

One question is whether these threshold effects might instead reflect a static response bias—a shift in willingness to respond “same” versus “mirror.” Additional signal detection analyses (see SI Fig. 4) argue against this possibility. Sensitivity (d') converged with the drift-rate findings, with women showing lower d' than men, most clearly under speed emphasis; criterion (c), by contrast, showed no gender by condition interaction, unlike the decision thresholds. This dissociation indicates that the gender difference under time pressure is not a static bias toward one response, but a property of how evidence is accumulated over time—the dynamic component that diffusion modeling isolates and that signal detection theory, by construction, cannot capture.

Next, we consider psychological mechanisms that might drive this context-dependent reduction in caution for women.

4.2. Affect and stereotype threat

Decision thresholds are known to shift with real-time monitoring of confidence and uncertainty. Individuals continue accumulating information until subjective confidence reaches a criterion level (Hausmann & Läge, 2008), and manipulations that disrupt confidence or heighten uncertainty can alter thresholds. For example, explicit confidence ratings can increase subsequent caution (Li et al., 2024), whereas low confidence can lead to lower thresholds and premature termination of evidence accumulation (Y. Liu & Lourenco, 2022).

Women often report lower confidence and greater anxiety than men on mental rotation tasks (Arrighi & Hausmann, 2022; Estes & Felker, 2012; Liu & Lourenco, 2022). Time pressure may amplify these affective responses, prompting earlier termination of evidence accumulation. Although we did not directly measure affective states in the present study, prior research shows that monitoring one's own performance and affect can influence decision thresholds (Li et al., 2024). Thus, women's reduced thresholds under speed emphasis may reflect affectively driven disengagement from extended evidence accumulation. Future research

should consider testing whether moment-to-moment confidence or anxiety mediates these threshold differences.

Confidence is also relevant to the longstanding finding that women are less likely to guess on mental rotation tasks (Voyer & Saunders, 2004). Although our findings do not directly index guessing behavior, reduced willingness to seek further evidence under uncertainty may contribute to this pattern. In extreme cases, premature decision termination may function as an “escape strategy” to alleviate task-related anxiety (Choe et al., 2019; Daker et al., 2023; Lourenco & Liu, 2023).

Post-error behavior provides converging evidence for this account (see SI). Meta-analytic evidence indicates that the coupling between anxiety and neural error monitoring is stronger in women than men (Moser et al., 2016), which suggests that women could show greater behavioral adjustment following errors. We found the opposite, however, such that men slowed after errors, adopting an adaptive, caution-increasing adjustment, whereas women showed essentially no post-error slowing (SI Fig. 5). This pattern suggests that heightened error reactivity in women does not necessarily translate into increased caution; instead, the absence of post-error adjustment mirrors women's reduced decision thresholds under time pressure, consistent with error signals driving affect-related disengagement rather than compensatory monitoring.

Affective explanations also extend to drift rates. According to Attentional Control Theory (ACT), anxiety disrupts efficient information processing (Eysenck et al., 2007) and consistent with this view, higher anxiety predicts lower drift rates (L. Liu et al., 2022; Y. Liu & Lourenco, 2022). Women's lower drift rates under time pressure may therefore reflect anxiety-related disruptions in processing efficiency. Notably, gender differences in drift rates in the present study appeared even on non-rotation trials (0° mirror trials) in both timed and untimed conditions, suggesting that women may experience greater task-related anxiety irrespective of explicit time constraints. This interpretation aligns with claims that ostensibly “untimed” tasks may still be experienced as implicitly time pressured, and that men and women may interpret timing instructions differently (Goldstein et al., 1990). More broadly, this is consistent with the argument that features of trial design and administration modulate the magnitude of gender differences in mental rotation (Jost & Jansen, 2024); our findings identify time pressure as one such feature, operating at the level of decision processes rather than overall performance alone.

Stereotype threat provides a broader framework for understanding affective dynamics. Simply reminding women of gender stereotypes regarding spatial ability can heighten anxiety and impair performance (Moè, 2009). Although few studies have examined stereotype threat through DDM, its effect could plausibly manifest in lower drift rates, reduced response caution, or both. Future work might directly test how stereotype messaging modulates diffusion parameters during spatial decision-making.

