A Computational Model of Mental Rotation in Aphantasia

Shiu-Cheng Wang[1,2] and Tomer Ullman[1]
[1]Department of Psychology, Harvard University
[2]Neuro-X Institute, École Polytechnique Fédérale de Lausanne (EPFL)

Abstract

People with aphantasia seem unable to create visual imagery, yet they perform mental manipulation of objects in the mind. Different frameworks have different views on how this is possible. Mental rotation is a particularly useful test case for comparing these frameworks, as people with aphantasia show systematic differences and similarities on this task compared to people without aphantasia. Here, we propose a computational model of mental rotation based on the 'broken rendering' account of aphantasia (Balaban & Ullman, 2025), which separates physical/spatial representations from graphical, picture-like ones. Our model proposes that both people with and without aphantasia are similar in their identification and rotation of objects, but differ in the comparison stage: People with and without aphantasia can use spatial/physical similarity to compare rotated objects, but people with typical imagery can also render a rotated object back into a 2D, pixel-like representation, and carry out a fast-but-less-reliable comparison between such visual depictions. Both sources of comparison (physical and visual) provide evidence to a single drift-diffusion decision process. We used our model to generate simulated experimental data, which we compared to the detailed findings of Kay et al., 2024. We found that our model captures the qualitative and quantitative patterns of behavior in a mental rotation task for people with and without aphantasia, including the canonical increase in response time (and decrease in accuracy) with angle growth, the slower but more accurate responses in aphantasia, and the opposite within-group correlations. This lends support to the broader framework of 'broken rendering' over 'different strategies' and 'lack of meta-access' as approaches to aphantasia. The findings hold across a wide set of parameter settings, and the model makes novel and testable predictions for further study in mental rotation.

Keywords: aphantasia, mental imagery, mental rotation, intuitive physics, computational modeling

Correspondence concerning this article should be addressed to Tomer Ullman, tullman@fas.harvard.edu

Introduction

We are not passive audience members in the theater of the mind, we are directors. We can manipulate the mental images that come before us, moving them about and changing them. One of the most well-studied mental manipulations of this sort is mental rotation: people seem able to take an imagined item and turn it around in various ways (Shepard & Metzler, 1971). A compelling feature of mental rotation is the linear relationship between the time it takes people to mentally rotate an object, and the real angular distance that same object would travel, were it to exist in reality (Cooper & Shepard, 1973; Shepard & Cooper, 1982/1986; Shepard & Metzler, 1971). This linear relationship is often taken as evidence that the format of mental imagery resembles that of perception (Kosslyn et al., 2006), and that the manipulation of mental images proceeds through mental simulation (Shepard & Cooper, 1982/1986). The mental rotation task has been studied in many ways over the decades, including neurally (Tomasino & Gremese, 2016; Zacks, 2008), behaviorally (Just & Carpenter, 1976; Shepard & Cooper, 1982/1986), developmentally (Frick et al., 2014), and computationally (Khazoum et al., 2026; Provost & Heathcote, 2015). The surrounding claims regarding the format of mental imagery have also been hotly debated (Just & Carpenter, 1976; Kosslyn et al., 2006; Pylyshyn, 2002). The main question for us here is: how can a director manipulate something that they cannot see?

When we imagine things, many of us seem to *see* something. But, it seems that some of us cannot see things in the mind's eye at all. Over the past decade, there has been a tremendous amount of research and interest in aphantasia, a condition in which people cannot voluntarily create visual mental images (Kay et al., 2022; Keogh & Pearson, 2018; Milton et al., 2021; Wicken et al., 2021; A. Zeman et al., 2015). For people with aphantasia, the stage before the mind's eye seems dark or non-existent. And yet, people with aphantasia are apparently able to direct the goings-on in the mental stage: many studies suggest that people with aphantasia are able to carry out a wide range of imagery tasks (Bainbridge et al., 2021; Milton et al., 2021; Pounder et al., 2022; A. Z. Zeman et al., 2010; Zhao et al., 2022). How can this be?

The views on how people with aphantasia carry out mental imagery tasks can be roughly sorted into 3 camps: (1) **Different strategies**. In this view, people with aphantasia use non-imagery based strategies to carry out mental imagery tasks (Kay et al., 2024; Reeder et al., 2024). If this were true, it would not be surprising. People use various mental strategies to carry out the same operations. But, such a view needs to account for the cases in which people with and without aphantasia seem to be broadly *similar*. While it is not unreasonable to suggest that different strategies converge in their final output, it does seem like a suspicious coincidence that different strategies would just-so-happen to produce similar behavioral profiles on the same task (including in mental rotation), and a person who holds this view would need to spell out the *specific* strategies people with aphantasia supposedly rely on for this to be a testable alternative; (2) **Same format, but lack of meta-access**. In this view, people with and without aphantasia both create mental images in the same visual-perceptual format, and operate on them in similar ways. It is just that people with aphantasia lack meta-access to these perceptual images (Michel et al., 2024; Nanay,

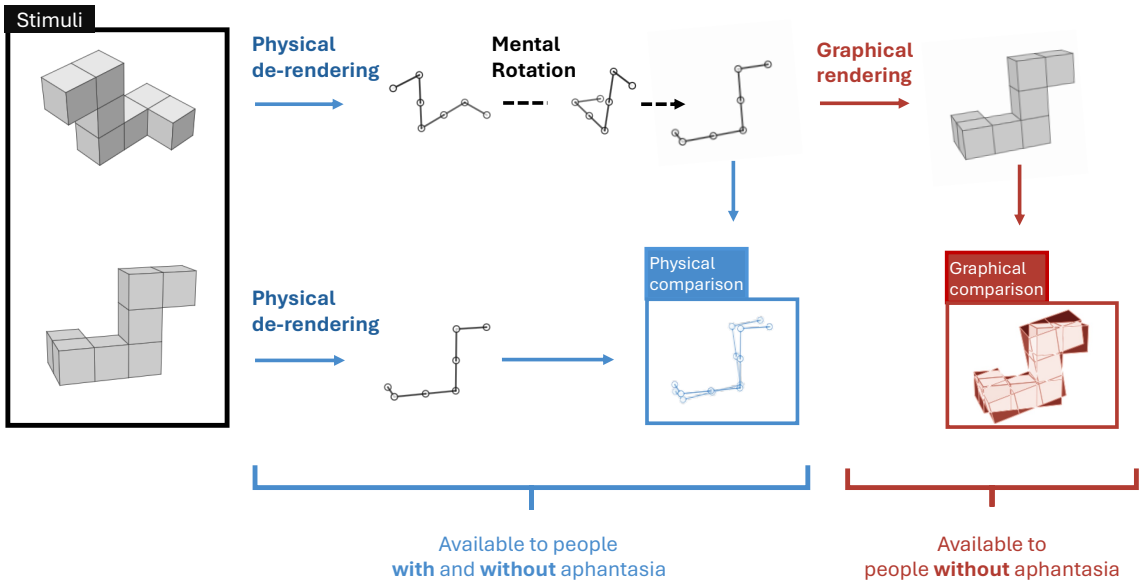


Figure 1

Schematic of proposed computational model of mental rotation in people with and without aphantasia. *According to this proposal (based on Balaban & Ullman, 2025), when tasked with a mental rotation task such as determining whether two images are of the same object (Stimuli), both people with and without aphantasia can "de-render" the objects into a physical/spatial representation, and perform mental operations such as rotation on the physical representations. Physical representations can be compared for alignment and similarity judgements, e.g. by comparing the centers of nodes along a wire-frame representation. In addition, people without aphantasia may 'render' a physical representation back into a graphical, 2D, picture-like representation. Such pictorial, graphical representations may be compared using similarity functions suitable for pixel-based comparisons. Each of these comparisons (physical and graphical) is combined into the evidence signal in a decision-making model.*

2021). Such a view easily accommodates similar behavioral patterns between people with and without aphantasia, and also finds recent support in neuroscience (Chang et al., 2025; Liu et al., 2025). However, it does not easily account for the experiments that show people with aphantasia truly do not seem to create visual images, including those based on binocular rivalry speed-up (Keogh & Pearson, 2018), skin conductance studies (Wicken et al., 2021), imagery pupillary light response (Kay et al., 2022), or implicit priming (Purkart et al., 2025). Nor does this view account for the fact that even within the similarities between people with and without aphantasia, there are systematic differences between them (Kay et al., 2024; Milton et al., 2021; Pounder et al., 2022), such as the case of mental rotation; (3) **Broken rendering**. This view (Balaban & Ullman, 2025; Phillips, 2025) be-

gins with a separation between *physical/spatial* imagery and *visual/graphical* imagery. In this view, when people create mental images, they go through a hierarchical process akin to the process that creates images in computer simulations and animations. First, there is the spatial/physical scaffolding of rough object bodies and their relations to one another. Only later, it is an optional, separate operation to create a 2D representation of what such a scene looks like from a particular perspective (an operation termed *rendering* in computer graphics). People with aphantasia do not go through the rendering operation, and so do not create a 2D, pixel-like representation of what a scene looks like from a particular view. But, rendering is not necessary for many mental imagery tasks, and most tasks attributed to mental imagery can be carried out through spatial/physical representations. While visual images may be non-functional in many tasks, some downstream processes that rely on 'picture-like' formats may be affected, e.g. certain emotional responses (Balaban & Ullman, 2026).

Mental rotation is a particularly useful case study for teasing apart the different proposals for how aphantasia might work. It is particularly well-studied, and research shows that people with aphantasia can generally perform mental rotation tasks (Kay et al., 2024; Milton et al., 2021; Pounder et al., 2022; A. Z. Zeman et al., 2010; Zhao et al., 2022), but their performance shows both similarities and differences compared to people without aphantasia. Models of aphantasia should account for both the similarities and differences. Kay et al., 2024, with the largest sample size to date (N=95 for aphantasia, N=114 for control), found the following 5 characteristics in particular:

1. People with and without aphantasia show the canonical pattern: the reaction time grows linearly with the angular distance between two mental objects, and the accuracy drops as this angular distance grows.

2. People with aphantasia show a longer reaction time on this task.

3. People with aphantasia have better accuracy on this task.

4. There is a positive correlation between accuracy & reaction time in the control group.

5. There is a *negative* correlation between accuracy & reaction time in the aphantasia group.

These patterns can also be seen in Figure 2 (top panels). Note that while a classic accuracy/reaction-time tradeoff can explain points 2-4, it falls apart at point 5. That is, it is *not* the case that people with aphantasia are simply taking longer with mental rotation tasks and, therefore, being more accurate.

