# 1. Introduction

Our ability to understand and reason from sensory information depends on our ability to represent it in our mind; however, the precise nature of these representations has been the subject of a long-standing debate. Concerning the processing of visual information, two opposing conceptions have emerged: one suggesting that it relies on symbolic representations, in a non-visual, propositional format; the other on pictorial representations, in a depictive format, similar to a weakened version of perception (Palermo et al., 2022; Pearson, 2019). However, representational formats are not restricted to this dichotomy, since other strategies have been proposed, such as spatial or sensorimotor representations (Palmiero et al., 2022; Reeder et al., 2024).

To explain the processing of visual and spatial information, mental model theory (Johnson-Laird, 2001, 2006) succeeded in integrating spatial and depictive formats within the same framework, respectively with mental models and visual mental images. Numerous studies on human reasoning, mainly conducted by Johnson-Laird’s team, suggested that this cognitive ability relies on mental models, i.e. representations that faithfully indicate the positions and spatial relationships between the elements presented in a problem (Johnson-Laird, 2010; Krumnack et al., 2011; Thevenot & Perret, 2009). These small-scale reproductions of reality are considered similar to schemas and diagrams (Engel, 1994).

Mental models can thus spatially represent any type of situation, regardless of visual details such as colour, shape or texture (Johnson-Laird & Byrne, 1991). Within their framework, a visual mental image is viewed as one of many possible representations of the visualisable aspects of a given mental model, itself constructed from propositional representations (Sima et al., 2013). However, the view that reasoning is underpinned solely by spatial representations is challenged by mental imagery theory, according to which visual mental images play a central role in reasoning processes (Kosslyn, 1994). Mental imagery theory also distinguishes between visual and spatial mental images (see Sima et al., 2013 for a review). According to Kosslyn et al. (2006), the analysis of visual mental images can provide additional information that are complementary to the spatial image and leads to new understandings of the problems. Accordingly, they never specify that spatial mental images are crucial for reasoning.

Three- and four-term series problems have made an important contribution in the investigation of the role of spatial and visual representations in reasoning (e.g., Albrecht et al., 2015; De Soto et al., 1965; Knauff et al., 2003; Knauff & Johnson-Laird, 2002; Sima et al., 2013). These problems are deductive relational reasoning problems in which the relationship between two elements A and C must be inferred from the relationships between A and B, and between B and C. They therefore consist of several statements (called premises) followed by a conclusion to which participants must respond true or false. In their experiment, Knauff & Johnson-Laird (2002) varied the nature of the relations used in these problems, employing visual, visuo-spatial, or control pairs. They proposed that visual relations “automatically” solicit visual imagery and that these images contain irrelevant details that would disrupt reasoning processes, summarizing this as the “Visual-Imagery-Impedance Hypothesis” (VIIH). Slower responses to visual problems were indeed observed compared to visuo-spatial or control problems. This pattern of results has been replicated several times for response times (Knauff et al., 2003; Knauff & Johnson-Laird, 2002; Knauff & May, 2006; Tse et al., 2017) and was also observed for the accuracy levels, with a reduced rate of correct responses for the visual category (Knauff & May, 2006).

If the VIIH proves to be true, then a population that does not experience mental images should be immunized against this effect. Aphantasia, reduced or absent visual imagery, offers a unique opportunity to test this hypothesis. This condition could affect 3-4 % of the global population (see for instance Dance et al., 2022; Palermo et al., 2022; Wright et al., 2024) and can be understood as a form of neutral neurodivergence, without causing any impairment in their socioprofessional functioning (Monzel et al., 2023). Indeed, aphantasics are often unconscious of their own condition and seem to use alternative strategies in their daily life. Interestingly, numerous studies indicate that they can accurately perform many tasks that were previously thought to rely on visual imagery, including working memory tasks (e.g., Keogh et al., 2021; Knight et al., 2022; Reeder et al., 2024), mental comparison tasks (Liu & Bartolomeo, 2023) or tasks assessing visual and verbal declarative memory (Pattern Recognition Memory and Verbal Recognition Memory, respectively, see Pounder et al., 2022). The differences that have been observed in aphantasics, are most of the time reflected by increased response times, particularly in tasks that require fine visual working memory (Jacobs et al., 2018; Monzel & Reuter, 2024; Pounder et al., 2022). The fact that aphantasics performed as well as a control population suggests the use of alternative strategies. Reeder et al. (2024), after administering a visual working memory task, asked their participants to indicate the frequency of use of five information retention strategies: visual, spatial, verbal, semantic and sensorimotor. Their results showed that aphantasics rely on non-visual alternative strategies, without this affecting their performance. Complete aphantasics predominantly used spatial strategies, while hypophantasics (i.e. individuals whose visual imagery is dim and vague rather than absent) used a combination of spatial and sensorimotor strategies.