4.3. Rotational strategies

A final explanation for the observed pattern involves strategy use. Prior work suggests that women are more likely to adopt an analytical (piecemeal) strategy, whereas men favor a holistic strategy (Boone & Hegarty, 2017; Geiser et al., 2006; Hirnstein et al., 2009; Voyer et al., 2020). The analytical strategy, which involves comparing components piece by piece, is thought to be more time-consuming than the holistic strategy, placing greater demands on sequential information extraction. Under time pressure, an analytical strategy may therefore reduce the effective rate of evidence accumulation.

Women's reduced drift rates—especially on trials requiring substantial rotation (100°) or flip transformations (0° mirror trials)—are consistent with this possibility. By contrast, non-rotation 0° same trials—which require minimal rotation and on which strategies should converge—yielded weak and inconsistent gender differences. Our parameter recovery analysis (presented in SI) confirmed that model-estimated parameters were reliable even given the low error rates on

the 0° same trials. Thus, the attenuated difference is unlikely to reflect estimation failure. The absence of a gender difference may therefore be best understood as an expected convergence of strategies on the least demanding trials.

This difficulty-dependent pattern aligns with Lohman's (1986) speed–accuracy analysis, in which women's accuracy reached asymptote sooner than men's on the most rotated trials, whereas the rate of accuracy gain did not differ by gender. Lohman interpreted the apparent gender difference in rotation speed as a downstream consequence of these accuracy-ceiling differences on the hardest trials. Our findings of drift rates provide a trial-level analogue such that reduced processing efficiency for women was concentrated on more demanding trials, suggesting the difference reflects the quality of rotation on hard problems rather than a global speed deficit.

An additional nuance is that gender differences were not uniform across same versus mirror trials. Mirror trials can be more effortful than same trials because they require an additional “flip” transformation (Hamm et al., 2004) or because of the mismatch between perceptual representations of mirrored objects and top-down expectations of their configuration (Paschke et al., 2012). Such effortful conditions may interact with strategy differences or affective responses to produce heterogeneous patterns across trial types.

Moreover, the neural processes underlying these strategies appear heterogeneous across men and women rather than uniform: women's mental rotation has been associated with stronger somatosensory, body-based engagement than men's, which can itself be less efficient (Bersier et al., 2024). Although we did not directly measure strategy use in the present study, it is likely that the drift-rate differences reflect a mixture of processes rather than a single strategic switch. Finally, the strategy account should be regarded as one probable contributor rather than a definitive mechanism. Future studies should consider linking DDM parameters to independently measured strategy profiles.

4.4. Methodological considerations and potential limitations

One potential limitation of the current modeling approach is that we did not directly assess the starting point parameter; we assumed no a priori response bias. Although this assumption is common in mental rotation paradigms—where stimulus processing is required before a correct response can be reached—it remains possible that subtle response biases exist (see Ratcliff & McKoon, 2008, for concerns on multi-stage processing). To address this possibility, we additionally performed a response-coded DDM on our data (upper boundary: same response; lower boundary: different response) to assess any starting point biases pressing the two response keys (see SI). Importantly, neither men nor women exhibited starting point response biases, suggesting that response preferences did not drive the observed gender differences. According to Brosnan et al. (2010), gender differences in accuracy occur mostly on “same” trials with larger rotations, even when the amount of same/mirror trials was disproportionately allocated in the task, suggesting that a preferential bias toward one response alone should not explain gender differences. If response biases nevertheless exist, understanding how the abovementioned factors influence decision processes remains an important direction for further research.

A further limitation concerns online testing. Although stimulus presentation was scaled to each participant's display and posterior predictive checks indicated adequate model fit at the aggregate level, uncontrolled variability in screen characteristics, viewing conditions, and potential distractions may have added noise to the RT measurement, limiting the precision with which individual component processes—particularly non-decision time, which absorbs display and motor latencies—can be isolated. Future research should complement online testing with controlled laboratory administration to reduce measurement noise.

5. Conclusion

This study provides new insight into the mechanisms underlying the gender differences in mental rotation. By decomposing performance into components of response caution and processing efficiency, DDM revealed gender-specific differences in decision thresholds and drift rates that are not easily captured by traditional accuracy and RT measures. These findings underscore the importance of considering decision processes—not just spatial performance—when interpreting gender disparities in spatial cognition. As in other domains where DDM has clarified cognitive processes underlying group differences (Forstmann et al., 2016), our results demonstrate its utility in understanding how task demands, affective context, and strategy may interact to shape mental rotation performance. Understanding how evidence accumulation and decision-making vary across individuals has important implications for reducing gender disparities in spatial cognition and related STEM pathways. Future work integrating DDM with direct measurements of confidence, anxiety, stereotype threat, and strategy use will be critical for developing a more complete mechanistic account of performance differences.