How do the different views on aphantasia account for the pattern of similarities and differences in mental rotation? A 'Different Strategies' view could in principle account for it, but one would need a specific alternative for what different strategy people with aphantasia are using, and to show how it specifically leads to both the suspicious coincidence

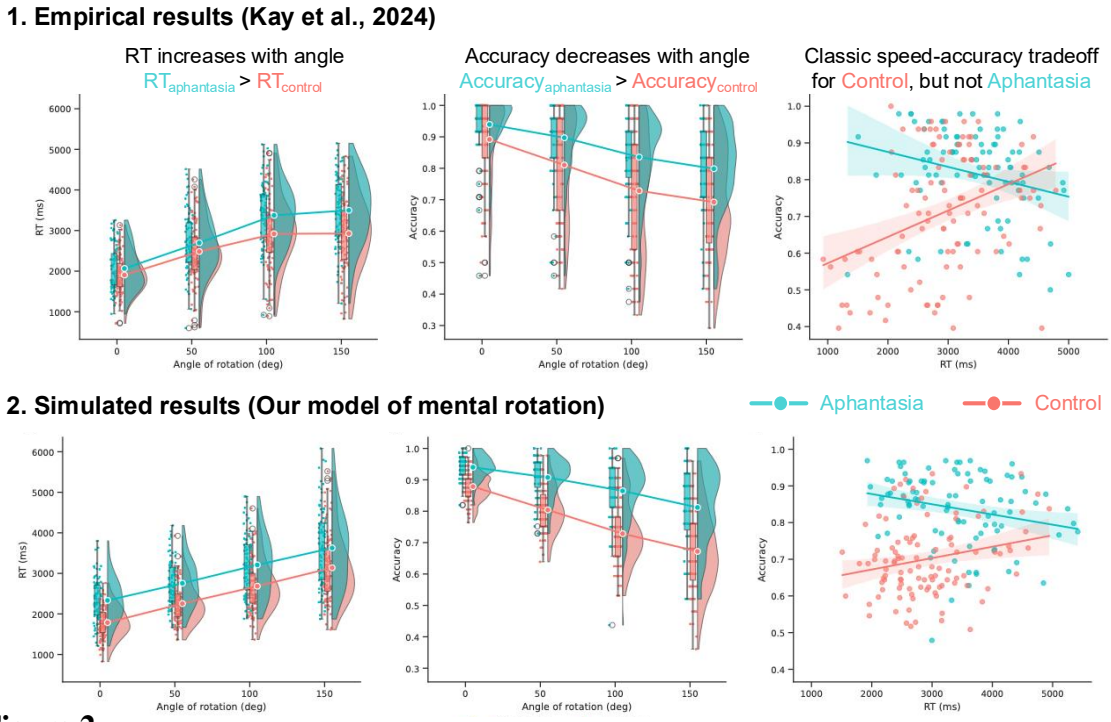


Figure 2

Comparison of empirical findings and our model. Top panels: *A recreation of the results of Kay et al., 2024, testing mental rotation in people with and without aphantasia. Each dot denotes a participant's mean RT or accuracy. For the Accuracy/RT comparison (right-most panel), these are averaged over the* 100° *and* 150° *conditions, and shaded bands show 95% confidence intervals from bootstrap estimation when fitting the least-squares lines.* ***Bottom panels:*** *Example results from one simulated experiment, using our model of mental rotation. The results in the main text refer to 1000 such simulations. The model captures the main patterns of the original study.*

that people with and without aphantasia show the canonical pattern (item #1), but also the differences in items #2-5. A 'Lack of Meta-Access' view can easily explain the similarity (item #1), but why would 'lack of meta-access' lead to the particular differences (patterns #2-#5)? A 'Broken Rendering' view could in principle account for all of these patterns (Balaban & Ullman, 2025), but one would need to put forward a specific, quantitative model based on this proposal to be convincing.

Here, we put forward a specific, testable, computational model of mental rotation in aphantasia, based on the 'broken rendering' proposal (see Figure 1). This model modifies the classic information-processing model of mental rotation proposed by Shepard and Cooper, 1982/1986. Specifically, we suggest that people with and without aphantasia are *similar* in their mental operations when it comes to determining the identity and orientation of the stimuli (Stage 1), that they are *similar* in rotating the objects (Stage 2), and that

they *differ* when it comes to checking whether the rotated objects are the same (Stage 3). For Stage 3, we suggest that people with and without aphantasia can use spatial/physical alignment to determine whether the physical-representations of the objects match. But, in addition to this, people without aphantasia can "render" a rotated object, and so can use a graphical-similarity comparison as an added source of information for making their final decision. Such an added source of information can potentially lead to faster judgements for people without aphantasia, but also introduce additional error[1].

The rest of this paper is organized as follows: We first detail our computational model, including the background assumptions and relation to previous proposals. The aim is to explain the model's important details, while not bogging down the readers who prefer a higher-level summary of the model's assumptions. To strike a balance, we specify some of the particulars of the model in Appendix A at the end of this paper. We then compare the results of our computational model with those of Kay et al., 2024, since it offers the most detailed and well-powered recent examination of mental rotation in aphantasia. To summarize these results upfront, we found that our model can indeed reproduce patterns #1-5 qualitatively and quantitatively. We end with a discussion of the additional testable predictions that our model makes, which current empirical findings do not speak to (yet), and the ways to test them. We also consider how our model relates to the ongoing debates surrounding aphantasia, and mental imagery more broadly.

Computational Model

Theoretical Framework

Following Shepard and Cooper, 1982/1986, we focus on three stages of a mental rotation task: stimulus identification, mental rotation, and comparison. Continuing the broken-rendering account (Balaban & Ullman, 2025), we assume the first two stages are the same in people with and without aphantasia. For stimulus identification, this assumption is backed up by evidence that shows similar visual working memory capacity for people with and without aphantasia (Keogh et al., 2021). For the mental rotation stage, the assumption of similarity between people with and without aphantasia is a hypothesis, based on the separation between "physics" and "graphics" and the suggestion that physical scene representation is intact in aphantasia. For the stimulus comparison stage, we propose that people with and without aphantasia can both make similarity judgements on the basis of physical object representations (the same simplified physical/spatial representations that are being rotated in the second stage). However, in addition to this, people without aphantasia can "render" physical/spatial representations in 2D graphical representations, and use those as an additional source of evidence for similarity judgements.

We model the actual *rotation* stage of the mental rotation task by aligning the probe figure and the reference figure to the degree that best minimizes the least-square distances

[1] Our proposal relies on a simplified version of the Shepard and Cooper 7-step model of mental rotation, which includes parts that are not relevant for our purposes. For example, we do not model the motor execution of the decision, and assume it is shared between people with and without aphantasia.

in the three-dimensional space. The alignment is only about the vertical axis, as that is the only axis of rotation in the stimulus set (Ganis & Kievit, 2015). We also introduce noise to account for imprecision in people's representations or mental operations. Specifically, we add a small random error $\phi \sim \mathcal{N}(0, s^2_{\text{jit}})$ about the same vertical axis of rotation. We considered four discrete levels of s_{jit}, $s_{jit} \in \{20, 25, 30, 35\}^\circ$.

Turning to the comparison stage: The two sources of evidence (physical/spatial and imagistic/graphical) feed into a single drift diffusion model (Ratcliff & McKoon, 2008). In what follows, we use the subscript S to denote the spatial/physical similarity route, and the subscript I to denote imagistic/graphical similarity route. We denote the "spatial dissimilarity" function as f_S, and "graphical" dissimilarity function as f_I. There are various (dis)similarity functions that can be used for either f_S or f_I, and our proposal is not strictly tied to one particular function. Here, specifically, we used:

- f_S: The input is the 3D representation of the objects shown in the stimuli (post rotation and alignment, if relevant). The output is the spatial/physical dissimilarity between the two objects. To make this concrete, we used the root-mean-square distance between the two arrays of cube centers, but other spatial/physical similarity metrics can be considered.

- f_I: The input is the 2D, pixel-based representation of the objects shown in the stimuli (post rotation, and re-rendering, if relevant). The output is the visual/graphical dissimilarity between the two objects. To make this concrete, we used the root-mean-square difference of pixel intensity between the two pixel images, but other visual similarity metrics can be considered.

To reiterate, we assume people with and without aphantasia make decisions of 'same' or 'different' in mental rotation based on the spatial route through f_S, and that in addition to this people without aphantasia have an additional source of information through f_I. Since the input to f_S and f_I rely on the same rotated object, their evidence is correlated, but it is not identical. We weight the additional information of f_I by $\alpha_j \geq 0$, which is specific a person's reliance on visual imagery. The aphantasia group has $\alpha_j = 0$.

In summary, the architecture of our model includes two routes of evidence that feed into a single decision process, modeled via drift-diffusion.

A hierarchy of populations, people, and trial-by-trial differences

Our computational model describes the populations of people performing mental rotation in different trials. It has three levels:

- **Population.** We assume there are two sub-populations/groups, people with and without aphantasia. People in the aphantasia group have zero reliance on the visual route described above, but share all the other population-level distributions.

- **Person.** Each simulated participant has individual-level characteristics describing their degree of reliance and proficiency with spatial and visual comparison (Appendix A). In line with standard drift-diffusion models for decision making, participants also have individual differences when it comes to non-decision time, and overall speed.

- **Trial.** Each trial presents a participant with a new pair of objects in a mental rotation task. The model draws read-outs for this pair, and then runs a simulation of a drift-diffusion decision with this participant's specific characteristics.

The available data from Kay et al., 2024 include per-participant summaries, mean correct RT, and accuracy for each angle $\times$ {mirror versus same} combination. Given this, we simulate the whole experiment, and compute these same summaries for each simulated participant. We then compare the summary statistics from our model with the experimental data.

