Consciousness and Cognition 133 (2025) 103907

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/yccog

Full Length Article

Mental rotation is a weak measure of people's propensity to visualise

Derek H. Arnold[a,*], Loren N. Bouyer[a], Blake W. Saurels[a], Elizabeth Pellicano[b], D. Samuel Schwarzkopf[c,d]

[a] Perception Lab, School of Psychology, The University of Queensland, Australia
[b] Clinical, Educational and Health Psychology, University College London, United Kingdom
[c] Experimental Psychology, University College London, United Kingdom
[d] School of Optometry & Vision Science, The University of Auckland, New Zealand

A R T I C L E I N F O

Keywords:
Visual Imagery
Mental rotation
Aphantasia

A B S T R A C T

In mental rotation (MR) tasks people can be asked to decide if pairs of objects depicted from different viewpoints are the same, or different. A common response strategy is to visualise one of the two objects rotating, until it is visualised from the same viewpoint as the other object. However, some people, Congenital Aphants, assert that they cannot visualise, and yet they perform similarly on MR tasks. This could mean that Congenital Aphants are mistaken about their inability to visualise. Alternatively, we reasoned that MR tasks might be an unreliable metric of people's propensity to rely on visualisation in MR tasks. In a sample of the general population, we had people report on their response strategies on a trial-by-trial basis. Neither people's overall propensity to visualise nor their propensity to visualise on different viewpoint trials was related to viewpoint-contingent changes in MR task performance. There was only a weak association between viewpoint-contingent changes in MR task performance and viewpoint-contingent *changes* in the proportion of visualising trials. Overall, our data suggest that MR task performance is a weak measure of people's propensity to visualise.

1. Introduction

In a mental rotation task people can be asked to quickly decide if two images depict the same or different objects. Crucially, the objects are shown from the same or from different viewpoints (Shepard & Metzler, 1971; Ganis & Kievit, 2015; see Fig. 1). When the objects are depicted from different viewpoints, many people report using a strategy where they visualise one of the two objects rotating in their mind's eye, until it is seen from the same perspective as the other image – thereby helping them to decide if the two images depict the same or different objects. We refer to this response strategy as a Mental Rotation (MR). Viewpoint-contingent changes in performance, in terms of the speed and accuracy of decisions, are popularly regarded as a reliable metric of people's propensity to visualise objects rotating during MR tasks (e.g. Ganis & Kievit, 2015; Just & Carpenter, 1985; Pounder et al., 2022; Shepard & Metzler, 1971).

Some findings, however, suggest that MR tasks may not provide a reliable metric of people's propensity to visualise during MR tasks. Prominent recent examples relate to an imagery deficit. Aphantasia is a term coined to refer to an inability to have voluntary

* Corresponding author.
E-mail address: d.arnold@psy.uq.edu.au (D.H. Arnold).

https://doi.org/10.1016/j.concog.2025.103907
Received 29 October 2024; Received in revised form 23 April 2025; Accepted 14 July 2025
Available online 17 July 2025
1053-8100/

imagined visual sensations – or to visualise (Zeman et al., 2010; Zeman et al., 2015). A number of studies have found that some Aphants can perform similarly to non-aphants on tasks that are commonly believed to measure the propensity of people to visualise (e.g. Jacobs et al., 2018; Pounder et al., 2022; Zhao et al., 2022), including MR tasks (e.g. Jacobs et al., 2018; Marks, 1999; Pounder et al., 2022; also see Logie et al., 2011). For instance, a recent study compared the performance of a group of Aphants on a MR task to a comparison group who could visualise. They found that Aphants were overall slower, but no less accurate in terms of their task performance (Kay et al., 2024). Participants were also quizzed about their response strategies, and Aphants more often reported using non-visualising analytic strategies. From this, the authors argued that their data suggested that similar levels of MR performance could be achieved using different response strategies.

The similar levels of performance achieved by some Aphants on MR tasks (Jacobs et al., 2018; Kay et al., 2024; Marks, 1999; Pounder et al., 2022) has encouraged theorists to question the nature of aphantasia – suggesting that Aphants might be able to visualise, but lack conscious awareness of this capacity (Nanay, 2021; Siena & Simons, 2024; Pounder et al., 2022). We think another possibility should be considered. Perhaps MR tasks are an unreliable metric of people's propensity to visualise during MR tasks, and therefore of people's propensity to visualise generally. If this were true, findings that Aphants and non-aphants can perform similarly on MR tasks would not provide strong evidence that both groups have relied on visualisations to perform the task (also see Kay et al., 2024; Khooshabeh et al., 2013; Marks, 1999).

Evidence suggesting that MR tasks might not be a reliable metric of people's propensity to visualise has also been found in studies of the general population. In this context, when people have been quizzed about how they perform MR tasks, a number of different strategies have been identified, including analytic response strategies that do not involve visualisations (Bethell-Fox & Shepard, 1988; Hegarty, 2018; Khooshabeh et al., 2013). This could involve a cross comparison of the two object images, until the participant finds a feature (often a directional turning point) that is inconsistent across the two images (which signals that the two objects are different). Evidence suggests that, in some circumstances, a non-visualising response strategy can result in *superior* performance on a MR task (Hegarty, 2010). If people who can visualise often use non-visualising response strategies when doing MR tasks, then viewpoint-contingent changes in MR task performance may not generally provide a reliable metric of people's propensity to visualise.