Although not evidenced by the latter study, the representations participants rely on are likely to influence their performance in several tasks while reflecting distinct cognitive styles. Within this perspective, each cognitive style may entail specific advantages as well as disadvantages - precisely what Blazhenkova & Kozhevnikov (2009) set out to explore with their Object-Spatial Imagery and Verbal Questionnaire (OSIVQ). This scale measures self-reported preferences for mentally representing the visual characteristics of an object or scene (i.e., shape, colour, texture, etc.); for schematic images indicating spatial positions and relationships between objects, or for a verbal and semantic representation. The exploration of these cognitive styles and their impact on real-world activities showed that scientists, engineers, and architects were found to rely more on spatial imagery, while visual artists employ more object imagery. Furthermore, investigation of verbal and semantic style preference showed that humanities specialists scored highest on verbal items (Blazhenkova & Kozhevnikov, 2010; Kozhevnikov et al., 2010). Turning now to the field of visual imagery extremes, it is suggested that aphantasia reflects a particular cognitive style, illustrating a “semantic and factual” mode of information processing, compared to an “episodic and sensorially rich” mode regarding hyperphantasia, the opposite of aphantasia, characterised by visual imagery “as vivid as real perception” (Pearson, 2019). Accordingly, it has been observed that aphantasia is more associated with scientific occupations (“Computer and Mathematical”/“Life, Physical and Social Sciences”), while hyperphantasia is more linked to artistic professions (Zeman et al., 2020).

This body of results served as a source of inspiration for Delem et al. (2025), who conducted a study comparing people with aphantasia with non-aphantasic individuals on a range of behavioral tasks and questionnaires, including reasoning tasks and the OSIVQ (Blazhenkova & Kozhevnikov, 2009). This study used a clustering method, derived from Machine Learning, which identified three groups characterised by distinct OSIVQ scores. A first cluster, composed solely of non-aphantasics, was distinguished by a higher score in visual-object imagery; a second mixed cluster (aphantasics and non-aphantasics) showed a preference for spatial imagery; while a final cluster, composed solely of aphantasics, favoured verbal strategies. Differences also appeared in behavioural tasks. Concerning control participants, those with strong spatial imagery outperformed those with strong visual imagery in tasks requiring reasoning and working memory, such as spatial span and verbal reasoning tasks (similarities subtest of the WAIS-IV, Wechsler et al., 2008). However, all participants (controls and aphantasics) performed similarly on tasks involving non-verbal reasoning (Raven Standard Progressive Matrices, Bilker et al., 2012) and reverse digit span. Thus, this study demonstrates two important results. First, although no differences were observed in cognitive ability between the three clusters, vivid visual imagery seems to impair verbal reasoning, in line with the VIIH’s predictions. Secondly, it demonstrates the heterogeneity of aphantasia: some individuals showed a preference for a spatial mode of representation, while others rely on a more verbal and semantic processing, without this affecting their performances in several tasks. Secondly, it demonstrates the heterogeneity of aphantasia: some individuals showed a preference for a spatial mode of representation, while others opted for a more verbal and semantic processing, without this affecting their performances in several tasks.

In the field of aphantasia, numerous studies have focused on the impairments associated with the condition. However, it may be valuable to demonstrate that aphantasia, like any other “cognitive style”, also possesses notable strengths. Administering a reasoning task similar to that of Knauff & Johnson-Laird (2002) may offer further support for this argument. This type of task has previously been proposed to congenitally blind individuals, a population that also lacks mental imagery. Knauff & May (2006) demonstrated that, although observed in a population of sighted individuals (blindfolded or unblindfolded), the visual-imagery-impedance effect (VIIE) was absent in congenitally blind participants since they showed the same level of performance regardless of the nature of the problems in terms of accuracy and response time. When blind participants were asked about the strategies they used, they reported using only spatial strategies.