Author note

The authors declare no conflicts of interest.

The data and analysis from the experiments are available at https://osf.io/zy6p3/?view_only=324f3d29be9a47efa8076abe91ad57db.

CRediT authorship contribution statement

Yaxin Liu: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stella F. Lourenco:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Ethics declaration

Written informed consent to take part in the study and to publish the article has been obtained from all participants or their legal representatives. The privacy rights of participants have been observed.

This study was performed in compliance with relevant laws, regulatory frameworks and guidelines where the research took place.

This study was approved by the Emory University.

(Approval No. CR003-IRB00003388)

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2026.106723>.

Data availability

Data are available at OSF (<https://doi.org/10.17605/OSF.IO/ZY6P3>) and Mendeley Data (doi: <https://data.mendeley.com/datasets/pt7pz7mxm3/1>).

References

- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28(11), 1547–1562. <https://doi.org/10.1177/0956797617723724>
- Arrighi, L., & Hausmann, M. (2022). Spatial anxiety and self-confidence mediate sex/gender differences in mental rotation. *Learning & Memory*, 29(9), 312–320. <https://doi.org/10.1101/lm.053596.122>
- Atit, K., Power, J. R., Pigott, T., Lee, J., Geer, E. A., Uttal, D. H., ... Sorby, S. A. (2022). Examining the relations between spatial skills and mathematical performance: A