Here we address the following question. Are MR tasks generally a reliable metric of people's propensity to visualise during MR tasks, and therefore of people's capacity and propensity to visualise generally? While our question was motivated by Aphantasia studies (e.g. Jacobs et al., 2018; Kay et al., 2024; Marks, 1999; Nanay, 2021; Pounder et al., 2022; Siena & Simons, 2024), we are interested in how MR tasks are generally mediated, so we addressed our question by taking a sample of the general population. Crucially, in addition to performing a MR task, we asked people to report on their response strategies on a trial-by-trial basis. To preface our results, we found that viewpoint-contingent changes in MR task performance are *not* predicted by the proportions of trials on which people report using visualisations as a response strategy. This is true when the propensity to visualise is estimated by averaging across all trials, and when the propensity to visualise is estimated by averaging across different viewpoint trials. We could only predict viewpoint-contingent *changes* in the accuracy (not the speed) of MR decisions if we considered *changes* in response strategies, on same as opposed to different viewpoint trials.

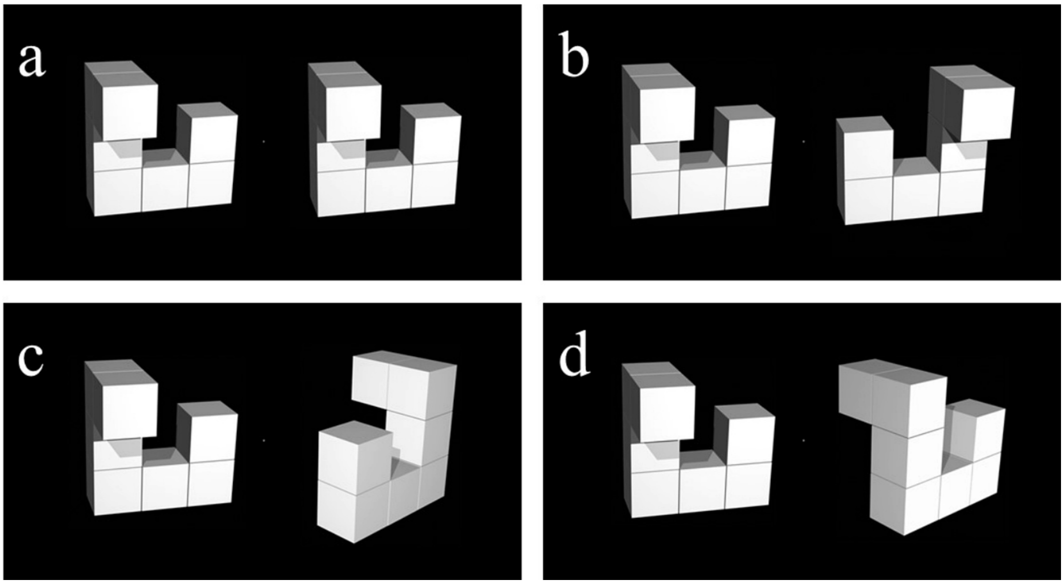

Fig. 1. a) Images depicting the same object viewed from the same viewpoint. b) Images depicting mirror-reversed objects, viewed from the same viewpoint. c) Images depicting the same object, viewed from viewpoints rotated 150°. d) Images depicting mirror-reversed objects, viewed from viewpoints rotated 150°. All images are taken from a set of images depicting three-dimensional shapes for investigating mental rotation – published by Ganis & Kievit, 2015.

2. Materials and methods

All data relating to this study are publicly available via UQ eSpace.

2.1. Ethics

Ethical approval was obtained from the University of Queensland's (UQ) Ethics Committee. The experiment was performed in accordance with UQ guidelines and regulations for research involving human participants. Each participant provided informed consent to participate, and they were informed they could withdraw from the study at any time without prejudice or penalty.

2.2. Participants

A total of 231 people commenced, but 40 failed to complete the study. We excluded a further 96 people on the following basis.

The task response sequence was 2-part, with an additional response required when people reported that they had ***not*** visualised a rotation. This meant that there was an incentive to report visualisations to finish the study quickly. We therefore excluded *all* participants who had either failed to complete the study, or who had reported visualising an object rotating when test images had depicted the same object seen from the same viewpoint.

The final sample size included 95 participants (M age = 20 years, SD = 4.2, range = 17 – 38). Of these participants, 67 identified as women, 27 as men, with one 'other'. All participants were undergraduate students at The University of Queensland, who took part in the study for course credit.

2.3. Stimuli

This study was conducted online, via Qualtrics (Qualtrics, Provo, UT). After providing informed consent to take part, participants were first asked to respond to some demographic questions (Age, Gender identity). After, they completed Mental Rotation trials, followed by the Vividness of Visual Imagery Questionnaire v2 (VVIQ-2, Marks, 1995).

The VVIQ-2 (Marks, 1995) provides a subjective assessment of the vividness of a participant's imagined visual experiences to cued scenarios (e.g. Think of the rising sun. Consider carefully the picture that comes before your mind's eye. The sun is rising above the horizon into a hazy sky). VVIQ-2 scores range from a minimum of 32 (which indicates that the participant has reported having "No image at all, you only "know" that you are thinking of the object" in response to each of 32 scenarios) up to a maximum of 160 (which indicates that the participant reported their imagery was "Perfectly clear and as vivid as normal vision" in response to each scenario).

Images for the mental rotation task were selected from an image set depicting three-dimensional shapes (Ganis & Kievit, 2015). There were four conditions, with pairs of images depicting the same or different (mirror reversed along the vertical axis) objects, depicted from the same or from different viewpoints (i.e. rotated horizontally by 150°) – a 2 × 2 design (see Fig. 1 for an example of each type of test image). We selected 8 images for each experimental condition, so there were 32 trials in total, which were all completed in a uniform randomised order (i.e. in the same for all participants).