In addition, Gazzo Castaneda & Knauff (2013) used the OSIVQ to divide participants into three groups according to their cognitive style, distinguishing between object-visualisers, spatial-visualisers, and verbalisers. In this study the category-dependent VIIE depended on the cognitive style of the participants. Overall object-visualisers showed the weakest performance compared to verbalisers and spatial-visualisers, i.e. longer response times regardless of the nature of the problems. According to the authors, verbalisers were unaffected by the visual characteristics of the problems, while object-visualisers, tending to rely on visual representations independently of the material, displayed the VIIE for all problems. Spatial-visualisers, on the other hand, showed a pattern similar to the classic VIIE, although not significant. However, these results require cautious interpretation, especially due to the small number of participants in each group (verbalisers = 9; spatial-visualisers = 6; object-visualisers = 13).

The purpose of our study was to administer a task similar to Knauff & Johnson-Laird (2002) to aphantasic participants in order to test the VIIH and to provide some insight into the role of visual and spatial representations in reasoning. We hypothesized that aphantasics, unable to generate mental images, would not show the VIIE on response times, unlike control participants. Additionally, using the OSIVQ, we used a classification distinguishing all participants (aphantasic and non-aphantasic) according to their cognitive style. This second, more exploratory stage of the study, aimed to investigate a possible difference in performance depending on cognitive style, inspired by the work of Gazzo Castaneda & Knauff (2013). Finally, we constructed a questionnaire based on the work of Reeder et al. (2024) to explore the alternative strategies used by aphantasics. We hypothesised that they would rely more on spatial and verbal strategies than visual ones when solving the reasoning task.

# 2. Methods

No part of the study procedures or analysis plan was preregistered prior to the research being undertaken. We report all data exclusions, all inclusion/exclusion criteria, all manipulations, and all measures in the study.

## 2.1 Questionnaires

### 2.1.1 Vividness of Visual Imagery Questionnaire (VVIQ)

The Vividness of Visual Imagery Questionnaire (VVIQ, Marks, 1973) is a self-report questionnaire consisting of sixteen items, each asking participants to imagine a particular scene and rate the vividness of their mental imagery using a Likert scale ranging from 1 (“No image at all, you just know you’re thinking about the object”) to 5 (“Perfectly clear and vivid as if it were normal vision”). The final score is between 16 and 80. The total score of 32, conventionally used as a threshold to define aphantasia, is equivalent to a score of 2 (“*vague and faint*”) for each item in the questionnaire. The internal reliability (Cronbach’s ) of the VVIQ is .88 (McKelvie, 1995).

### 2.1.2 Object, Spatial and Visual Imagery Questionnaire (OSIVQ)

The Object, Spatial and Visual Imagery Questionnaire (OSIVQ, Blazhenkova & Kozhevnikov, 2009) is a self-assessment scale for preferences in terms of information processing and representation, based on a theory that distinguishes three dimensions: visual-object imagery, visuo-spatial imagery, and verbal strategies. 15 items assess each dimension, for a total of 45 items, which the participants rate from 1 (“strongly disagree”) to 5 (“strongly agree”). Four items are negatively formulated and therefore reversed for analysis. For each dimension, values are added together to obtain a score ranging from 15 to 75, and thus identify the cognitive style of each participant (distinguished as object-visualisers, spatial-visualisers or verbalisers). Cronbach’s of the object, spatial and verbal scales are .83, .79 and .74 respectively (Blazhenkova & Kozhevnikov, 2009).

### 2.1.3 Raven’s Standard Progressive Matrices (RSPM-18)

Raven’s standard progressive matrices test assesses abstract reasoning and was employed as a control task in our study. This reduced version, consisting of two sets of nine items, accurately predicts the total score for all 60 items and has been validated by Bilker et al. (2012). Each of the short forms had correlations of with the long form, and respective Cronbach’s of .80 and .83. It reduces test time by up to 75%. Each item presents a 3x3 or 4x4 matrix with a missing figure. Participants must identify the logical rules underlying the configuration of each matrix to complete it by choosing the correct figure from several options. The difficulty of the task increases with each item.