Relationship to previous models. We note that ours is certainly not the first model of mental rotation. As mentioned, our overall framework builds on and modifies the model of Shepard and Cooper, 1982/1986. That model did not provide a process-level account connecting to a decision-making process, with differences between the comparison functions available to people with and without aphantasia in a way that can account for the data we consider. A more recent process-level model of mental rotation by Provost and Heathcote, 2015 fit the rate of evidence accumulation from behavioral data, but does not specify how the evidence is collected, nor relate to aphantasia. An even more recent model, by Khazoum et al., 2026, proposed a deep-learning model of mental rotation, which showed behavioral results comparable to those of humans. However, this model too did not consider varying similarity functions that split between graphics and physics, nor try to account for mental rotation in aphantasia.

Further model details. A full technical specification of the model includes further details on how evidence is computed from the stimuli, specific choices of parameters regarding the drift-diffusion, and how individual (simulated) participants differ. We give these full details in Appendix A.

Experiment: Simulating mental rotation in aphantasia

Stimuli

We use the custom rendition of three-dimensional cube objects made by Ganis and Kievit, 2015, to match the same stimulus set used in Kay et al., 2024. An example is shown in Figure 3. The objects were built in Python for our purpose of obtaining their 3D structure (cube centers), while the rendering choices, such as lighting and view point, remain as close to Ganis and Kievit, 2015 as possible. To err on the side of caution, we excluded 6 stimuli pairs from the original 48 due to them being either achiral, or not mirrors of each other. This leaves us with a set of 42 image pairs.

We computed 1000 *same* and 1000 *mirror* trials for each pair of figures, and at each level of imprecision, $s_{jit} \in \{20, 25, 30, 35\}^\circ$. These samples are the model's "readout bank", and all following simulated experiments draw from them.

The two routes already differ in their discriminability before setting any parameters (Figure A1). At $s_{jit} = 25^\circ$, physical/spatial comparison separates *same* and *mirror* pairs better than imagistic/graphical comparison for all 42 figures, with a median discriminability of 2.90 against 1.06 (Figure B1; Appendix B).

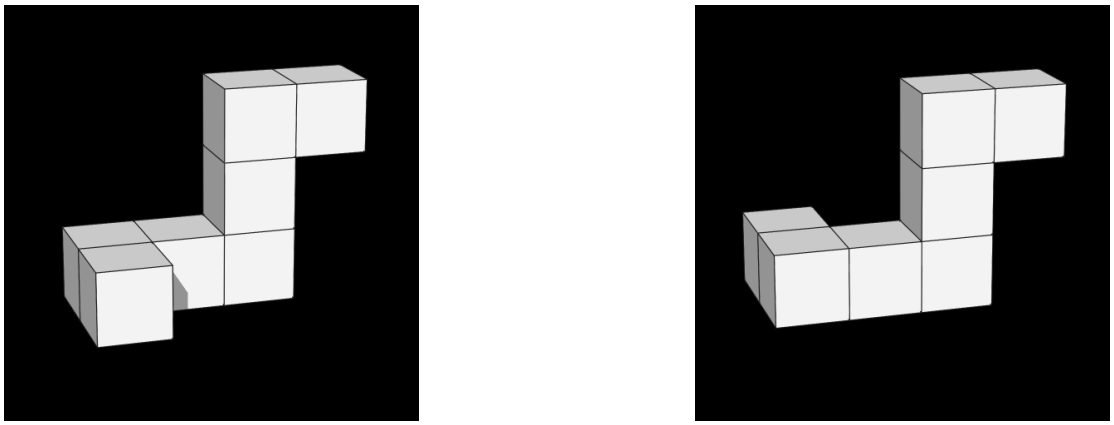


Figure 3

Example stimuli. *One example used in our simulated experiments, to match the set used in Kay et al., 2024. Note that in comparing these two objects, the answer should be 'different'.*

Method

We simulated 1000 experiments at a representative parameter setting, chosen from a fixed rule (Appendix E). Each simulated experiment reuses the original experimental design of Kay et al., 2024, including 91 participants with aphantasia, and 107 controls. A simulated participant in either group was tested on 12 randomly drawn figures out of the set of 42. Each of the figures was presented with a same pair and a mirror pair, crossed with four angle disparity values, $\theta \in \{0, 50, 100, 150\}^\circ$, which totals 96 trials per participant. In each trial, a readout pair is sampled from the bank of the figure at the predetermined imprecision level, $s_{jit} = 25^\circ$.

We analyzed our simulated experiments following the same procedure as in Kay et al., 2024. The accuracy of a participant at each angle is the proportion of their correct responses. A participant's RT is counted for analysis when the response is correct, consistent with Kay et al., 2024, though we note Shepard and Cooper, 1982/1986 further excluded the *mirror* trials. The between-subject correlations are calculated on each participant's means in the 100° and 150° conditions. Following Kay et al., 2024, we use Spearman's correlation

to measure the correlation between RT and accuracy in the aphantasia group, and Pearson's correlation for the control group, with both tested at $\alpha = 0.0125$.

We relied on the data in Kay et al., 2024 to inform our model in two ways. First, the shared rotation stage used the group-level summary statistics from Kay et al., 2024's data (including the mean RT, accuracy at each angle, and the spread of RT across participants, see Appendix A). We assume that this stage is the same for the two groups, and it can only attenuate the within-group correlations but not alter the specific difference between the two groups. Second, our choice of representative parameter settings favored parameter settings with similar size in effect to Kay et al., 2024. We emphasize however that the resulting patterns do not come simply from taking the original data. The main findings we describe below (increase of RT with angle, decrease of accuracy with angle, aphantasia group is slower but more accurate, and the opposite signs in within-group correlations) hold in 96.9% of a broad region of parameter settings (Appendix D).

Results

Figure 2 shows the data from Kay et al., 2024 compared to a typical simulated experiment. As can be seen, the two agree with one another on the main patterns. Moving to a more formal comparison: we compared our model's results to human data in two ways, which are not necessarily consistent. The first analysis examined where the empirical data fell in the distribution of our 1000 simulated experiments (Palminteri et al., 2017). For each statistic, we estimate p by taking the 1000 simulated values of the statistic as the (approximate) distribution the model predicts, (Table 1). In the second analysis (summarized in Table 2), we examined whether applying the same statistical tests on our experiments gives the same results as the human data (Wilson & Collins, 2019). The model values below are medians, with 2.5% to 97.5% of the 1000 simulated experiments shown in brackets.

People with aphantasia are slower but more accurate. Our model showed the same pattern that both groups are slower and less accurate at larger angles, which is from the built-in shared rotation stage. Kay et al., 2024 found that people with aphantasia were slower than controls by 351 ms, and more accurate by 8.7%, across all angles. Our model shows a slowdown of 549 ms ([374, 711]) and increase in accuracy of 10.9% ([8.8%, 12.9%]). We note that this accuracy difference between the two groups is ever-so-slightly larger than observed ($p = 0.036$), but we do not take this to be meaningful.

We note a main difference in RT between our model and people, at small angles (0° and 50°). Specifically, our model predicts an RT difference between people with and without aphantasia throughout, whereas the empirical findings show an increase in the RT difference, due to a very small (or possibly non-existent) RT difference at small angles. Such a difference is easily accounted for in principle: For small angles (or even no rotation, 0°), mental-rotation may not be necessary. Since both images are still available to the participants perceptually, it is possible in principle for people with aphantasia to make use of graphical similarity functions to compare the two images. Our model is more similar to a situation in which one (or both) of the original images are removed from perception. We can easily amend our model to account for this, but preferred to keep things simpler in the

Statistic	Kay et al.	Our model	p (size)
RT difference (ms), all angles	351	549 [374, 711]	**.018**
over 100°–150°	513	549 [327, 771]	.75
Accuracy difference, all angles	.087	.109 [.088, .129]	**.036**
over 100°–150°	.107	.145 [.117, .172]	**.010**
Aphantasia r, Spearman	$-.33$	$-.25$ [$-.43, -.06$] ($-.72$)	.45
Controls' r, Pearson	$+.36$	$+.20$ [$.00, +.38$] ($+.53$)	.078
Aphantasia b (ms)	-198	-201 [$-369, -50$]	.95
Controls' b (ms)	$+307$	$+163$ [$+2, +309$]	.054

Table 1

Comparing Kay et al., 2024's data to the null distribution from simulated experiments of our model. *We tested the statistics Kay et al., 2024 reported against the null distribution from our model. The table lists the median and central 95% of* 1000 *simulated experiments of their size, with between-participant speed differences estimated from their data (comparison stage alone in parentheses for correlations). All differences are aphantasia minus controls. Both correlations and bs are over* 100°-150°. *p (size): two-sided p value of Kay et al.'s statistics, with the model's distribution as the null distribution. With* 1000 *experiments, the smallest possible p is* .002.

main analysis. We discuss this further in the Discussion.

Opposite speed-accuracy correlations between the two groups. With between-participant speed spread matching the data, the aphantasia group's RT-accuracy Spearman's correlation is -0.25, with 95% of experiments showing a correlation going from -0.43 to -0.06. This is comparable to the empirical data correlation of -0.33 ($p = 0.45$). The Pearson's correlation for the simulated control group is 0.20, with 95% of experiments showing a correlation that ranges from 0.00 to 0.38. This is comparable to the empirical data correlation of 0.36 ($p = 0.078$). Kay et al., 2024's data is around the upper limit of the model's range.

From the comparison stage alone, without the added speed spread from the shared rotation stage, the correlations are stronger than observed, -0.72 for the aphantasia group ([-0.80, -0.60]) and +0.53 for the control group ([0.37, 0.65]). As the shared rotation stage cannot adjust the RT spread differentially in the two groups, we use $b = r \cdot SD(RT)$, the standard deviation of RT that is related to accuracy within the two groups, which the rotation stage leaves unchanged (see Appendix E). The model reproduces b in aphantasia (-201 ms for the model, vs. -198 ms for people; $p = 0.95$), but only half of it in the control group (163 ms for the model, vs. 307 for people, $p = 0.054$).

Applying Kay's statistical tests on our simulated data matches most of their conclusions. We run the same statistical tests on each of our simulated experiments, to see whether the conclusions agree (Table 2). Overall, the model achieved 10 of 14 conclusions in Kay et al., 2024's results regarding group and angle in at least 96% of the experiments.

The four non-matched conclusions all concern small angles, which can be explained as discussed above. For RT, all of the simulated experiments replicate the main effects of group and angle, but not their interaction, where only 5% of the simulated experiments are significant. Multiple comparisons show larger RT for aphantasia at larger angles, 100° and 150°, for almost all experiments. However, our model does not reproduce the non-significant RT differences at 0° and 50° (only showing in less than 1% of the simulations). For accuracy, all of the simulated experiments replicate the main effects of groups, that of angles, their interaction, and multiple comparison except at 0° (only 2% of the simulated experiments are non-significant at 0°).