2.4. Procedure

Before mental rotation trials, participants were given the following instructions:

> *"You are about to see a number of pictures. Each depicts two objects. You task is to decide if they are the same object, depicted twice, or if the two objects are different.*
> *Sometimes the two objects will be shown from the same angle, and sometimes the object on the right will be rotated relative to the object on the left. When the two objects are shown from different angles, half the time they will be the same object, and half the time they will be different objects.*
> *Many people find it helps to imagine one of the two objects rotating in their mind, to see if it matches the other object after they have seen it rotate. Other people rely on an intuition, that the objects are the same or different. Still other people judge if prominent features and turning points of the objects would match if one of the two objects were physically turned – but they do not visualise either object as having turned.*
> *Your task on each trial is to judge as quickly and as accurately as possible whether the two objects are the same or different. You will then be asked how you came to that decision."*

After reading these instructions, participants completed 4 practise trials, one for each condition. On practise trials feedback was given regarding the accuracy of same/different decisions. Participants received no feedback on experimental trials. The images used for practise trials were not re-used in the formal experiment.

Test image presentations were shown until the participant had chosen the 'same' or the 'different' response option and had then clicked on an advance icon. The timing of the last response option selection (participants could change their responses) was taken as the response time for that trial (not the timing of the advance icon click). After this response sequence participants were asked, "*Did you visualise (have an imagined visual experience of) one of the objects rotating while making your decision?*". If they responded *No*, they were additionally asked "*Which of these best describes how you made your decision*". There were four response options for this question, **1)** *I could immediately tell*, **2)** *I guessed*, **3)** *I made a careful cross comparison of the two objects' features (without having any imagined experience)*, and **4)** *I used a strategy not described here*. If participants answered **4)**, at the end of the block of trials they were reminded that they had reported using a strategy which we had not described, and they were asked to describe that strategy (or strategies) in a text entry box.

3. Results

In all correlational analyses we excluded data points based on Mahalanobis distance calculations, (see McLachlan, 1999) using a significance level of 0.9. We also conducted a Shapiro-Wilk's test to assess if each dataset conformed with the normality assumption. When they did, a Pearson's correlation coefficient was calculated, otherwise a Spearman's rank correlation coefficient was calculated. Pearson's correlation coefficients are abbreviated as r in text and figures, and Spearman's rank correlations are abbreviated as rho.

Analyses of response times (RTs) relate to individual conditional median RTs for correct decisions. When describing the strength of correlations, we adhere to descriptors outlined by Dancey & Reidy (2007), where correlations $<= 0.3$ are described as weak, correlations > 0.3 and $<= 0.6$ are described as moderate, and correlations > 0.6 are described as strong. While it is common practise to conduct split-half analyses of similar datasets, we have chosen not to because such analyses are subject to statistical artefacts relating to regression to the mean that can result in the analyses having less statistical power than the X/Y correlational analyses that we report (see Cohen et al., 2003; Holmes, 2009; Shanks, 2017; Stoll et al., 2022).

All reported correlational analyses are focussed on trials on which participants either reported visualising an object rotating, or on which participants reported having relied on a non-visualising analytic response strategy. Participants also reported that 'I could instantly tell', most often on baseline trials (M = 0.61, SD = 0.18) and less frequently on different viewpoint trials (M = 0.07, SD = 0.13). None of the participants included in analyses ever reported that they had 'guessed' a response.

3.1. *People were slower and less accurate when making decisions about rotated objects*

First, we address commonly accepted hallmarks of visualisations during MR tasks – the increased time and difficulty associated with evaluating objects depicted from different, as opposed to the same viewpoint (viewpoint-contingent changes in performance). Consistent with these suggested hallmarks , we found that individual median response times (RTs) were *longer* when objects were depicted from different viewpoints (M = 4.86, SD = 2.13) as opposed to objects depicted from the same viewpoint (M = 2.22, SD = 0.8, paired $t_{92} = 17.3$, $p < 0.001$, $BF_{10} > 1$ million; see Fig. 2a). Similarly, when objects were depicted from different viewpoints, people

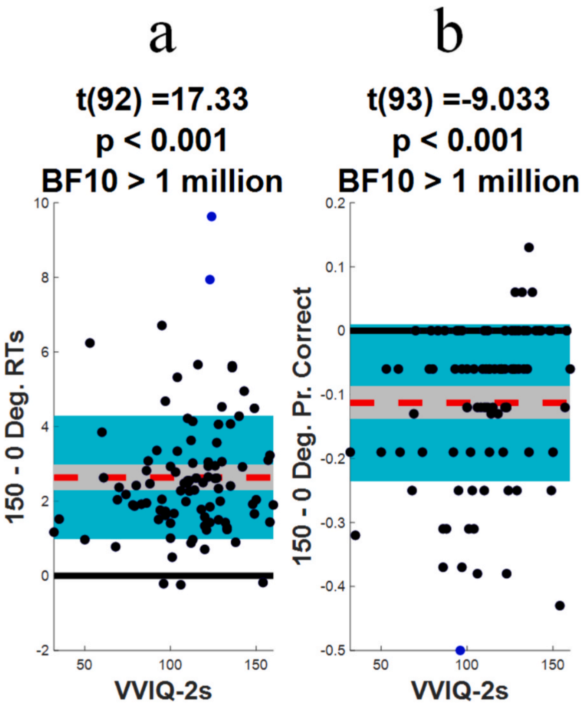

Fig. 2. (a) X/Y scatter plot of viewpoint differences (150° − 0° trials) in RTs (Y-axis) and VVIQ-2 scores (X-axis). (b) X/Y scatter plot of viewpoint differences (150° − 0° trials) in proportion correct same/different object decisions (Y-axis) and VVIQ-2 scores (X-axis). In both subplots, bold black horizontal lines depict 0 differences. Dotted red horizontal lines depict conditional average differences. Shaded aqua regions depict +/−1 S.D. from the average difference, and shaded grey regions depict +/− 2 S.E.s from the average difference. Blue data points are excluded from analyses as outliers (see main text). In both plots, plotting datapoints according to VVIQ-2 scores on the X-axis is just a convenient means of minimising overlap between plotted data points. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

made fewer correct same/different object decisions (M = 0.86, SD = 0.15), compared to when objects were depicted from the same viewpoint (M = 0.97, SD = 0.06, paired $t_{93} = -9$, $p < 0.001$, $BF_{10} > 1$ million; see Fig. 2b). These analyses confirm that our data are characterised by commonly accepted hallmarks of visualisation during MR tasks, and that our experimental protocols were sufficiently sensitive to detect Viewpoint-contingent changes in the speed and accuracy of same/different object decisions.