### 2.1.4 Strategies questionnaire

A questionnaire was constructed in order to investigate the resolution strategies used by the participants, based on the work of Reeder et al. (2024). These authors identified five different strategies for the retention of information in working memory: visual, spatial, verbal, semantic, and sensorimotor. In this study, the same strategies were investigated, considering the central role of working memory in our reasoning task. After the latter, the participants had to judge the frequency of use of each strategy on the following scale: I only used this strategy; I mainly used this strategy; I used this strategy as much as others; I used this strategy in a secondary way compared to one or more dominant strategies; I did not use this strategy. These responses were numerically coded for analysis, with 5 being “I only used this strategy” and 1 being “I did not use this strategy”. Each strategy was defined to ensure that participants understood our terms: Visual (I used a mental image representing the order of letters as a whole or some specific details); Verbal (I rehearsed the order of the letters in my head); Spatial (I used a grade on which I placed the letters, and/or associated each letter with a position on the screen/wall/desk/etc.); Semantic (I gave meaning to the order of the letters, and I used a mental image/sound associated with that meaning); Sensorimotor (I associated the direction of my eyes with the position of each letter and/or I imagined pointing at the location of each letter). If none of these options suited them, the participants could provide more detailed descriptions of their methods.

## 2.2 Reasoning Task

The reasoning task was constructed on the basis of tasks used in previous studies employing 4-term series problems (Cortes et al., 2021; Knauff & Johnson-Laird, 2002; Tse et al., 2017). In these tasks, participants are asked to solve inference problems. Several statements, called premises, are presented one by one, followed by a conclusion. The participant must determine whether this conclusion follows logically from the premises presented previously by answering with “True” or “False”. The problems used are described as “determinate”, which means that from the premises only one configuration can be constructed, so the conclusion cannot be ambiguous.

These inference problems vary in the nature of the relation used. These have been validated by pilot studies aimed at identifying the nature of each relation by asking participants to rate the ease with which they represented each antonym pair visually and spatially (Knauff & Johnson-Laird, 2002; Tse et al., 2017). Thus, several relations were identified: visual (easy to visualise), spatial (easy to represent spatially), visuo-spatial (easy to represent visually and spatially) and control (difficult to represent visually or spatially). Recently, it has been suggested that “spatial” and “visuo-spatial” relations should be grouped together in a single category (Cortes et al., 2021). This adjustment was taken into account in the task design used in this study. The relations were divided into three categories: visual, spatial and control. Three relationships were chosen for each of the categories ([Table 1](#tbl-relations)), and we used these 9 relationships to construct 27 relational reasoning problems. Previous studies have shown that there was no difference between accepting valid inference (problem with a correct conclusion) and rejecting invalid conclusions (problem with a wrong conclusion) (Knauff & Johnson-Laird, 2002; Knauff & May, 2006). Thus, this factor was not controlled, and this task consisted of 13 problems with valid inference and 14 problems with invalid inference.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Relationships employed in the reasoning task, categorized according to their category (visual, spatial, control).   | Visual | Spatial | Control | | --- | --- | --- | | Cleaner/Dirtier | To the left/To the right | Slower/Faster | | Curlier/Straighter | In front/Behind | Braver/More cowardly | | Thinner/Thicker | Above/Below | Calmer/More agitated | |

Our task uses only 4-term series problems, each consisting of three premises and one conclusion. Moreover, these problems are said to be “semi-continuous”, which means that the subject of a premise is identical to the object of the preceding premise, except for the last premise preceding the conclusion.

For this task, we also choose to use letters instead of animal names or first names as in previous studies (e.g., Cortes et al., 2021; Knauff & Johnson-Laird, 2002). The objective was that visualisation of the relationships between letters varies only according to the nature of the category (control, visual or spatial). In the field of syllogistic reasoning, one of the categories of the NeuBAROCO – a database of more than 300 syllogisms – consists solely of letters from the alphabet which are considered neutral and not influenced by beliefs (Ando et al., 2023). Following their evidences, the letters A, B, C and D were used in our study.