- meta-analysis. *Psychonomic Bulletin & Review*, 29(3), 699–720. <https://doi.org/10.3758/s13423-021-02012-w>
- Bersier, N. M., Arbuth, S., Ionta, S., & Rumiati, R. I. (2024). Mental rotation-related neural interactions between gender and cognitive strategy. *Imaging Neuroscience*, 2. https://doi.org/10.1162/imag_a_00310. imagc-2-00310.
- Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, 33(1), 10–16. <https://doi.org/10.1016/j.tins.2009.09.002>
- Boone, A. P., & Hegarty, M. (2017). Sex differences in mental rotation tasks: Not just in the mental rotation process! *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(7), 1005–1019. <https://doi.org/10.1037/xlm0000370>
- Brosnan, M., Walker, I., & Collomosse, J. (2010). The effect of explicitly varying the proportion of “same” and “different” responses on sex differences in the Shepard and Metzler mental rotation task. *European Journal of Cognitive Psychology*, 22(2), 172–189. <https://doi.org/10.1080/09541440902743710>
- Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, 125(3), 367–383. <https://doi.org/10.1037/0033-2909.125.3.367>
- Caviola, S., Carey, E., Mammarella, I. C., & Szucs, D. (2017). Stress, time pressure, strategy selection and math anxiety in mathematics: A review of the literature. *Frontiers in Psychology*, 8, 1488. <https://doi.org/10.3389/fpsyg.2017.01488>
- Cheung, C.-N., Sung, J. Y., & Lourenco, S. F. (2020). Does training mental rotation transfer to gains in mathematical competence? Assessment of an at-home visuospatial intervention. *Psychological Research*, 84(7), 2000–2017. <https://doi.org/10.1007/s00426-019-01202-5>
- 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. *Science Advances*, 5(11), Article eaay1062. <https://doi.org/10.1126/sciadv.aay1062>
- Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology*, 24(3), 329–340. <https://doi.org/10.1016/j.jenvp.2004.08.006>
- Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448–474. <https://doi.org/10.1257/jel.47.2.448>
- Daker, R. J., Slipenky, M. S., Green, A. E., & Lyons, I. M. (2023). Evidence for avoidance tendencies linked to anxiety about specific types of thinking. *Scientific Reports*, 13(1), 3294. <https://doi.org/10.1038/s41598-023-29834-z>
- Epstein, J. N., Karalunas, S. L., Tamm, L., Dudley, J. A., Lynch, J. D., Altaye, M., ... Atluri, G. (2023). Examining reaction time variability on the stop-signal task in the ABCD study. *Journal of the International Neuropsychological Society*, 29(5), 492–502. <https://doi.org/10.1017/S1355617722000431>
- Estes, Z., & Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. *Archives of Sexual Behavior*, 41(3), 557–570. <https://doi.org/10.1007/s10508-011-9875-5>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Feldman, J. S., & Huang-Pollock, C. (2021). A new spin on spatial cognition in ADHD: A diffusion model decomposition of mental rotation. *Journal of the International Neuropsychological Society*, 27(5), 472–483. <https://doi.org/10.1017/S13556177220001198>
- 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>
- Ganis, G., & Kievit, R. (2015). A new set of three-dimensional shapes for investigating mental rotation processes: Validation data and stimulus set. *Journal of Open Psychology Data*, 3. <https://doi.org/10.5334/jopd.ai>
- Geiser, C., Lehmann, W., & Eid, M. (2006). Separating “rotators” from “nonrotators” in the mental rotations test: A multigroup latent class analysis. *Multivariate Behavioral Research*, 41(3), 261–293. <https://doi.org/10.1207/s15327906mbr4103.2>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4). <https://doi.org/10.1214/ss/1177011136>
- Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition*, 18(5), 546–550. <https://doi.org/10.3758/BF03198487>
- Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, 12(2), 229–246. <https://doi.org/10.1023/A:1009027516424>
- Hamm, J., Johnson, B., & Coballis, M. (2004). One good turn deserves another: An event-related brain potential study of rotated mirror-normal letter discriminations. *Neuropsychologia*, 42(6), 810–820. <https://doi.org/10.1016/j.neuropsychologia.2003.11.009>
- Hausmann, D., & Läge, D. (2008). Sequential evidence accumulation in decision making: The individual desired level of confidence can explain the extent of information acquisition. *Judgment and Decision making*, 3(3), 229–243. <https://doi.org/10.1017/S1930297500002436>
- Hedge, C., Powell, G., Bompas, A., Vivian-Griffiths, S., & Sumner, P. (2018). Low and variable correlation between reaction time costs and accuracy costs explained by accumulation models: Meta-analysis and simulations. *Psychological Bulletin*, 144(11), 1200–1227. <https://doi.org/10.1037/bul0000164>
- Heil, M., & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly Journal of Experimental Psychology*, 61(5), 683–689. <https://doi.org/10.1080/17470210701822967>
- Herz, D. M., Little, S., Pedrosa, D. J., Tinkhauser, G., Cheeran, B., Foltyn, T., ... Brown, P. (2018). Mechanisms underlying decision-making as revealed by deep-brain stimulation in patients with Parkinson’s disease. *Current Biology*, 28(8), 1169–1178. e6 <https://doi.org/10.1016/j.cub.2018.02.057>