3.2. Viewpoint-contingent changes in RTs and decisional accuracy were not predicted by reliance on MR as a response strategy, overall or selectively on different viewpoint trials

Common assumptions are that people should be able to instantly discern if objects depicted from a common viewpoint are the same or different, and that they rely on visualising MRs as a response strategy when objects are depicted from different viewpoints (e.g. Shepard & Metzler, 1971; Ganis & Kievit, 2015). It follows that Viewpoint-contingent decreases in the speed and accuracy of same/different object decisions should be predicted by **1)** reliance on visualising as a response strategy on different viewpoint trials, and **2)** by overall reliance on visualising as a response strategy (as visualising objects rotating should *only* be used on different viewpoint trials).

In our data, decreases (150° − 0° trials) in the speed (rho = 0.15, CIs −0.08 to 0.35, p = 0.18; see Fig. 3c) and changes in the accuracy (rho = 0.12, CIs −0.1 to 0.33, p = 0.28; see Fig. 3d) of same/different object decisions were not associated with the proportion of *all* trials on which people reported relying on visualising as a response strategy. If we regard the overall proportion of trials on which people reported relying on visualising objects rotating as an estimate of each person's propensity to visualise, these data show that a propensity to visualise is not associated with viewpoint-contingent changes in decisional speed or accuracy. The individual

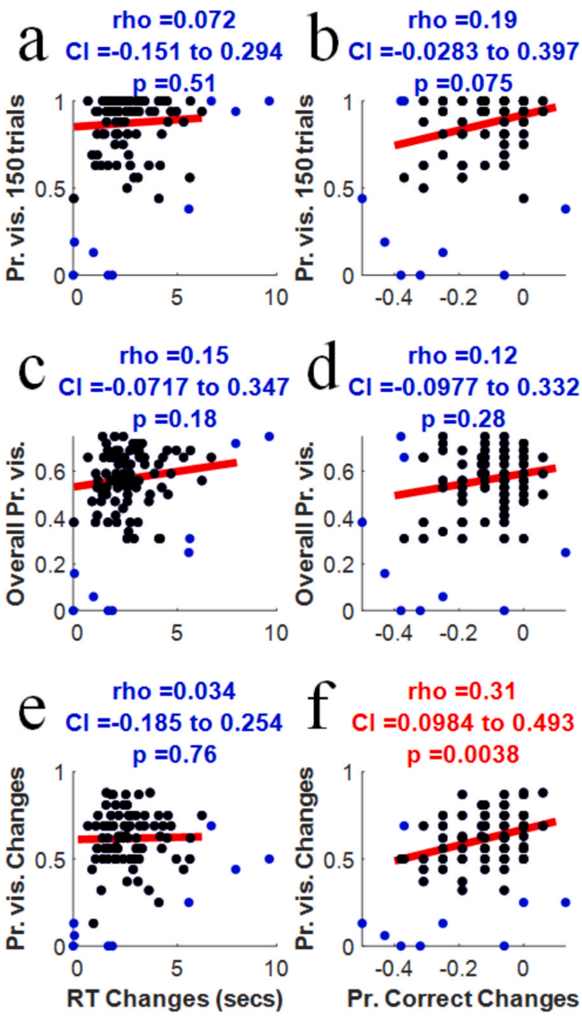

Fig. 3. a) X/Y scatter plot of the proportion of visualisations reported on 150° trials (Y-axis) and changes in RTs (150° − 0° trials). b) X/Y scatter plot of the proportion of visualisations reported on 150° trials (Y-axis) and changes in the proportion of correct MR decisions (150° − 0° trials). c-d) Details are as for Fig. 2a-b, except that data on Y axes relate to proportions of visualisations reported across all trials. e-f) Details are as for Fig. 2a-b, except that data on Y axes relate to changes in proportions of reported visualisations reported (150° − 0° trials).

average proportion of *all* trials on which visualising was used as a response strategy was 0.54 (SD = 0.17, see Fig. 3c-d).

Decreases in the speed (rho = 0.072, p = 0.51, see Fig. 3a) and changes in the accuracy (rho = 0.19, p = 0.075, see Fig. 3b) of same/different object decisions were also unrelated to the proportion of *different* viewpoint trials (150° trials) on which people reported relying on visualisation as a response strategy. The individual average proportion of different viewpoint (150°) trials on which visualisations were reportedly used was 0.83 (SD = 0.24, see Fig. 3a-b).

3.3. Viewpoint-contingent changes in the proportion of correct same/different object decisions were predicted by an increased reliance on visualisation as a response strategy

While decreases (150° − 0° trials) in the speed of same/different object decisions were not associated with individual changes (150° − 0° trials) in the proportion of trials on which people reported relying on visualisation as a response strategy (rho = 0.034, CIs −0.18 to 0.26, p = 0.76; see Fig. 3e), these changes were moderately positively associated with an increase (150° − 0° trials) in the proportion of correct same/different object decisions (rho = 0.31, CIs = 0.1 to 0.49, p = 0.004; see Fig. 3f). This means that people who relied more on visualisation as a response strategy, *selectively* on different viewpoint trials, were less likely to make incorrect same/different object decisions on different viewpoint trials. This implies that visualising an object rotating is a superior response strategy, relative to other strategies, when objects are depicted from different viewpoints.