Thus, the version of the task used in this study consisted of 27 semi-continuous 4-term series problems, using letters. The problems were divided into three categories: visual, spatial and control. Each category had three relationships (i.e., antonyms pairs) and three problems were created from each relationship. Each letter appeared the same number of times and each antonym appeared as many times as its inverse (e.g., “Clean” and “Dirty” are used six times each throughout the experiment).

Each premise and the conclusion were presented on separate screens, and the participants had to press the space bar to read the next premise or the conclusion (self-paced design; see [Figure 1](#fig-procedure)). The premises were presented in blue letters, while the conclusion was in red. Participants were asked to evaluate whether the given conclusion follows from the premises, pressing the D (yes) or K (no) keys. Reading time for each premise, response time to the conclusion, and the response were recorded.

Four practice trials were presented before the experimental phase. Each used a different relation from those used in the experimental phase (e.g., silent/noisy, young/old). The latter consisted of 27 problems, presented randomly. For all nine problems, a pause screen suggested that participants take a break, although they could take one after each problem.

|  |
| --- |
| Figure 1: Reasoning Task Procedure. Example of a whole problem. The trial begins with a fixation cross displayed for 500 ms, followed by premise 1. The participant scrolls through the premises at their own pace by pressing the space bar. Finally, the conclusion is presented and the participant responds with ‘True’ (key d) or ‘False’ (key k). |

## 2.3 Online Data Collection

The questionnaires and the reasoning task were computerised using SurveyJS and jsPsych (Leeuw et al., 2023), open-source JavaScript libraries dedicated to the creation of online questionnaires and experiments respectively. They were hosted on a JATOS server (Lange et al., 2015) owned by the University Lumière Lyon 2. The study was conducted online and shared through a website dedicated to a multi-experiment research project on aphantasia (https://innerexperiencelab.com). Participants were recruited through social networks or websites dedicated to sharing scientific experiments, as well as on aphantasia-specific Facebook or Instagram pages (@aphantasiaclub). Participants could complete several experiments on the website, some of them containing the same questionnaires (e.g., all included the VVIQ). To allow for a more fluid experience, participants were given the option to skip questionnaires they had already filled in another experiment and were not required to complete all questionnaires and tasks in a single setting. A unique code allowed them to easily reconnect and resume the experiment. An anonymised identifier was created for each participant based on their code and a private encryption key, enabling us to gather all data from a single participant across all online experiments proposed on the website.

As part of this study, participants were presented with the VVIQ, the OSIVQ, the reasoning task, the strategy questionnaire, the RSPM-18 and a feedback questionnaire about the experiment.

## 2.4 Participants

Participants had to be French speakers and had normal or corrected vision. 137 participants completed the main reasoning task of the experiment. As participants were allowed to skip certain questionnaires, not all had a complete dataset. We excluded participants who had not completed the VVIQ, OSIVQ or RSPM-18 in any of our online experiments, those who admitted to cheating or distraction in the post-experiment feedback form, as well as those whose accuracy was below 50% in the main task. The final sample used for analyses comprised 104 complete datasets. Using the most widely used threshold to define aphantasia (VVIQ 32), this sample contained 47 aphantasics (34 female, 1 other gender, = 18.79, = 4.56, = 35.26, = 12.97) and 57 “typical imagers” (VVIQ 32, 45 female, 1 other gender, = 57, = 12.89, = 31.63; = 10.67).

Participation was voluntary and not remunerated. The study was carried out following the recommendations of the French Law (Loi Jardé n◦2012- 300) and informed consent was obtained from all participants following the Declaration of Helsinki.