	Test	Kay et al.	Model	Same conclusion
RT	group	$F = 13.5$, $p < .001$	$F = 39.9$ [19.0, 70.7]	100%
	angle	$F = 506.7$, $p < .001$	$F = 415$ [336, 502]	100%
	group × angle	$F = 14.6$, $p < .001$	$F = 0.5$ [0.0, 4.6]	5%
	at 0°	$d = .22$, n.s.	$d = .79$ [.59, 1.01]	< 1%
	at 50°	$d = .29$, n.s.	$d = .80$ [.56, 1.03]	< 1%
	at 100°	$d = .62$, $p < .001$	$d = .80$ [.51, 1.10]	99%
	at 150°	$d = .77$, $p < .001$	$d = .80$ [.42, 1.17]	96%
Accuracy	group	$F = 22.8$, $p < .001$	$F = 111$ [66, 178]	100%
	angle	$F = 144.9$, $p < .001$	$F = 242$ [201, 291]	100%
	group × angle	$F = 4.9$, $p = .002$	$F = 24.9$ [13.6, 40.5]	100%
	at 0°	$d = .32$, n.s.	$d = .62$ [.46, .77]	2%
	at 50°	$d = .58$, $p = .002$	$d = 1.00$ [.75, 1.28]	100%
	at 100°	$d = .72$, $p < .001$	$d = 1.39$ [1.04, 1.75]	100%
	at 150°	$d = .71$, $p < .001$	$d = 1.77$ [1.34, 2.22]	100%
Correlation	aphantasia, $r < 0$	$r_s = -.33$, $p = .001$	$r_s = -.25$ [−.43, −.06]	47%
	controls, $r > 0$	$r = +.36$, $p < .001$	$r = +.20$ [.00, +.38]	34%
	both	yes (68%)[a]	—	16%

Table 2

Results of running the same statistical tests on Kay et al., 2024's and model data from 1000 ***simulated experiments.****For RT and accuracy, mixed ANOVAs of group as between-participant variable and angle as within-participant variable was used. Bonferroni corrections are applied to comparisons of the groups at each angle (Cohen's d, aphantasia minus controls). The within-group correlation is also tested. The table shows the median and the central 95% of all simulated experiments. Same conclusion column shows the percentage of experiments that reach Kay et al.'s conclusion, with type I error level at 0.0125 for the correlations, and 0.05 for the others. [a]Percentage of 5000 bootstrap drawn from Kay et al.'s own participants in which both correlations are significant.*

For the within-group correlations, 47% of the simulated experiments with Kay et al., 2024's sample size find significance in the aphantasia group, 34% in the control group,

and both 16% (Type I error rate = 0.0125). A bootstrap analysis using Kay et al., 2024's original participants with replacement finds 74% for aphantasia, 92% for controls, and 68% for both, but this could be over-optimistic. At conventional Type I error = 0.05, our simulated experiments give 70%, 55%, and 39%, against 86%, 98%, and 84% in bootstrap sampling.

Discussion

Mental rotation is just one example of a mental operation that people can carry out on objects of the mind (Shepard & Cooper, 1982/1986). But it is a particularly important one, and one that has been studied, turned over, and examined from many angles. It is also a particularly useful case-study for comparing different frameworks that try to account for the phenomena of aphantasia. What framework can account for the fact that people with aphantasia show systematic similarities and differences in mental rotation, when compared to people who can create visual imagery?

Here, we proposed a process-level account of mental rotation that can account for how it is carried out in aphantasia. Our model was based on the more general proposal of 'Broken rendering' (Balaban & Ullman, 2025), which itself is based on a basic separation in the mind between "physics" and "graphics". Our model suggests that people with and without aphantasia share many of the same operations of mental rotation: both types of people recognize and de-render the stimuli into physical/spatial representations, both rotate it using physical-based operations, and both can compare the results using physics/spatial-based similarity functions. However, in addition to this, people without aphantasia can 'render' the rotated object in a 2D graphical representation, which allows for an additional comparison route via similarity functions that expect pixel-like, 2D visual input. We further operationalized these two routes (physics and graphics) as sources of evidence for a drift-diffusion decision model.

We found that our model reproduced the main behavioral characteristics shown by people with and without aphantasia when doing a mental rotation task (Kay et al., 2024), including both the similarities and the differences. The model, like people, shows increased overall RT and lower accuracy with rotation angle, a higher RT and higher accuracy for aphantasia, and a speed-accuracy tradeoff for control, but not aphantasia. While the model needs to specify a diverse set of parameters to account for various things from the speed of rotation to the noise and variance in each route, we found similar conclusions in 97% of a wide-yet-reasonable parameter space (Appendix D). This suggests that the conclusions are based on the overall model structure, rather than a hyper-specific choice of parameters.

We take the ability of our model to capture diverse findings in mental rotation to be more generally supportive of the 'broken rendering' account. The general notion is that people with and without aphantasia can carry out spatial/physical operations supported by mental simulation, akin to how a game/physics engine can be used to run a simulated world forward without necessarily rendering it graphically (Balaban & Ullman, 2025; Ullman et al., 2017). How does this 'broken rendering' account fare against the other two main alternatives, roughly categorized as 'lack of meta-access' and 'different strategies'? The 'lack

of meta-access' account can explain the similarities in performance between people with and without aphantasia, but we are hard-pressed to see how it accounts for the specific differences. The 'different strategies' account could perhaps come up with dual strategies that can account for the similarities and differences, but if so it would be less parsimonious: our view does not require an entirely different strategy, but merely one reasonable difference in an additional available representation. Further, we ourselves could not come up with a truly alternative, reasonable 'dual strategy' model that just-so-happens to fit the similarities and differences in the empirical data. This does not mean such models/strategies don't exist, but we believe the onus is now on supporters of that view to come up with them and put them to a rigorous comparison with the data and our model.

While our model is generally in line with the findings of Kay et al., 2024, we do recognize that it diverges from it for small angles, and most especially for cases where there is no angle-difference between the two objects that are shown for comparison. For example, our model predicts that people with aphantasia will be slower to respond when the difference between the two objects is 0°, but empirically it appears that there is no RT difference in such cases. However, this difference can be easily accounted for: When the two stimuli images do not differ in angle, the information is already in 2D visual format and is available to people with and without aphantasia, without the need for a 'rendering' process. In such a case, people with aphantasia could use the same similarity functions as people without aphantasia, and we should expect little to no difference between them. This is in line with other studies, which show that for perceptual tasks people with and without aphantasia perform equally well (Bainbridge et al., 2021). Our model in its current form assumes that objects are de-rendered into a physical representation and then (for people without aphantasia) rendered back. This is more akin to an experimental set-up in which the two objects are not presented simultaneously, and one has to be kept in memory. We could simply augment our model with the additional assumption that for small angles no re-rendering is necessary for either group, but opted to keep the model specification cleaner and simpler for our present purposes.

Beyond broadly matching the available data, our model makes specific predictions that can be examined with further behavioral studies. A first such prediction continues the point above: if stimuli were presented to the participants such that the first object was removed from direct perception, followed by a delay prior to the presentation of the second object, we would expect a bigger RT difference between the aphantasia and control groups even in small angles. A different prediction relates to the fact that our model suggests that the difference in RT between people with and without aphantasia is due to the comparison stage, rather than the recognition or rotation stages. The current data clumps all of these stages together, but it is possible to design experiments that tease these stages apart. Indeed, Shepard and Cooper, 1982/1986 introduced such studies. We would propose then to run equivalent studies that separate the different stages of mental rotation, and the response of people with and without aphantasia for each stage. We expect to find that much of the RT difference between the groups is due to the comparison stage rather than the rotation stage.

While the bulk of our studies here have focused on the comparison between aphan-

tasia and control groups, we note that our model is also a model of mental rotation in people *without* aphantasia, and is interesting to consider on its own as such. In particular, while our current model is set up to 'render' an object at the end of the rotation, it is interesting to consider whether such a description matches people's reports and behavior, and how rotation can be modified to render objects at intermediate outputs. To explain what we mean by this, consider first that it is possible for a game/physics-engine to rotate a body step-by-step, but either with or without rendering what the body looks like visually at different sub-steps. Rendering at different sub-steps is possible, but slows the process down (we note as an aside that we did not account for the processing time of rendering in the model here, which could introduce an added constant on top of the results). Now, consider the visual imagery associated with mental rotation in people. Consider for example an image with a high level of internal visual detail, such as a zebra or a birthday cake with many colored candles. Next, try to mentally rotate that image by an unconventional angle (e.g. 60 degrees, or 150 degrees). Based on anecdotal examination, it seems even many people without aphantasia might agree with the suggestion that they do not 'see' the internal details of the object being rotated but rather have only the impression of the bulk shape, outline, or axis rotating, and it is only once the rotation is finished that they picture the individual features. Such an impression is in line with our model here. One could then go further and have people try to render the image at sub-steps of the rotation, which we then expect to lead to specific, predictable slowdowns, akin to those of a physics/game-engine tasked in a similar way.

Our model has focused on mental rotation and aphantasia. It accounts for both the similarities and differences between people with and without aphantasia in this task, and suggests novel studies to test this framework further. Because the model was based on the 'broken rendering' proposal, we also see it applying more broadly. The broader suggestion is that people with and without aphantasia create and manipulate objects of the mind in a similar, hierarchical way, with much of the work being done by spatial/physical operations. It is a separate operation to then 'render' what a scene or object looks like from a particular perspective. This operation may not be available to all of us, but also may not be terribly crucial. When we wonder how a director can move around the scenery on the mind's stage without seeing it, we should remember that even for those of us who can see the set, it was we who set it up before the curtains lifted.

References

Bainbridge, W. A., Pounder, Z., Eardley, A. F., & Baker, C. I. (2021). Quantifying aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory. *Cortex*, *135*, 159–172. https://doi.org/10.1016/j.cortex.2020.11.014

Balaban, H., & Ullman, T. D. (2025). Physics versus graphics as an organizing dichotomy in cognition. *Trends in Cognitive Sciences*, *29*(11), 985–996. https://doi.org/10.1016/j.tics.2025.05.003

Balaban, H., & Ullman, T. D. (2026). Emotion may indirectly link rendering and social reasoning. *Trends in Cognitive Sciences*.