3.4. People's overall propensity to visualise did not predict better performance on different viewpoint trials, as people with vivid imagery often visualised inappropriately

Our data have suggested that people's overall propensity to visualise, as given by the proportion of all trials on which they reportedly visualised, does not predict the speed or accuracy of same/different object decisions (see Fig. 3c-d). This somewhat contrasts with another finding, that increased use of visualisation, *selectively* on different viewpoint trials, was associated with fewer incorrect same/different object decisions (see Fig. 3f). So, why was overall use of visualisation not advantageous, when it was advantageous on different viewpoint trials? The answer must be that some people were using visualisation as a response strategy inappropriately, to no advantage, on same viewpoint baseline trials.

In Fig. 4a we plot the proportion of same viewpoint baseline trials on which people reported visualising (Y-axis), in relation to people's VVIQ-2 scores (X-axis) – which provide an estimate of the typical vividness of people's visualisations (Marks, 1995). First, it is clear that people often reported visualising on baseline trials (M = 0.25, S.D. 0.14), when it should have been immediately apparent if objects were the same or different. Moreover, the propensity to visualise on these trials was weakly positively associated with VVIQ-2 scores (rho = 0.25, CIs 0.04 to 0.44, p = 0.02; see Fig. 4a). So, people who reported having vivid visualisations were more likely to use visualisation inappropriately, as a response strategy on baseline trials.

We can show that visualising an object rotating was a disadvantageous response strategy on same viewpoint baseline trials by considering people's RTs (X-axis) in relation to the proportions of baseline trials on which these people reported visualising (Y-axis). There was a weak positive association (rho = 0.28, CIs 0.08 to 0.46, p = 0.009; N = 86, see Fig. 4b). So, visualising on same viewpoint baseline trials was a disadvantage, at least in terms of decision speeds. There was a ceiling effect in these data, relating to proportions of correct decisions (M = 0.97, S.D. 0.06), so a lack of an association between proportions of visualising trials and decisional accuracy (rho = 0.01, p = 0.92) on these trials is unremarkable.

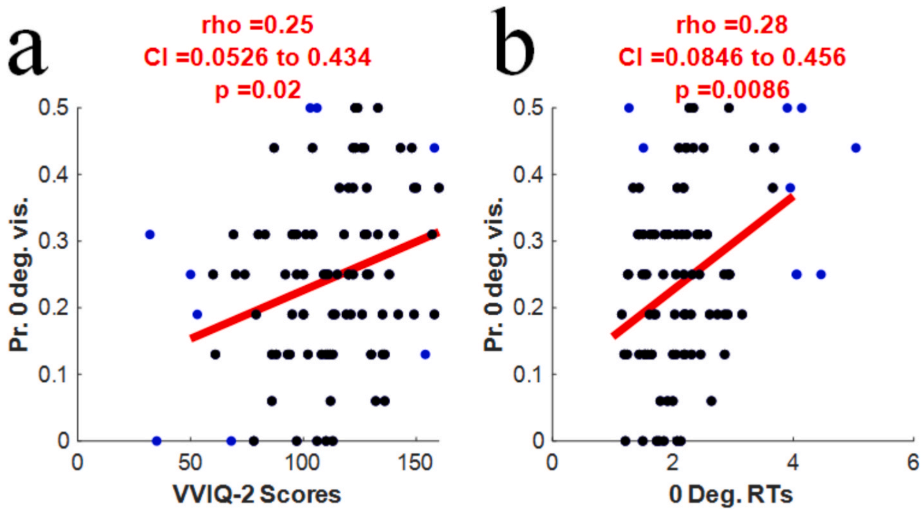

Fig. 4. (a) X/Y scatter plot of the proportion of visualisations reported on 0° trials (Y-axis) and individual VVIQ-2 scores (X-axis). (b) X/Y scatter plot of the proportion of visualisations reported on 0° trials (Y-axis) and individual RTs (X-axis). Other details are as described for Fig. 3.

3.5. Participants often reported using a non-visualising analytic response strategy

We can estimate people's propensity to use an analytic response strategy from the proportion of all trials on which they endorsed the response option, '*I made a careful cross comparison of the two objects' features (without having any imagined experience)*'. Overall, people reported using an analytic response strategy on 0.12 of all trials (SD = 0.14), and on 0.10 (SD = 0.18) of trials where objects had been shown from different viewpoints (see Fig. 5a). The majority of our participants (0.73) reported using a non-visualising analytic response strategy on at least some trials.

3.6. VVIQ-2 scores and analytic responses

People who reported having vivid imagery were overall weakly *less* likely to report using an analytic response strategy (r = -0.28, CIs −0.47 to −0.05, p = 0.01; see Fig. 5a). They were also weakly less likely to report using an analytic response strategy on different viewpoint trials (r = -0.24, CIs −0.43 to −0.03, p = 0.025; see Fig. 5b). Note that these data are not simply the inverse of people reporting having visualised as a response strategy, as people also reported that they had immediately known if objects were the same or different on 0.61 (S.D. 0.18) of same viewpoint (0°) trials, and on 0.07 (S.D. 0.13) of different (150°) viewpoint trials.

3.7. What if we had been more conservative, and used a statistical correction for multiple tests?

So far, we have not adjusted alpha levels to correct for multiple tests when considering the nominal significance of a reported test. Overall, we have reported on 11 statistical tests, so we could have applied a Bonferroni adjusted alpha level of 0.0045. If we had compared tests to this corrected alpha level, our tests for the hallmarks of visualising during MR tasks (i.e. slower response times and more errors on different viewpoint tests, see Fig. 2 and associated text) and for a positive association between viewpoint-contingent *changes* in proportions of wrong decisions and viewpoint-contingent *changes* in the likelihood of relying on visualisation as a response strategy (see Fig. 3f and associated text) would still be regarded as nominally significant. No other tests would be regarded as nominally significant against a Bonferroni adjusted alpha level of 0.0045.