## 2.5 Data Analysis

Data analysis was programmed in R language (version 4.5.1, R Core Team, 2025) on RStudio (version 2025.5.1.513, Posit team, 2025). The data and code have been structured in an R package available on GitHub (Delem, 2025) to maximise the reproducibility of the analyses. The online documentation of the package (<https://m-delem.github.io/aphantasiaReasoningViie/>) contains further detailed information on all aspects of the data analysis process, including power analysis, sample description, data preparation, confirmatory and exploratory analyses. All source data (N = 137), a self-contained version of the package and PDF versions of the supplementary materials are also available on the Open Science Framework ([https://osf.io/hfbcp/?view\_only=0ff6ba4fba3e46b281de0c3cc4f94a55](https://osf.io/hfbcp/?view_only=0ff6ba4fba3e46b281de0c3cc4f94a55%7D)) as an alternative to GitHub.

### 2.5.1 Power analyses

As the experiment was conducted online, the sample size was mostly limited by the time resources of the project. We estimated the statistical power conferred by various sample and effect sizes using a simulation approach, where power is defined as the proportion of cases where a model detects an existing effect that we simulated. Instead of choosing a fixed sample size, this approach allowed us to have a full picture of the power of our models across a range of sample and effect sizes. For any given (reasonable) sample size, we managed to estimate the smallest effect size we could detect with good power, allowing for more flexibility in data collection.

We based the power analyses on the VIIE observed by Tse et al. (2017) on response times, as they had the closest paradigm to ours and well-documented statistics. Their data allowed us to find good parameters to simulate realistic data based on our theoretical model and hypothesised effects. We hypothesised an increase in RT for the typical imagery group in visual problems that would not be present in aphantasics. We simulated data for sample sizes between 10 and 200 participants, with varying effect sizes for the VIIE (RT difference between the visual and control/spatial categories) and group interaction (nullifying the VIIE for the aphantasia group), ranging from 0.5s to 2.5s effects (Tse et al. (2017) observed a 2.5s VIIE, so we examined pessimistic scenarios). The reproducible code used to perform these analyses is available in the supplementary materials online. The complete results of the power analyses are displayed in [Figure 2](#fig-power).

|  |
| --- |
| Figure 2: Results of the power analyses by simulation. The semi-transparent dots represent the proportion of successful detections of a simulated effect among 350 simulations for each combination of sample and effect size (a total of 147,000 simulated datasets). The smooth opaque lines represent non-linear models fitted to the power curves for ease of reading. Two horizontal lines indicate important thresholds at 80% and 90% statistical power. These curves allow us to evaluate the power conferred by a sample to detect various effect sizes or, conversely, to predict the sample size needed to accurately trace any estimated effect size. |

Using a conventional VVIQ = 32 threshold, our final sample allowed us to have 50 participants per group for the main analyses. According to the simulations, this would allow to detect VIIEs (or interactions) 1.9s with 90% power, or 1.6s with 80% power. Alternatively, these analyses offer another perspective for understanding the results of the models (notably statistical significance) based on effect and sample sizes: to detect a significant 2.5s VIIE with 90% power, only 35 participants per group are required, but at least 70 per group are required to detect a 1.5s VIIE reliably, and more than 200 are required to detect a 0.9s VIIE.

### 2.5.2 OSIVQ clustering

In addition to the usual VVIQ groups, we decided to use the OSIVQ sub-scales to analyse the link between cognitive styles, mental imagery and VIIE. Following the object-spatial-verbal model (Blazhenkova & Kozhevnikov, 2009) and the methodology proposed by Delem et al. (2025), we divided the sample in visualiser, spatialiser and verbaliser sub-groups using clustering algorithms on the three OSIVQ sub-scale scores. To consider the data from several angles and aim for a robust partition, we used a “consensus” method between various algorithms (Gaussian Mixture Modelling, Partitioning Around Medoids and Fuzzy C-means Clustering) using the *diceR* package (Chiu & Talhouk, 2025). The consensus solution was determined using hierarchical clustering on the results of the other algorithms.

### 2.5.3 Accuracy models

We fitted Generalised Linear Mixed Models (GLMMs) with binomial distributions and logit links using the *glmmTMB* package (McGillycuddy et al., 2025) to predict accuracy with a grouping variable (VVIQ groups, OSIVQ clusters), Category (visual, spatial, or control) along with their two-way interactions as fixed categorical predictors. Varying slopes and intercepts (“random effects”) have been added for each participant by category and for each problem by grouping variable.