Burle, B., Possamaï, C.-A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: An electromyographic and distributional analysis. *Psychological Research*, *66*(4), 324–336. https://doi.org/10.1007/s00426-002-0105-6

Chang, S., Zhang, X., Cao, Y., Pearson, J., & Meng, M. (2025). Imageless imagery in aphantasia revealed by early visual cortex decoding. *Current Biology*, *35*(3), P591–599. https://doi.org/10.1016/j.cub.2024.12.012

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In *Visual information processing* (pp. 75–176). Elsevier.

Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., & Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *eLife*, *3*, e03005. https://doi.org/10.7554/eLife.03005

Frick, A., Möhring, W., & Newcombe, N. S. (2014). Development of mental transformation abilities. *Trends in Cognitive Sciences*, *18*(10), 536–542. https://doi.org/10.1016/j.tics.2014.05.011

Ganis, G., & Kievit, R. A. (2015). A new set of three-dimensional shapes for investigating mental rotation processes: Validation data and stimulus set. *Journal of Open Psychology Data*, *3*(1), e3. https://doi.org/10.5334/jopd.ai

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* John Wiley.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Kay, L., Keogh, R., Andrillon, T., & Pearson, J. (2022). The pupillary light response as a physiological index of aphantasia, sensory and phenomenological imagery strength. *eLife*, *11*, Article e72484. https://doi.org/10.7554/eLife.72484

Kay, L., Keogh, R., & Pearson, J. (2024). Slower but more accurate mental rotation performance in aphantasia linked to differences in cognitive strategies. *Consciousness and Cognition*, *121*, 103694. https://doi.org/10.1016/j.concog.2024.103694

Keogh, R., & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, *105*, 53–60. https://doi.org/10.1016/j.cortex.2017.10.012

Keogh, R., Wicken, M., & Pearson, J. (2021). Visual working memory in aphantasia: Retained accuracy and capacity with a different strategy. *Cortex*, *143*, 237–253. https://doi.org/10.1016/j.cortex.2021.07.012

Khazoum, R., Fernandes, D., Krylov, A., Li, Q., & Deny, S. (2026). A deep learning model of mental rotation informed by interactive VR experiments. *Proceedings of the 43rd International Conference on Machine Learning*, *306*.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The Case for Mental Imagery*. Oxford Psychology Series. https://doi.org/10.1093/acprof:oso/9780195179088.001.0001

Liu, J., Zhan, M., Hajhajate, D., Spagna, A., Dehaene, S., Cohen, L., & Bartolomeo, P. (2025). Visual mental imagery in typical imagers and in aphantasia: A millimeter-

scale 7-t fmri study. *Cortex*, *185*, 113–132. https://doi.org/10.1016/j.cortex.2025.01.013

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide, 2nd ed.* Lawrence Erlbaum Associates Publishers.

Michel, M., Morales, J., Block, N., & Lau, H. (2024). Aphantasia as imagery blindsight. *Trends in Cognitive Sciences*, *29*(1), 8–9. https://doi.org/10.1016/j.tics.2024.11.002

Milton, F., Fulford, J., Dance, C., Gaddum, J., Heuerman-Williamson, B., Jones, K., Knight, K. F., MacKisack, M., Winlove, C., & Zeman, A. (2021). Behavioral and neural signatures of visual imagery vividness extremes: Aphantasia versus hyperphantasia. *Cerebral Cortex Communications*, *2*(2), tgab035. https://doi.org/10.1093/texcom/tgab035

Nanay, B. (2021). Unconscious mental imagery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1817), Article 20190689. https://doi.org/10.1098/rstb.2019.0689

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences*, *21*(6), 425–433. https://doi.org/10.1016/j.tics.2017.03.011

Phillips, I. (2025). Spared spatial imagery solves the puzzle of aphantasia. *Trends in Cognitive Sciences*, *29*(11), 961–962. https://doi.org/10.1016/j.tics.2025.05.004

Pitt, M. A., Kim, w., Navarro, N. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, *113*(1), 57–83. https://doi.org/10.1037/0033-295X.113.1.57

Pounder, Z., Jacob, J., Evans, S., Loveday, C., Eardley, A. F., & Silvanto, J. (2022). Only minimal differences between individuals with congenital aphantasia and those with typical imagery on neuropsychological tasks that involve imagery. *Cortex*, *148*, 180–192. https://doi.org/10.1016/j.cortex.2021.12.010

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

Purkart, R., Delem, M., Ranson, V., Andrey, C., Versace, R., Cavalli, E., & Plancher, G. (2025). Are there unconscious visual images in aphantasia? development of an implicit priming paradigm. *Cognition*, *256*, 106059. https://doi.org/10.1016/j.cognition.2024.106059

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, *25*(2), 157–182. https://doi.org/10.1017/S0140525X02000043

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

Reeder, R. R., Pounder, Z., Figueroa, A., Jüllig, A., & Azañón, E. (2024). An example reference for a single author article. *Cognition*, *251*(1), 105907. https://doi.org/10.1016/j.cognition.2024.105907

Shepard, R. N., & Cooper, L. A. (1986). *Mental images and their transformations*. The MIT Press. (Original work published 1982)

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Tomasino, B., & Gremese, M. (2016). Effects of stimulus type and strategy on mental rotation network: An activation likelihood estimation meta-analysis. *Frontiers in Human Neuroscience*, *9*(693). https://doi.org/10.3389/fnhum.2015.00693

Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017). Mind Games: Game Engines as an Architecture for Intuitive Physics. *Trends in Cognitive Sciences*, *21*(9), 649–665. https://doi.org/10.1016/j.tics.2017.05.012

Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology*, *78*, 148–174. https://doi.org/10.1016/j.cogpsych.2015.02.005

Wicken, M., Keogh, R., & Pearson, J. (2021). The critical role of mental imagery in human emotion: Insights from fear-based imagery and aphantasia. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1946), Article 20210267. https://doi.org/10.1098/rspb.2021.0267

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, Article e49547. https://doi.org/10.7554/eLife.49547

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of cognitive neuroscience*, *20*(1), 1–19. https://doi.org/10.1162/jocn.2008.20013

Zeman, A., Dewar, M., & Sala, S. D. (2015). Lives without imagery – congenital aphantasia. *Cortex*, *73*, 378–380. https://doi.org/10.1016/j.cortex.2015.05.019

Zeman, A. Z., Della Sala, S., Torrens, L. A., Gountouna, V.-E., McGonigle, D. J., & Logie, R. H. (2010). Loss of imagery phenomenology with intact visuo-spatial task performance: A case of 'blind imagination'. *Neuropsychologia*, *48*(1), 145–155. https://doi.org/10.1016/j.neuropsychologia.2009.08.024

Zhao, B., Della Sala, S., Zeman, A., & Gherri, E. (2022). Spatial transformation in mental rotation tasks in aphantasia. *Psychonomic Bulletin & Review*, *29*(6), 2096–2107. https://doi.org/10.3758/s13423-022-02126-9

Appendix A
Further details on formal model specification

Dissimilarity Distributions

For each figure, imprecision level, and route $r \in \{S, I\}$, there is one readout bank. Each readout bank gives the separation of the two trial types and a decision criterion:

$$\begin{aligned}
\Delta\mu_r &= \text{mean}(f_r^{\text{mirror}}) - \text{mean}(f_r^{\text{same}}),\\
\sigma_r &= \sqrt{\tfrac{1}{2}\big(\text{Var}(f_r^{\text{same}}) + \text{Var}(f_r^{\text{mirror}})\big)},\\
d_r &= \Delta\mu_r / \sigma_r,\\
c_r &= \arg\max_c \ \tfrac{1}{2}\big(\Pr(f_r^{\text{same}} < c) + \Pr(f_r^{\text{mirror}} > c)\big).
\end{aligned}$$

d_r is the discriminability of the route, the equivalent of d_a in signal detection theory, when the signal distribution and noise distribution have unequal variances (Green & Swets, 1966; Macmillan & Creelman, 2005). The criterion c_r is the threshold that maximizes accuracy when same and mirror trials are equally likely. It is estimated by sweeping the threshold over the bank, due to the heteroscedasticity of the distributions. The underlying rationale of the readout bank is the assumption that people have a self-model of these quantities for the figures they see. Therefore, these constants reflect the properties of the stimuli and rotation imprecision.

Visual-route unreliability κ

A factor $\kappa \geq 1$ scales the visual route's intrinsic noise relative to its algorithmic readout f_I, i.e., the readout using 160×160 pixels, for each participant. The effective discriminability is d_I/κ_j, so κ_j spreads out the visual readout's distributions. $\kappa = 1$ means the reliability is the same as algorithmic readout. $\tilde{f}_I = f_I + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \sigma_I^2(\kappa_j^2 - 1))$, which gives the degraded readout $\tilde{f}_I$ a standard deviation of $\sigma_I \kappa_j$. The effective discriminabilities are:

$$\tilde{d}_S = d_S, \quad \tilde{d}_I = \frac{d_I}{\kappa_j} \tag{A1}$$

Evidence to decision

For each simulated trial, three quantities are passed sequentially: the readout, the evidence, and the drift. The readout $\tilde{f}_r$ is the dissimilarity measurement. The evidence z_r links the readout value to the drift. The drift v_r is the rate that evidence drives the decision.