3.8. What if we had used less conservative correlational analyses?

Our correlational analyses are conservative, in that we have used Mahalanobis distance calculations (see McLachlan, 1999) to exclude participants as outliers if their data were representative of the group. This precaution was designed to ensure our data were representative of the general population. If we had not excluded participants on this basis, results would only have been slightly impacted. Two correlations that were not nominally statistically significant at an uncorrected alpha of 0.05 (Pr. Different Viewpoint Trials Involving visualising X Pr. Viewpoint-contingent Changes in Response Accuracy, see Fig. 3a and related text; and Pr. All Trials Involving visualising X Pr. Viewpoint-contingent RT Changes, see Fig. 3c and related text) would have been regarded as nominally significant.

3.9. Descriptions of strategies that participants felt had not been available as a response option

Fifteen participants included in data analyses took the opportunity to report using a response strategy that we had not described. Of these descriptions, we regarded seven as having described another form of visualisation. Three of these involved imagining viewing objects from different viewpoints (e.g. "*I imagined myself standing looking at each object from a certain perspective and comparing them*"). One described a more complex form of visualisation ("*Following one side of the shape from one end to other. Do so to both and compare*

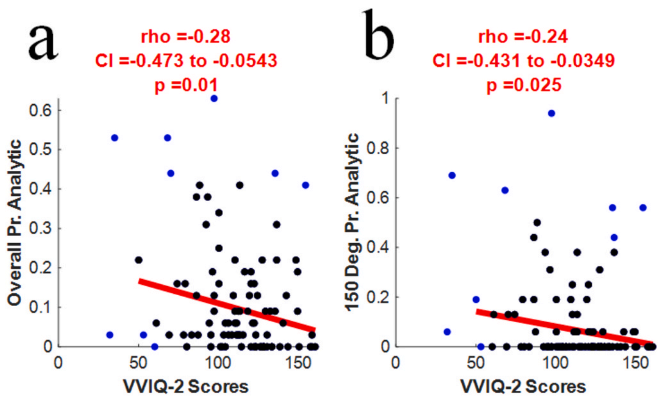

Fig. 5. X/Y scatter plot of the overall proportion of trials on which participants reported using a non-visualising Analytic response strategy (Y-axis) and individual VVIQ-2 scores (X-axis). (b) Details are as for Fig. 5a, except that individual proportions different viewpoint (150°) trials on which participants reported using a non-visualising Analytic response strategy are depicted on the Y-axis.

visualised line"), and one described a form of motor imagery ("*Careful examination of the 3D shape, alongside hand gestures to see the direction of the cubes*"). The other two were differently worded descriptions of visualisations (e.g. "*I rotated it around*").

We regarded another seven responses as having described a different analytic strategy. Three highlighted symmetry detection (e.g. "seeing they were a reflection of each other, which means they cannot be rotated to appear the same"), three described a counting strategy (e.g. "*Count blocks going in each direction*"), and one described an analytic strategy involving the comparison of prominent features ("…*if they are have a bend on the same side when they are 180 different then they are different*"). The final description described confusion ("*The image on the right is not fully displayed, so I can't confirm if they are the same*"). These data highlight that people adopt a range of visualising and analytic response strategies when completing a mental rotation task.

4. Discussion

Our data suggest that mental rotation tasks are a weak measure of people's propensity to visualise. Participants often reported relying on non-visualising response strategies (also see Bethell-Fox & Shepard, 1988; Hegarty, 2010, 2018; Kay et al., 2024; Khooshabeh et al., 2013), and they frequently reported having visualised on same viewpoint baseline trials (see Fig. 4a) when this was a disadvantageous response strategy – at least in terms of speed (see Fig. 4b). Crucially, there was no relationship between proportions of different viewpoint trials that reportedly involved visualisations and viewpoint-contingent changes in either the speed (see Fig. 3a) or accuracy (see Fig. 3b) of same/different object decisions. Viewpoint-contingent *changes* in decisional accuracy were weakly positively associated with Viewpoint-contingent *changes* in reported reliance on visualisation as a response strategy (see Fig. 3f). However, trial-by-trial information regarding response strategies is seldom recorded in MR studies, so this weak inter-relationship would rarely be accurately measured.

One interesting facet of our data is that participants often reported visualising on same viewpoint baseline trials (see Fig. 4a), when this strategy was disadvantageous (see Fig. 4b). These visualisations were selective to trials depicting *different* objects, as we had taken the precaution of excluding participants who had ever reported visualising on trials where the same object had been depicted from the same viewpoint (for reasons we explain below, see subsection titled 'Why did we exclude so many participants?').

A clear implication of our data is that trials that involve different objects depicted from the same viewpoint will likely involve some attempts to visualise. These data have often been treated as a visualisation free 'baseline', that can either be subtracted from data from all other conditions (e.g. Zhao et al., 2022), or which can be used as an anchor point, theoretically devoid of the time required to visualise, from which a slope involving datapoints from other conditions can be calculated to estimate MR speeds (e.g. Gardony et al., 2014; Kail & Park, 1990; Shepard & Metzler, 2971; Zeman et al., 2010). In either case, the inadvertent inclusion of visualisations within a supposedly visualisation free baseline could distort estimates of the propensity to visualise, and of the impact of visualisations.