### 2.5.4 Response time models

For RT analysis, we removed incorrect trials (587 trials, 21%), trials where a screen (premise or conclusion) was displayed for less than 0.6s or more than 30s (thresholds determined by examining RT distributions) and trials where total response times did not fall within +/- 2 standard deviations of the mean response time of individual participants (317, 14% of correct trials), in accordance with the methodology adopted by Tse et al. (2017). 1904 trials (from 104 participants) remained in the final dataset.

We fitted GLMMs with Gamma distributions and identity links to account for the skewed distributions of RTs, using the *glmmTMB* package (McGillycuddy et al., 2025). The models included a grouping variable (VVIQ groups, OSIVQ clusters), Category (visual, spatial, or control) along with their two-way interactions as fixed categorical predictors. Varying slopes and intercepts (“random effects”) have been added for each participant by category and for each problem by grouping variable (following the same structure as the accuracy models).

### 2.5.5 Mixed models contrast analyses

Due to the way that variance is partitioned in GLMMs (Rights & Sterba, 2019), there does not exist an agreed-upon way to calculate standard effect sizes for individual terms such as main effects or interactions in these models. Thus, in line with general recommendations on how to report effect sizes (e.g., Pek & Flora, 2018), we report and analyse unstandardised effect sizes for post-hoc tests in the form of estimated marginal contrasts (i.e. differences in model-estimated marginal means, hereinafter denoted ), in seconds for RTs or as odds ratios for accuracies. To answer our hypotheses, we planned to analyse contrasts between groups, contrasts between categories for each group separately, and interaction contrasts testing the differences in category contrasts between the groups.

### 2.5.6 Strategy models

Ordinal cumulative link regression models were fitted using the *ordinal* package (Christensen, 2023) to predict the score (on a question about the use of a given strategy) with a grouping variable (VVIQ groups, OSIVQ clusters), Strategy (visual, verbal, spatial, semantic or sensorimotor) and their two-way interaction as fixed categorical predictors. We planned to analyse the contrasts between groups for each strategy separately.

### 2.5.7 Non-linear RT models

In addition to the total response times per trial usually analysed in VIIE studies, we collected response time data for each trial phase (three premises and the conclusion). We explored whether the VIIE could be specific to certain phases, or whether it was a difference in the *dynamics* of reasoning across trial phases rather than an overall difference in speed, by modelling response times throughout a trial with non-linear models.

We fitted generalised additive models using the *mgcv* package (Wood, 2011) to predict response times with a grouping variable (VVIQ groups, OSIVQ clusters), Category (visual, spatial or control) and their interaction as fixed categorical predictors, as well as “smooth” (non-linear) terms that capture the evolution of RTs across trial terms (premise 1/2/3 and conclusion) for each grouping and category, each problem for each grouping, and each participant for each category (the latter two being equivalent to “random effects” in mixed models).

## 2.6 Results

Due to the large number of fitted models and contrast analyses performed, we chose to detail only the significant contrasts relevant to our hypotheses and to simplify the presentation of non-significant effects with *p*-values only, to facilitate reading and interpretation. Comprehensive tables with all the tests conducted on the models can be found in the supplementary materials.

### 2.6.1 Description of the groups

We planned on analysing visual imagery groups using VVIQ thresholds and cognitive style groups based on a clustering of OSIVQ sub-scale scores. For the VVIQ groups, we aimed for the fine-grained classification proposed by Reeder et al. (2024) that defines aphantasia as VVIQ = 16, hypophantasia as VVIQ , typical imagery as VVIQ and hyperphantasia as VVIQ . However, this has been limited by our sampling process, as to date it remains difficult to find ways of specifically recruiting people with hypo- or hyperphantasia, as opposed to “complete” aphantasics and typical imagers. There were only 4 hyperphantasics in the final sample (N = 104), so we grouped them with the typical imagers. The final sample comprised only 17 hypophantasics. This group size conferred low power, but close to being acceptable for detecting significant effects observed in the reasoning literature (75 80% power for 2.3 2.5s effect sizes). Thus, we decided to conduct the analyses using three different classifications:

1. A 2-group VVIQ classification that groups full aphantasia and hypophantasia in a broad “aphantasia” group (VVIQ 32) and compares it to a broad “typical” imagery group (VVIQ 32). This classification is the most used in the literature.
2. A 3-group VVIQ classification that additionally separates aphantasia and hypophantasia for more fine-grained groups, as described above.
3. A 3-cluster classification using the “visualiser”, “spatialiser” and “verbaliser” OSIVQ cognitive styles, based on participants’ most “dominant” sub-scale score. The clusters were identified through a consensus of several clustering algorithms.