Each readout bank has different signal and noise (*mirror* and *same*) distributions. The raw value of the readout in each trial is only meaningful if considering the bank distributions. We define the evidence of a route as its readout relative to the criterion, divided by

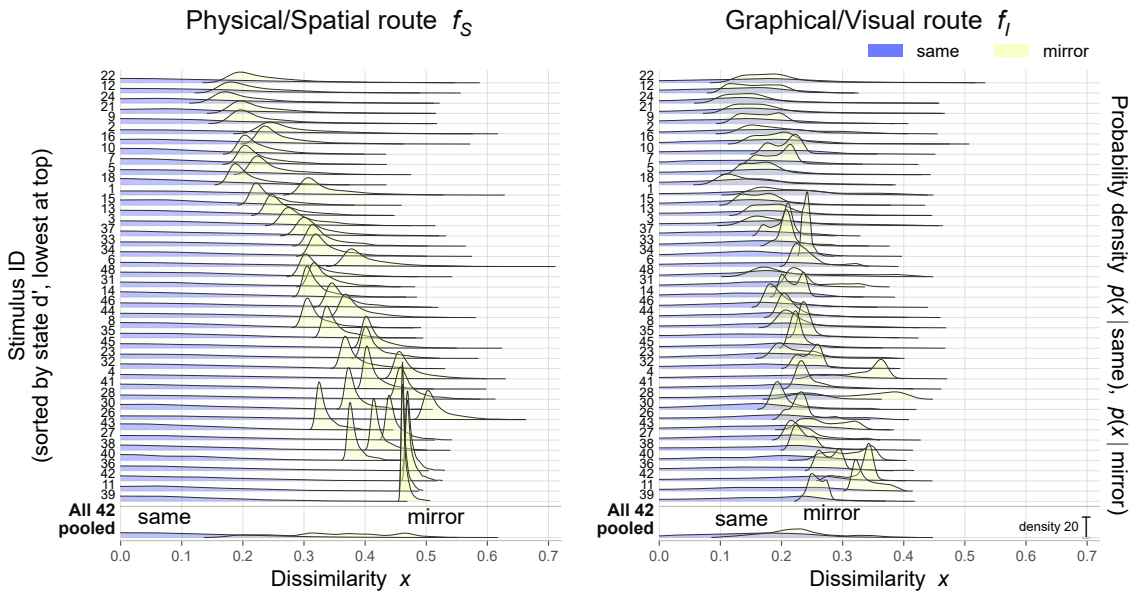


Figure A1

Same versus mirror dissimilarity for every stimulus. *Each row shows the probability density of dissimilarity on same and mirror trials of one stimulus figure,* $s_{jit} = 25°$. ***Left.*** *The physical/spatial route. The median* d' *across the stimuli is 2.90.* ***Right.*** *The graphical/visual route. The median* d' *across the stimuli is 1.06.*

the separation (mean difference) between the two trial types:

$$z_r = \frac{\tilde{f}_r - c_r}{\Delta\mu_r}. \tag{A2}$$

The sign indicates which response the evidence favors, positive for *mirror* and negative for *same*. The magnitude shows the strength of the evidence. The evidence averages to $+\frac{1}{2}$ on mirror trials and $-\frac{1}{2}$ on same trials. The standard deviation of each route is $1/\tilde{d}_r$.

We set the drift to be derived by evidence in the following form:

$$v_r = \frac{z_r}{\tilde{d}_r}. \tag{A3}$$

This definition connotes the core assumption of the model: a speed-accuracy tradeoff between comparison routes. A comparison route that discriminates the trial types better contributes to a smaller drift on average, but varies less from trial to trial. The mean drift is $\frac{1}{2\tilde{d}}$ and the trial-level standard deviation is $\frac{1}{\tilde{d}^2}$. The signal-to-noise ratio is $\frac{\tilde{d}}{2}$ (see Appendix A). A more reliable comparison route thus drives slower, but more accurate decisions; a less reliable one drives faster, but more erroneous decisions. This implies the visual comparison route is faster but less reliable, and the spatial, structural comparison route is slower

and more reliable. This is how the model expresses the trade-off. Appendix F shows that the resulting patterns depend on this division rather than the exact form formulated.

We model people's ability of spatial comparison as a gain g_j on the signal in the spatial evidence. The ability is applied in the same direction as the evidence:

$$v_S = \frac{z_S + (g_j - 1)\,\mathrm{sign}(z_S)/2}{\tilde{d}_S}. \tag{A4}$$

When $g_j = 1$, there is no gain in ability in the spatial route. A large gain value drives the decision faster to the direction of the evidence. As the direction is usually correct, a more capable person would be both faster and more accurate. The ability gain follows the direction of the evidence, not of the correct answer, so that the generative model does not act as an oracle. This also adds occasional error, and Appendix A shows that the gain can therefore only raise accuracy up to that of the spatial route evidence itself, $\Phi(\frac{\tilde{d}_S}{2})$, the probability that the evidence has the correct sign.

We combine the drifts of the two routes by the following:

$$v = v_S + \alpha_j v_I \tag{A5}$$

Note that v_S is the same for both groups. Only people without aphantasia have the additional source of information from the visual comparison route. This route makes the responses faster but less accurate. For people with aphantasia, $\alpha_j = 0$, $v = v_S$. Adding different drifts into a single diffusion is standard practice (Drugowitsch et al., 2014; Ulrich et al., 2015).

The decision is a drift-diffusion process (Ratcliff & McKoon, 2008). A decision variable X starts in the middle of the two bounds, 0 and a. For each small time step Δt, the variable X accumulates:

$$X_0 = \frac{a}{2}, \quad X_{n+1} = X_n + v\Delta t + s\sqrt{\Delta t}\,\eta_n, \quad \eta_n \sim \mathcal{N}(0,1). \tag{A6}$$

X moves by the drift v times the step length, plus the diffusion of independent Gaussian noise. The drift, v, varies across trials, as each trial draws a new pair of readouts from the bank. The response is *mirror* when reaching a and *same* for reaching 0. The maximum decision time is 8 seconds. If this is exceeded, the process responds with the bound that X is closer to. The decision time T_{dec} is then $N\Delta t$.

There is also a non-decision time $\tau_{0,j}$, for the part of response not accumulating evidence:

$$RT = \tau_{0,j} + T_{dec}, \quad \tau_{0,j} = t_0 + t_{0,sd}\,\zeta_j, \quad \zeta_j \sim \mathcal{N}(0,1). \tag{A7}$$

Each participant j has their own non-decision time $\tau_{0,j}$, with population mean $t_0 = 520$ ms and standard deviation $t_{0,sd} = 65$ ms. Other parameters, bound separation $a = 2.4$, noise $s = 1$, Δt = 3 ms, are equal for all participants.

Individual differences

Each simulated participant j draws from three characteristic distributions that contribute to group-level correlation between RT and accuracy:

$$
\begin{aligned}
g_j &\sim \mathscr{TN}(\bar{g},\ g_{sd};\ [0.05,\infty)), && \text{both groups,}\\
\alpha_j &\sim \begin{cases} \mathscr{TN}(\bar{\alpha},\ \alpha_{sd};\ [0,\infty)), & \text{control,}\\ 0, & \text{aphantasia,} \end{cases}\\
\kappa_j &\sim \mathscr{TN}(\bar{\kappa},\ \kappa_{sd};\ [1,\infty)), && \text{controls.}
\end{aligned}
$$

Each variable has a characteristic effect on the within-group RT–accuracy correlation:

- Ability g_j is a difference in how well one can exploit the information from evidence. More capable participants are faster and more accurate, which produces a negative RT-accuracy correlation.
- Visual reliance α_j and visual route unreliability κ_j both trade accuracy for speed. The higher reliance on the visual route or the less reliable the visual route, the faster, the less accurate one is. This produces a positive RT-accuracy correlation.

Participants in the aphantasia group differ only in spatial route ability. Participants in the control group differ in all three variables, while visual reliance and visual route unreliability dominate the variation. These differences among individuals are what produce the opposite signs of the two correlations (Table D3).

The shared rotation stage

The model so far only concerns the comparison stage, which creates the group differences. The shared rotation stage, which we hypothesize to involve only 3D spatial representation, also contributes to the overall performance and reaction time pattern while exerting no group difference. Therefore, for the rotation stage, we describe rather than model the reaction time and accuracy pattern. We use a few shared constants that are fitted to Kay et al. (2024)'s group level data. This helps to produce the canonical RT-angle linear relationship, and also the increased error as angle disparity rises (Shepard & Cooper, 1982/ 1986; Shepard & Metzler, 1971). We add them to each participant's mean correct RT, $\overline{RT}_j(\theta)$, and accuracy at each angle:

$$
\begin{aligned}
\widetilde{RT}_j(\theta) &= \overline{RT}_j(\theta) + \phi_j \rho\, \theta + \psi_j \rho_0,\\
\widetilde{acc}_j(\theta) &= 1 - k(\theta)[1 - acc_j(\theta)], \quad k(\theta) = k_0 + k_1 \theta
\end{aligned}
\tag{A8}
$$

where:

- ρ is the mean rotation rate (ms/degree), and ϕ_j is the participant's rotation rate relative to the mean,
- ρ_0 is the mean time spent not on the comparison stage, and ψ_j is the participant's time on it relative to the mean,
- $k(\theta)$ is an error correction factor that shrinks the error and scales linearly with angle,

The factors ϕ_j and ψ_j are drawn independently, each from a normal distribution truncated at 0 and rescaled to a mean of 1. σ_ϕ and σ_ψ denote their standard deviations.

Speed

People differ in how fast they rotate (Provost & Heathcote, 2015; Shepard & Cooper, 1982/1986). Factor ϕ_j sets the rotation rate for participant j, and ψ_j sets the time they spend outside the comparison stage. These two factors are drawn independently, which is supported by Kay et al. (2024)'s data, where a participant's RT at 0° is not correlated with the slope of RT over angles ($r = 0.03$ in aphantasia and $r < 0.01$ in controls), although Provost and Heathcote (2015) found $r = 0.82$.

Errors

The comparison stage reproduces the reaction time-accuracy correlations, but two error patterns are absent: First, the participants from Kay et al. (2024)'s data have higher accuracy than the comparison stage model implies. Second, the errors do not scale with angle disparities in the comparison stage as the participants do. An error factor, $k(\theta)$, addresses both. The factor shrinks the error when smaller than 1. Though this is a descriptive choice, it could be justified by the literature (Burle et al., 2002).

We set the six constants from Kay et al. (2024)'s group-level data. The fitted values are $\rho = 8.5$ ms/deg, $\rho_0 = 815$ ms, $k_0 = 0.33$, $k_1 = 0.0042/deg$, $\sigma_\phi = 0.61$, and $\sigma_\psi = 0.46$. The error factor increases from 0.33 at 0° to 0.96 at 150°; error is most corrected when the angle disparity is small.

As stated above, the rotation stage is shared for everyone, so it does not drive the group difference. However, the rotation stage attenuates the correlation between RT and accuracy, because the speed factor ψ_j injects differences in speed that are independent of performance. The observed correlation is therefore smaller than the comparison stage alone. The error factor $k(\theta)$ only slightly changes the correlations as it weights 100° and 150° a little differently. This change is about 0.004 on average and at most 0.02, with no sign change in 1000 simulated experiments.