Some researchers have been mindful of the possibility that participants might attempt to visualise on baseline trials, so they have adopted the practise of excluding trials that depict different (mirror reversed) objects from analyses that attempt to estimate the speed of MRs (e.g. Hilton et al., 2022; Searle & Hamm, 2017). This, however, does not protect against the possibility of people using a non-visualising response strategy on the remaining trials. We did not exclude trials involving the same object depicted twice from the same viewpoint, as a key point of interest was the propensity of people to use different response strategies.

4.1. Were our RT measurements accurate?

One reasonable concern might relate to the fidelity of our RT data. We relied on an online survey platform, Qualtrics (Qualtrics, Provo, UT) to record times since stimulus onset at which participants had selected their final same/different object response. Participants were allowed to change their initial responses, if they decided these had been inaccurate. This may have contributed to the relatively slow average response times (same viewpoint M 2.22 s, different viewpoint M 4.86 s). However, our analyses were sensitive to Viewpoint-contingent changes in RTs (see Fig. 2a). Unlike our response accuracy data, Viewpoint-contingent changes in RTs were simply unrelated to changes in response strategies (see Fig. 3a, c and d).

4.2. Why did we exclude so many participants?

We and others (e.g. Fleming & Bowden, 2009; Griffin et al., 2022) have found that the quality of data gathered in online studies can be poor, relative to lab-based investigations. So, we took a precaution to identify and exclude any participant who might not have been fully engaged with our task.

Our task involved a 2-stage response cycle, with people first reporting on whether they had visualised an object rotating. They were only asked for a second response if they did not endorse this first option. This encouraged participants who were primarily motivated to rapidly finish the study to endorse the first option – even when this did not make sense (i.e. when test images showed the *same* object depicted from the *same* viewpoint, see Fig. 1a). We therefore removed *all* participants from analyses who had indicated that they had visualised an object rotating on any of these trials. Note that this will not only have removed people who were primarily motivated to rapidly finish the study, it will also have removed some people who were simply inattentive.

4.3. Do people have metacognitive insight into the strategies they use during MR tasks?

Our conclusions rely on the proposition that people can accurately report on the response strategies they have used when making MR task decisions. Evidence from metacognition studies of neurotypical people supports this notion, with people performing better on trials on which they reported having had a greater level of decisional confidence than was typical for that individual (e.g. Estes, 1998; Ordin et al., 2024). Indeed, evidence suggests that neurotypical people generally have good levels of metacognitive insight into their performance on cognitive tasks (see Fleming et al., 2012; Yeung & Summerfield, 2012; Keane et al., 2015).

4.4. Do aphantasics have metacognitive insight into the strategies they use during MR tasks?

It has been suggested that people with Aphantasia might be able to have imagined visual sensations, which guide their performance on mental imagery tasks, but lack insight into this capacity (Nanay, 2021; Siena & Simons, 2024; Pounder et al., 2022). We think this is unlikely. The key evidence that motivated this proposition, that Aphants can perform similarly to neurotypical people on MR tasks (e.g. Jacobs et al., 2018; Pounder et al., 2022; Zhao et al., 2022), can be accounted for without having to assume a deficit of metacognition unique to people with aphantasia. Our data highlight that people generally use a number of different response strategies when completing MR tasks, and these can result in similar (Bethell-Fox & Shepard, 1988; Hegarty, 2018; Kay et al., 2024; Khooshabeh et al., 2013) or even superior (Hegarty, 2010) levels of performance. So, it is possible that Aphants perform similarly on MR tasks by using a variety of non-visualising response strategies (Bouyer & Arnold, 2024; Dawes et al., 2020; Takahashi et al., 2023) at which they are practised (Bethell-Fox & Shepard, 1988; Hegarty, 2018; Kay et al., 2024; Khooshabeh et al., 2013).

We suggest that both neuro-typical people (Estes, 1998; Ordin et al., 2024; also see Arnold et al., 2024; Fleming et al., 2012; Yeung & Summerfield, 2012) and Aphants likely have good metacognitive insight into their effortful cogitations during MR tasks, and that they can report on these if asked to (also see Dawes et al., 2020; Dawes et al., 2024). Future studies that seek to understand the cognitive processes that underlie MR task performance, and to clarify how these relate to the capacity to visualise, should incorporate trail-by-trial reporting on response strategies.

5. Conclusions

While we have discussed how our findings might relate to Aphantasia, our study was focussed on how people *generally* perform MR tasks. In particular, we asked if MR tasks generally provide a reliable metric of people's propensity to use visualisation as a response strategy in MR tasks?

Our data suggest that Viewpoint-contingent changes in MR task performance cannot be predicted from people's general propensity to visualise (as given by the overall proportion of trials on which they report visualising) or from the proportions of different viewpoint trials that people reportedly visualise in. Rather, there was a *weak* interrelationship between Viewpoint-contingent *changes* in the proportion of visualising trials and Viewpoint-contingent *changes* in decisional accuracy. As this inter-relationship is weak, it provides a poor basis to predict behaviour at an individual level (i.e. knowing that an individual can or cannot visualise would not allow us to accurately predict how they would perform on a MR task). Moreover, as the vast majority of MR studies do not quiz people on their response strategies on a trial-by-trial basis, these studies are unlikely to provide a reliable metric of the dynamics of MR, or of people's propensity to rely on visualisation as a response strategy.

Author contributions

D.H.A conceived of the study, programmed the experiment, analysed data, created figures, and wrote the first draft of the manuscript. All other authors contributed to conceptual discussions informing the study design and edited successive drafts of the manuscript,

CRediT authorship contribution statement

Derek H. Arnold: Writing – original draft, Funding acquisition, Formal analysis, Conceptualization. **Loren N. Bouyer:** Writing – review & editing, Conceptualization. **Blake W. Saurels:** Writing – review & editing, Conceptualization. **Elizabeth Pellicano:** Writing – review & editing, Conceptualization. **D. Samuel Schwarzkopf:** Writing – review & editing, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by a Discovery Project Grant, DP250101200 funded by the Australian Research Council, awarded to D. H.A., D.S.S and E.P.