- Hirnshten, M., Bayer, U., & Hausmann, M. (2009). Sex-specific response strategies in mental rotation. *Learning and Individual Differences*, 19(2), 225–228. <https://doi.org/10.1016/j.lindif.2008.11.006>
- Ho, T., Brown, S., van Maanen, L., Forstmann, B. U., Wagenmakers, E.-J., & Serences, J. T. (2012). The optimality of sensory processing during the speed-accuracy tradeoff. *Journal of Neuroscience*, 32(23), 7992–8003. <https://doi.org/10.1523/JNEUROSCI.0340-12.2012>
- Ivanoff, J., Branning, P., & Marois, R. (2008). fMRI evidence for a dual process account of the speed-accuracy tradeoff in decision-making. *PLoS One*, 3(7), Article e2635. <https://doi.org/10.1371/journal.pone.0002635>
- Jansen-Osmann, P., & Heil, M. (2007). Suitable stimuli to obtain (no) gender differences in the speed of cognitive processes involved in mental rotation. *Brain and Cognition*, 64(3), 217–227. <https://doi.org/10.1016/j.bandc.2007.03.002>
- Jost, L., & Jansen, P. (2024). The influence of the design of mental rotation trials on performance and possible differences between sexes: A theoretical review and experimental investigation. *Quarterly Journal of Experimental Psychology*. <https://doi.org/10.1177/17470218231200127>
- Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability’s unique role. *Psychological Science*, 24(9), 1831–1836. <https://doi.org/10.1177/0956797613478615>
- Kerkman, D. D., Wise, J. C., & Harwood, E. A. (2000). Impossible “mental rotation” problems. *Learning and Individual Differences*, 12(3), 253–269. [https://doi.org/10.1016/S1041-6080\(01\)00039-5](https://doi.org/10.1016/S1041-6080(01)00039-5)
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2), 573–603. <https://doi.org/10.1037/a0029146>
- Larsen, A. (2014). Deconstructing mental rotation. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 1072–1091. <https://doi.org/10.1037/a0035648>
- Lauer, J. E., Yang, 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>
- Li, B., Hu, X., Shanks, D. R., Su, N., Zhao, W., Meng, L., ... Yang, C. (2024). Confidence ratings increase response thresholds in decision making. *Psychonomic Bulletin & Review*, 31(3), 1093–1102. <https://doi.org/10.3758/s13423-023-02380-5>
- Lighthall, N. R., Sakaki, M., Vasunilashorn, S., Nga, L., Somayajula, S., Chen, E. Y., ... Mather, M. (2012). Gender differences in reward-related decision processing under stress. *Social Cognitive and Affective Neuroscience*, 7(4), 476–484. <https://doi.org/10.1093/scan/nsr026>
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56(6), 1479–1498. <https://doi.org/10.2307/1130467>
- Liu, L., Wu, J., Geng, H., Liu, C., Luo, Y., Luo, J., & Qin, S. (2022). Long-term stress and trait anxiety affect brain network balance in dynamic cognitive computations. *Cerebral Cortex*, 32(14), 2957–2971. <https://doi.org/10.1093/cercor/bhab393>
- Liu, Y., & Lourenco, S. (2022). Affective factors affect visuospatial decision-making: A drift diffusion modeling approach. In , 44(44). *Proceedings of the annual meeting of the cognitive science society*. <https://escholarship.org/uc/item/9kd457qg>
- Liu, Y., & Lourenco, S. (2026). Dataset for “time pressure amplifies gender differences in mental rotation through shifts in evidence accumulation and response caution”. In , V1. *Mendeley data*. <https://doi.org/10.17632/pt7pz7mxm3.1>
- Lohman, D. F. (1986). The effect of speed-accuracy tradeoff on sex differences in mental rotation. *Perception & Psychophysics*, 39(6), 427–436. <https://doi.org/10.3758/BF03207071>
- 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*, 32(3), 187–196. <https://doi.org/10.1177/09637214231153072>
- Lyons, I. M., Ramirez, G., Maloney, E. A., Rendina, D. N., Levine, S. C., & Beilock, S. L. (2018). Spatial anxiety: A novel questionnaire with subscales for measuring three aspects of spatial anxiety. *Journal of Numerical Cognition*, 4(3), 526–553. <https://doi.org/10.5964/jnc.v4i3.154>
- Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdtke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, 10, 2767. <https://doi.org/10.3389/fpsyg.2019.02767>
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9780429029608>
- Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences*, 19(1), 21–27. <https://doi.org/10.1016/j.lindif.2008.02.002>
- Moser, J. S., Moran, T. P., Kneip, C., Schroder, H. S., & Larson, M. J. (2016). Sex moderates the association between symptoms of anxiety, but not obsessive compulsive disorder, and error-monitoring brain activity: A meta-analytic review. *Psychophysiology*, 53(1), 21–29. <https://doi.org/10.1111/psyp.12509>
- Nazareth, A., Huang, X., Voyer, D., & Newcombe, N. (2019). A meta-analysis of sex differences in human navigation skills. *Psychonomic Bulletin & Review*, 26(5), 1503–1528. <https://doi.org/10.3758/s13423-019-01633-6>
- Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, 5(5), 1. <https://doi.org/10.1167/5.5.1>
- Paschke, K., Jordan, K., Wüstenberg, T., Baudewig, J., & Leo Müller, J. (2012). Mirrored or identical—Is the role of visual perception underestimated in the mental rotation process of 3D-objects? A combined fMRI-eye tracking-study. *Neuropsychologia*, 50(8), 1844–1851. <https://doi.org/10.1016/j.neuropsychologia.2012.04.010>

- Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2. <https://doi.org/10.3389/neuro.11.010.2008>
- Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition*, 57(2), 176–184. <https://doi.org/10.1016/j.bandc.2004.08.052>
- Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition*, 66(3), 260–264. <https://doi.org/10.1016/j.bandc.2007.09.003>
- Pieters, R., & Warlop, L. (1999). Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of Research in Marketing*, 16(1), 1–16. [https://doi.org/10.1016/S0167-8116\(98\)00022-6](https://doi.org/10.1016/S0167-8116(98)00022-6)
- Provost, A., & Heathcote, A. (2015). Titrating decision processes in the mental rotation task. *Psychological Review*, 122(4), 735–754. <https://doi.org/10.1037/a0039706>
- Ramírez-Uclés, I. M., & Ramírez-Uclés, R. (2020). Gender differences in visuospatial abilities and complex mathematical problem solving. *Frontiers in Psychology*, 11, 191. <https://doi.org/10.3389/fpsyg.2020.00191>
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873–922. <https://doi.org/10.1162/neco.2008.12.06.420>
- Robert, M., & Chevrier, E. (2003). Does men's advantage in mental rotation persist when real three-dimensional objects are either felt or seen? *Memory & Cognition*, 31(7), 1136–1145. <https://doi.org/10.3758/BF03196134>
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703.
- Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed–accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging*, 25(2), 377–390. <https://doi.org/10.1037/a0018022>
- Theisen, M., Lerche, V., Von Krause, M., & Voss, A. (2021). Age differences in diffusion model parameters: A meta-analysis. *Psychological Research*, 85(5), 2012–2021. <https://doi.org/10.1007/s00426-020-01371-8>
- Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science*, 22(5), 367–373. <https://doi.org/10.1177/0963721413484756>
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47(2), 599–604. <https://doi.org/10.2466/pms.1978.47.2.599>
- Vats, D., & Knudson, C. (2021). Revisiting the Gelman–Rubin diagnostic. *Statistical Science*, 36(4). <https://doi.org/10.1214/20-STS812>
- Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). I. Spatial skills, their development, and their links to mathematics. *Monographs of the Society for Research in Child Development*, 82(1), 7–30. <https://doi.org/10.1111/mono.12280>
- Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: A meta-analysis. *Psychonomic Bulletin & Review*, 18(2), 267–277. <https://doi.org/10.3758/s13423-010-0042-0>
- 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. <https://doi.org/10.3758/BF03194024>
- Voyer, D., Rodgers, M. A., & McCormick, P. A. (2004). Timing conditions and the magnitude of gender differences on the Mental Rotations Test. *Memory & Cognition*, 32(1), 72–82. <https://doi.org/10.3758/BF03195821>
- Voyer, D., Saint-Aubin, J., Altman, K., & Doyle, R. A. (2020). Sex differences in tests of mental rotation: Direct manipulation of strategies with eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 46(9), 871–889. <https://doi.org/10.1037/xhp0000752>
- Voyer, D., & Saunders, K. A. (2004). Gender differences on the mental rotations test: A factor analysis. *Acta Psychologica*, 117(1), 79–94. <https://doi.org/10.1016/j.actpsy.2004.05.003>
- 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>
- Voyer, D., Voyer, S. D., & Saint-Aubin, J. (2017). Sex differences in visual-spatial working memory: A meta-analysis. *Psychonomic Bulletin & Review*, 24(2), 307–334. <https://doi.org/10.3758/s13423-016-1085-7>
- Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, 58(1), 140–159. <https://doi.org/10.1016/j.jml.2007.04.006>
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817–835. <https://doi.org/10.1037/a0016127>
- Warren, S. L., Zhang, Y., Duberg, K., Mistry, P., Cai, W., Qin, S., ... Menon, V. (2020). Anxiety and stress alter decision-making dynamics and causal amygdala-dorsolateral prefrontal cortex circuits during emotion regulation in children. *Biological Psychiatry*, 88(7), 576–586. <https://doi.org/10.1016/j.biopsych.2020.02.011>
- Wenzlaff, H., Bauer, M., Maess, B., & Heekeren, H. R. (2011). Neural characterization of the speed–accuracy tradeoff in a perceptual decision-making task. *Journal of Neuroscience*, 31(4), 1254–1266. <https://doi.org/10.1523/JNEUROSCI.4000-10.2011>
- White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, 63(4), 210–238. <https://doi.org/10.1016/j.cogpsych.2011.08.001>
- Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, 7. <https://doi.org/10.3389/fninf.2013.00014>
- Wiker, T., Pedersen, M. L., Fersmann, L., Beck, D., Norbom, L. B., Dahl, A., ... Tamnes, C. K. (2024). Assessing the longitudinal associations between decision-making processes and attention problems in early adolescence. *Research on Child and Adolescent Psychopathology*, 52(5), 803–817. <https://doi.org/10.1007/s10802-023-01148-8>
- Zhang, J., & Rowe, J. B. (2014). Dissociable mechanisms of speed–accuracy tradeoff during visual perceptual learning are revealed by a hierarchical drift-diffusion model. *Frontiers in Neuroscience*, 8. <https://doi.org/10.3389/fnins.2014.00069>