Although it has been suggested that there is trial-to-trial within-participant variability (Provost & Heathcote, 2015), this is inconsequential to our between-participant level behavioral pattern. With the variability magnitude previously found, it would only increase the between-participant standard deviation of mean RT by less than 2%.

Implied moments and interpretation

On mirror trials, let $\beta_j = (g_j - 1)/2$, $\zeta = \tilde{d}_S/2$, $p = \Phi(\zeta)$ (the probability the evidence lands on the correct side of the criterion).

Visual/graphical route.

$$\mathbb{E}[v_I] = \frac{1}{2\tilde{d}_I}, \qquad \mathrm{SD}(v_I) = \frac{1}{\tilde{d}_I^2}, \qquad \mathrm{SNR}_I = \frac{\tilde{d}_I}{2}. \tag{A9}$$

Spatial/physical route, with ability gain.

$$\mathbb{E}[v_S] = \frac{\frac{1}{2} + \beta_j(2p-1)}{\tilde{d}_S},$$
$$\mathrm{SD}(v_S) = \frac{1}{\tilde{d}_S}\sqrt{\frac{1}{\tilde{d}_S^2} + 4\beta_j^2\, p(1-p) + \frac{4\beta_j}{\tilde{d}_S}\,\varphi(\zeta)},$$

$$\mathrm{SNR}_S = \frac{\mathbb{E}[v_S]}{\mathrm{SD}(v_S)} = \frac{\frac{1}{2} + \beta_j(2p-1)}{\sqrt{\frac{1}{\tilde{d}_S^2} + 4\beta_j^2\, p(1-p) + \frac{4\beta_j}{\tilde{d}_S}\,\varphi(\zeta)}}. \tag{A10}$$

At $g_j = 1$ ($\beta_j = 0$) this reduces to $\mathbb{E}[v_S] = 1/(2\tilde{d}_S)$, $\mathrm{SD}(v_S) = 1/\tilde{d}_S^2$, the same form as the visual route. For $g_j > 1$ the boost raises the mean drift. Because the sign flips across the criterion, raising spatial ability increases the across-trial SD; consequently, as g_j increases, SNR_S first rises and then falls towards $(2p-1)/\left(2\sqrt{p(1-p)}\right)$, set by the route's discriminability. (This limit is the cost of using the evidence sign rather than the ground truth. A boost that used the ground truth would raise SNR indefinitely.) Note that with this spatial ability gain, the upper bound for accuracy is determined by what the spatial evidence gives, towards $p = \Phi(\tilde{d}_S/2)$.

In the control group, the total drift $v = v_S + \alpha_j v_I$ has mean $\mathbb{E}[v_S] + \alpha_j \mathbb{E}[v_I]$ and variance $Var(v_S) + \alpha_j^2 Var(v_I) + 2\alpha_j Cov(v_S, v_I)$. The covariance is positive as both routes read the same rotated and aligned objects.

Appendix B
Discriminability difference in the two comparisons

At an image size of 160×160 pixels, the discriminability of 3D spatial/physical route comparison is higher than that of the 2D visual/graphical route, at $s_{jit} = 25°$. This holds for all 42 figures (Figure B1). The median discriminability of spatial/physical route is $d_S = 2.90$ against $d_I = 1.06$ in visual/graphical route, which means a median ratio of 2.8. This ratio stays about the same at all imprecision levels (2.8–2.9 for $s_{jit} = 20°$ to $35°$). The lower reliability in the visual/graphical route is a property of the stimuli and comparison. For the representative setting, the visual/graphical route is less reliable still ($\bar{\kappa} = 1.29$, $\kappa_{sd} = 3.47$).

Appendix C

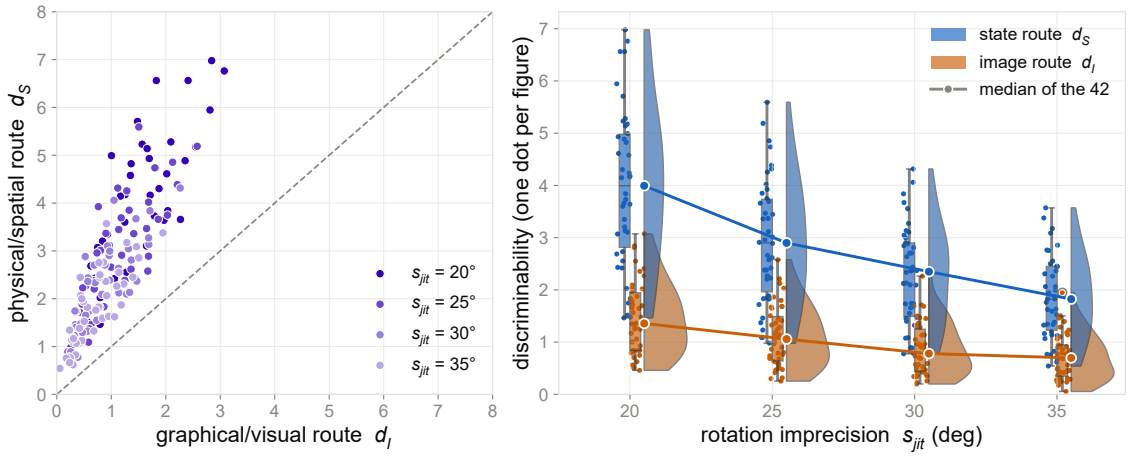


Figure B1

Discriminability ratio between two routes. Left*. Physical/spatial versus graphical/visual discriminability at $s_{jit} = 20°, 25°, 30°, 35°$. All figures lie above the diagonal, meaning physical/spatial route always discriminates better than the graphical/visual route.* ***Right****. The discriminability at each imprecision level. Each dot represents one figure. The median d' ratio stays between 2.8 and 2.9 across the imprecision levels.*

Imprecision of alignment as the source of uncertainty

We argue that rotation errors of this size ($s_{jit} = 25°$ and 30°) are plausible, and that it is plausible that the imprecision does not depend on angle disparity (Table C1). In a study where participants rotate one of the two Shepard-Metzler figures about a vertical axis in virtual reality, the median angle misalignment when they stopped the rotation was 15°, with 92% within ±45°, regardless of the initial presented angle disparity (Khazoum et al., 2026). At our main reported imprecision level, $s_{jit} = 25°$, the misalignment angle has a nearly identical distribution, with a median of 17° and 93% within ±45°.

Appendix D
Parameter space partitioning

Four sign criteria

Setting aside the pattern both groups share, slower and less accurate responses at larger angles, which belongs to the shared rotation stage, the specific differences between the two groups are as follows:

1. People with aphantasia are slower;
2. they are more accurate;
3. their RT and accuracy correlate negatively;
4. controls' RT and accuracy correlate positively.

	Khazoum et al.	Model	
	(measured)	$s_{\text{jit}} = 25°$	$s_{\text{jit}} = 30°$
Within ±45°	92%	93%	87%
Mean absolute	21°	20°	24°
Median absolute	15°	17°	20°
90th percentile, absolute	40°	41°	49°

Table C1

Misalignment between two objects at the end of the rotation stage. *Khazoum et al. (2026)'s data are from the 412 correct same trials with at least one rotation action, at* 60°, 120° *and* 180° *angle disparities. This is computed from their published trial data. For our model, the rotation error,* $\mathcal{N}(0, s_{jit}^2)$*, is about the vertical axis.*

These are the four main criteria for our model in the comparison stage. They are qualitative claims, so we care only about their signs. The shared rotation stage cannot produce or reverse these characteristics; it can only attenuate them. Pearson's and Spearman's correlations give the same sign in more than 99% of the simulated experiments from the model. Thus, the partition and selection use Pearson's for both groups.

Procedure

We explore the qualitative patterns produced by the model in the whole region of plausible parameter values by partitioning parameter space (Pitt et al., 2006), to see whether the characteristics are produced by the structural commitment of the model. We sampled 1024 parameter settings from a scrambled Sobol sequence over the parameter box in Table D1. We simulate 500 participants for both groups in each setting. The parameter range is as follows:

- each individual difference has a standard deviation from 0 to 5;
- state ability ranges from 0.5, below neutral, to 6 ($\bar{g} = 1$ is neutral, below 1 weakens the signal evidence);
- mean reliance on the visual comparison ($\bar{\alpha}$) from 0 to 3 times the spatial comparison weight;
- mean visual comparison unreliability ($\bar{\kappa}$) from 1 (none) to 10 times the discriminability of the algorithmic outputs.

This analysis gives us the percentage of the parameter box in which all four signs hold, the frequency of all $2^4 = 16$ combinations of sign patterns, and the percentage of the parameter box in which they still hold when individual differences are removed (g_{sd}, α_{sd}, or κ_{sd}), which are crucial for the RT-accuracy correlation pattern.

Parameter	Meaning	Range or status	Representative
$\Delta\mu_r, \sigma_r, c_r, d_r$	route calibration	from the stimuli	per figure
s_{jit}	rotation imprecision	$\{20, 25, 30, 35\}^\circ$	25°
$\bar{g},\ g_{sd}$	state ability	$[0.5, 6],\ [0, 5]$	4.63, 3.51
$\bar{\alpha},\ \alpha_{sd}$	imagery reliance (controls)	$[0, 3],\ [0, 5]$	0.30, 0.74
$\bar{\kappa},\ \kappa_{sd}$	image unreliability (controls)	$[1, 10],\ [0, 5]$	1.29, 3.47
$a, s, \Delta t$	diffusion	fixed	2.4, 1, 3 ms
$t_0, t_{0,sd}$	non-decision time	fixed	520, 65 ms
ρ, ρ_0	rotation and other time	Kay et al.'s RT means	8.5 ms/deg, 815 ms
k_0, k_1	error factor	Kay et al.'s accuracy	0.33, 0.0042/deg
σ_ϕ, σ_ψ	spread of ρ, of ρ_0	Kay et al.'s RT spread	0.61, 0.46

Table D1

Model parameters, their ranges, and representative values. *RT = response time; sd = standard deviation. Values for ρ, ρ_0, k_0, k_1, σ_ϕ, and σ_ψ were derived from Kay et al. (2024).*

Results

The four sign criteria hold almost everywhere in the parameter space box (Table D2). They hold for 96.9% (95% CI[95.6, 97.8]) of the parameter settings tested at $s_{jit} = 25^\circ$ (992 out of 1024). They hold for 96.3% (95% CI[94.9, 97.3]) at $s_{jit} = 30^\circ$.