Data availability

All data and analysis scripts will be made publicly available via UQ eSpace

References

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 12–23.

Bouyer, L. N., & Arnold, D. H. (2024). Deep Aphantasia: A visual brain with minimal influence from priors or inhibitory feedback? *Frontiers in Psychology, 15*, Article 1374349.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology*. Pearson education.

Dawes, A. J., Keogh, R., Andrillon, T., & Pearson, J. (2020). A cognitive profile of multi-sensory imagery, memory and dreaming in aphantasia. *Scientific Reports, 10*, 10022.

Dawes, A. J., Keogh, R., & Pearson, J. (2024). Multisensory subtypes of aphantasia: Mental imagery as supramodal perception in reverse. *Neuroscience Research, 201*, 50–59.

Fleming, S. Dolan, R. & Frith, C. (2012). Metacognition: Computation, biology and function *Philosophical Transactions of the Royal Society B: Biological Sciences* **367,** 1280–1286.

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 Files, 3*, 1–31.

Gardony, A. L., Taylor, H. A., & Brunyé, T. T. (2014). What does physical rotation reveal about mental rotation? *Psychological science, 25*, 605–612.

Hegarty, M. (2018). Ability and sex differences in spatial thinking: What does the mental rotation test really measure? *Psychonomic Bulletin and Review, 25*, 1212–1219.

Hegarty, M. (2010). Components of spatial intelligence. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 52, pp. 265–297). San Diego, CA: Academic Press.

Hilton, C., Raddatz, L., & Gramann, K. (2022). A general spatial transformation process? Assessing the neurophysiological evidence on the similarity of mental rotation and folding. *Neuroimage: Reports, 2*, Article 100092.

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: Some statistical considerations. *Brain Topography, 21*, 168–176.

Jacobs, C., Schwarzkopf, D. S., & Silvanto, J. (2018). Visual working memory performance in aphantasia. *Cortex, 105*, 61–73.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems. Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92*, 137–172.

Kail, R., & Park, Y. S. (1990). Impact of practice on speed of mental rotation. *Journal of Experimental Child Psychology, 49*, 227–244.

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, 103694*.

Keane, B., Spence, M., Yarrow, K., & Arnold, D. H. (2015). Perceptual confidence demonstrates trial-by-trial insight into the precision of audio-visual timing perception. *Consciousness & Cognition, 38*, 107–117.

Khooshabeh, P., Hegarty, M., & Shipley, T. F. (2013). Individual differences in mental rotation: Piecemeal versus holistic processing. *Experimental Psychology, 60*, 164–171.

Marks, D. F. (1995). New directions for imagery research. *Journal of Mental Imagery, 19*, 153–167.

Marks, D. F. (1999). Consciousness, mental imagery and action. *British Journal of Psychology, 90*, 567–585.

McLachlan, G. J. (1999). Mahalanobis distance. *Resonance, 4*, 20–26.

Nanay, B. (2021). Unconscious mental imagery. *Philosophical Transactions of the Royal Society of London B* 37620190689.

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.

Searle, J.A. & Hamm, J.P. (2017). Mental rotation: an examination of assumptions. *Wiley Interdisciplinary Reviews: Cognitive Science* **8,** 701–703.

Shanks, D. R. (2017). Regressive research: The pitfalls of post hoc data selection in the study of unconscious mental processes. *Psychonomic Bulletin & Review, 24*, 752–775.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Siena, M. J., & Simons, J. S. (2024). Metacognitive awareness and the subjective experience of remembering in aphantasia. *Journal of Cognitive Neuroscience, 36*, 1578–1598.

Stoll, S., Infanti, E., De Haas, B., & Schwarzkopf, D. S. (2022). Pitfalls in post hoc analyses of population receptive field data. *NeuroImage, 263*, Article 119557.

Takahashi, J., Saito, G., Omura, K., Yasunaga, D., Sugimura, S., Sakamoto, S., et al. (2023). Diversity of aphantasia revealed by multiple assessments of visual imagery, multisensory imagery, and cognitive style. *Frontiers of Psychology, 14*, Article 1174873.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Science, 367*, 1310–1321.

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*, 145–155.

Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery—congenital aphantasia. *Cortex, 73*, 378–380.

Zhao, B., Della Sala, S., Zeman, A., & Gherri, E. (2022). Spatial transformation in mental rotation tasks in aphantasia. *Psychonomic Bulletin & Review, 29*, 2096–2107.

Logie, R. H., Pernet, C. R., Buonocore, A., & Della Sala, S. (2011). Low and high imagers activate networks differentially in mental rotation. *Neuropsychologia., 49*, 3071–3077.

Fleming, C. M., & Bowden, M. (2009). Web-based surveys as an alternative to traditional mail methods. *Journal of Environmental Management, 90*, 284–292.

Griffin, M., Martino, R. J., LoSchiavo, C., Comer-Carruthers, C., Krause, K. D., Stults, C. B., & Halkitis, P. N. (2022). Ensuring survey research data integrity in the era of internet bots. *Quality & Quantity, 56*, 2841–2852.

Ordin, M., El-Dakhs, D. A. S., Tao, M., Chu, F., & Polyanskaya, L. (2024). Cultural influence on metacognition: comparison across three societies. *Humanities & Social Sciences Communications, 11*, 1492.

Estes, D. (1998). Young children's awareness of their mental activity: The case of mental rotation. *Child Development., 69*, 1345–1360.

Arnold, D. H., Clendinen, M., Johnston, A., Lee, A. L. F., & Yarrow, K. (2024). The precision test of metacognitive sensitivity and confidence criteria. *Consciousness & Cognition., 123*, Article 103728.