The parameter space partition also shows that the model always predicts slower responses for aphantasia, and only 0.2% of parameter settings do not produce higher accuracy in aphantasia, at $s_{jit} = 25^\circ$. The two exceptions have the lowest mean spatial ability in the tested range ($\bar{g} \simeq 0.5$, which makes the spatial evidence weaker than visual evidence), and the reversed accuracy differences are only 0.014 and 0.002. The other parameter settings that fail the sign criteria are those that fail to show the RT-accuracy correlation in either group. At $s_{jit} = 25^\circ$, the 20 cases that failed the correlation in the aphantasia group all have little variation in spatial state ability, and the 10 cases without a positive correlation in the control group all have a very unreliable visual route ($\bar{\kappa} > 6$).

Individual differences determine the correlation of RT-accuracy within the groups. We see this when ablating each individual-difference variable, that is, setting it to the population average for all individuals (Table D3). When ablating spatial ability differences, the negative correlation in the aphantasia group vanishes, with the median across parameter settings going from -0.58 to -0.02. For the control group, either kind of visual route difference contributes to their positive correlation. When both are removed, the median correlation drops from 0.36 to 0.01. Ablating either one still leaves the positive correlation intact: the median correlation is 0.31 when removing the visual route reliance differences, and 0.18 when removing the visual route unreliability.

Appendix E

Aph. slower	Aph. more accurate	Aph. $r < 0$	Ctl. $r > 0$	Share at 25° (30°)
yes	yes	yes	yes	96.9% (96.3%)
yes	yes	no	yes	2.0% (2.4%)
yes	yes	yes	no	1.0% (1.2%)
yes	no	yes	yes	0.2% (0%)
yes	yes	no	no	0% (0.1%)
any other pattern (11 of 16)				0% (0%)

Table D2

Parameter space partition: percentage of 1024 ***parameter settings showing each pattern of sign criteria.*** *Aph. = aphantasia; Ctl. = controls; r = within-group RT–accuracy correlation.*

Individual differences included	Aphantasia r	Controls' r
all (g, κ, α)	$-.58$ $[-.72, -.11]$	$+.36$ $[+.13, +.53]$
without visual route reliance variation ($\alpha_{sd} = 0$)	—	$+.31$ $[+.01, +.51]$
without visual route unreliability variation ($\kappa_{sd} = 0$)	—	$+.18$ $[+.04, +.33]$
without either visual route variation ($\alpha_{sd} = 0$ & $\kappa_{sd} = 0$)	—	$+.01$ $[-.05, +.07]$
without spatial ability variation ($g_{sd} = 0$)	$-.02$ $[-.08, +.04]$	$+.39$ $[+.14, +.59]$

Table D3

Within-group correlations come from individual differences across parameter space. *This shows the comparison stage's RT-accuracy correlation (Pearson's r,* 100°–150°*) when ablating one source of individual differences. It is tested across all* 1024 *settings at* $s_{jit} = 25°$*, with* 500 *simulated participants per group. An ablated characteristic is set to its population average. Medians across settings are shown, with the 10th and 90th percentiles. Visual route reliance and unreliability exist only in controls.*

Representative parameter setting selection

This section describes how we chose the parameter setting used for the detailed comparison with the behavioral results of Kay et al. (2024).

Magnitude targets

We set four magnitude targets from Kay et al. (2024)'s effect sizes, measured over 100°–150°, to achieve better quantitative mapping:

1. the RT difference between the two groups, 513 ms from Kay et al. (2024);

2. the accuracy difference between the two groups, 0.107;

3–4. b, how many milliseconds a group's RT changes per one standard deviation difference in accuracy ($b = r \times SD(RT) = \frac{Cov(RT,accuracy)}{SD(accuracy)}$). This quantifies the portion of RT spread that scales with accuracy. While r is attenuated when the shared rotation stage is added on top, b stays the same: -198 ms for aphantasia and 307 ms for controls.

A setting reaches a target if it lies within two standard errors of Kay et al. (2024)'s estimate (for accuracy difference, ± 0.05, about 2.5 standard errors).

Selection

The selection is done in two stages, separately at each imprecision level.

In stage 1, we tested 8192 parameter settings from a scrambled Sobol sequence (with a different random seed from the parameter space partitioning above), with 30 simulated participants per group, at each imprecision level s_{jit}. Among the settings for which the four sign criteria hold, we keep the 50 parameter settings per imprecision level that best match the magnitude targets. For each of these 50 settings, we run 100 simulated experiments at the sample sizes of Kay et al. (2024), and one with 1500 participants per group. The 100 simulated experiments give the probability that a single experiment of the sample size shows all four sign criteria. The $n = 1500$ check verifies that all four sign criteria hold.

In stage 2, we choose one of the 50 with a scoring rule that weighs two items: (1) with weight 0.6, the probability that a single experiment of Kay et al. (2024)'s sample size shows both within-group correlations as significant, which we estimate for each parameter setting from its b divided by $SD(RT)$ in Kay et al. (2024)'s data; and (2) with weight 0.4, how close the comparison stage gets to the magnitude targets. We report the parameter setting at $s_{jit} = 25°$, while $s_{jit} = 30°$ gives similar results.

The typical simulated experiment

One of the 1000 simulated experiments at the chosen representative setting is plotted in Figure 2. We kept the experiments whose means and group differences lie within the central 80% at every angle, and whose RT and accuracy differences over 100° and 150° lie within the central 50%. Among those, the one shown has the within-group correlations closest to the medians across the 1000 simulated experiments.

Appendix F
Robustness

The axis of rotation imprecision

Our current model uses vertical-axis perturbations. If the perturbation is about a random axis, at matched d_S, the results are essentially the same. At matched difficulty (the median d′ of the vertical axis at $s_{jit} = 25°$), the four sign criteria hold in 97.0% of the parameter space with a random axis, against 96.9% with the vertical axis; at $s_{jit} = 30°$,

the figures are 96.1% and 96.3%. The correlation between the two sensitivity matrices is $r = 0.998$.

The drift form

The critical assumption of the model is the mapping from evidence to drift (Equation A3), that the more cleanly a route separates trial types, the smaller the drift. To check how much the characteristics depend on this assumption, we replaced it with a family of mappings:

$$v_r = \frac{z_r}{d_{ref}} \left(\frac{d_{ref}}{\tilde{d}_r}\right)^k. \tag{F1}$$

d_{ref} is a constant that is the same for every route, figure, and participant. The spatial route's evidence includes the same ability gain, $z_S + (g_j - 1)\,\mathrm{sign}(z_S)/2$, as in Equation A4. The exponent k determines how drift depends on discriminability. When $k > 1$, the dependence is steeper: at $k = 2$, drift scales with $1/\tilde{d}_r^2$. At $k = 1$, the mapping is exactly $v_r = z_r/\tilde{d}_r$. At $k = 0$, drift does not depend on discriminability at all. At $k = -1$, drift is proportional to discriminability. This is the usual case for drift-diffusion models (Ratcliff & McKoon, 2008). At $k = -2$, drift is proportional to the log-likelihood ratio of the route. For normally distributed readouts with equal variances, it would be $\tilde{d}_r^2 z_r$, and the combination would be optimal if the routes were independent (Drugowitsch et al., 2014).

k changes how drift differs between routes (also figures and participants), therefore how much the two routes contribute relative to each other. A figure's image route contributes $\tilde{d}_S/\tilde{d}_I$ times as much drift as its spatial route at $k = 1$, the same amount at $k = 0$, and $(\tilde{d}_I/d_S)^2$ times as much at $k = -2$. We fix the constant $d_{ref} = 2.90$, the median discriminability of the spatial route at $s_{jit} = 25°$. This choice is arbitrary, but it should not matter, as it is the same for all forms (the constant effectively acts as the caution level in the drift-diffusion model).

With the setting described in the notes of Table F1, we found that the qualitative pattern holds as long as drift does not grow with discriminability. For $k \geq 0$, all four sign criteria were satisfied in at least 90% of the parameter settings. The pattern fails at $k = -1$ and $k = -2$, satisfying only 3.9% and 0%, respectively.

The model's pattern, therefore, does not depend on the exact form of the mapping. It only requires that the less reliable route does not lose its influence. The inverse mapping is one way of doing so.

		Share with each sign (%)				Median			
k	All four (%)	Slower	More acc.	$r_{\text{aph}} < 0$	$r_{\text{ctl}} > 0$	RT (ms)	Acc.	r_{aph}	r_{ctl}
-2.0	0.0 [0.0, 1.5]	58.2	45.3	100.0	0.4	4	$-.001$	$-.56$	$-.56$
-1.5	0.0 [0.0, 1.5]	77.7	41.4	98.8	0.8	12	$-.002$	$-.57$	$-.55$
-1.0	3.9 [2.1, 7.0]	97.7	71.5	98.4	6.2	61	.008	$-.56$	$-.50$
-0.5	51.6 [45.5, 57.6]	100.0	94.9	97.7	54.7	284	.068	$-.58$	.03
0.0	90.2 [86.0, 93.3]	100.0	99.6	96.1	94.5	615	.146	$-.59$	.37
0.5	95.3 [92.0, 97.3]	100.0	99.6	97.3	98.4	852	.198	$-.59$	.40
1.0	98.8 [96.6, 99.6]	100.0	100.0	99.6	99.2	964	.213	$-.59$	.36
1.5	95.7 [92.5, 97.6]	100.0	100.0	98.8	96.5	990	.222	$-.57$	.30
2.0	94.5 [91.0, 96.7]	100.0	100.0	100.0	94.5	988	.222	$-.54$	.26

Table F1

The pattern under different mappings from evidence to drift (Equation F1). *256 parameter settings at $s_{jit} = 25°$, 500 simulated participants per group. Left: percentage of settings showing all four signs (95% interval in brackets) and each sign. Right: medians across settings of the group differences (aphantasia minus controls) and of the within-group correlations between RT and accuracy (Pearson). All values are from the comparison stage at 100° and 150°, so the sizes are not comparable with the data. The model is $k = 1$; at $k = -2$ each route is weighted by its reliability.*