The descriptive statistics of the three VVIQ sub-groups and three OSIVQ clusters are displayed in [Table 2](#tbl-vviq-groups) and [Table 3](#tbl-osivq-clusters), respectively. The visualiser cluster consisted almost exclusively of typical imagers (N = 42) along with one aphantasic who scored surprisingly high on the OSIVQ object scale (4.93). The verbaliser cluster consisted mainly of aphantasics (N = 25) and hypophantasics (N = 14) as well as three typical imagers. The spatialiser cluster was more balanced, with 12 typical imagers, three hypophantasics and four aphantasics. The distribution of the groups in the clusters (visualiser-typicals, verbaliser-aphantasics and spatialiser-mixed) is very similar to that observed by Delem et al. (2025).

Analyses of variance (ANOVA) were conducted on the age and RSPM-18 scores to control for potential influences of age differences or any differences in abstract reasoning abilities between the groups. There were no statistically significant age or RSPM-18 differences between the two VVIQ groups (age: *F*(1, 102) = 2.44, *p* = 0.12; RSPM-18: *F*(1, 102) = 0.12, *p* = 0.73), the three VVIQ sub-groups (age: *F*(2, 101) = 1.33, *p* = 0.27; RSPM-18: *F*(2, 101) = 0.8, *p* = 0.45) or the three OSIVQ clusters (age: *F*(2, 101) = 1.87, *p* = 0.16; RSPM-18: *F*(2, 101) = 0.06, *p* = 0.94).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Descriptive statistics of each VVIQ sub-group. The ‘N’ row indicates the group size along with the number of females (F) and other genders (O). For all the other variables, the values indicate the means and standard deviations (in parentheses).   |  | Aphantasia | Hypophantasia | Typical | | --- | --- | --- | --- | | N | 30 (23 F, 1 O) | 17 (11 F, 0 O) | 57 (45 F, 1 O) | | Age | 34.63 (12.38) | 36.35 (14.29) | 31.63 (10.67) | | VVIQ | 15.97 (0.18) | 23.76 (4.31) | 57 (12.89) | | OSIVQ-Object | 1.39 (0.71) | 1.42 (0.33) | 3.38 (0.85) | | OSIVQ-Spatial | 2.43 (0.68) | 2.72 (0.77) | 2.93 (0.77) | | OSIVQ-Verbal | 3.23 (0.93) | 3.33 (0.96) | 3.04 (0.88) | | RSPM-18 | 16.17 (1.7) | 15.35 (1.9) | 15.72 (2.49) | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Descriptive statistics of each OSIVQ cluster. The “N” row indicates the cluster size along with the number of females (F) and other genders (O). For all the other variables, the values indicate the means and standard deviations (in parentheses).   |  | Visualiser | Verbaliser | Spatialiser | | --- | --- | --- | --- | | N | 43 (36 F, 1 O) | 42 (29 F, 1 O) | 19 (14 F, 0 O) | | Age | 30.81 (10.13) | 34.26 (13.1) | 36.63 (12.01) | | VVIQ | 58.21 (13.86) | 21 (9.34) | 39.32 (19.46) | | OSIVQ-Object | 3.81 (0.56) | 1.32 (0.27) | 2.07 (0.59) | | OSIVQ-Spatial | 2.98 (0.72) | 2.37 (0.61) | 3.07 (0.87) | | OSIVQ-Verbal | 3.16 (0.84) | 3.47 (0.86) | 2.37 (0.7) | | RSPM-18 | 15.72 (2.16) | 15.88 (1.88) | 15.74 (2.94) | |

### 2.6.2 Accuracy

The means and distributions of the accuracy in the two VVIQ groups, the three VVIQ sub-groups and the three OSIVQ clusters are displayed in [Figure 3](#fig-acc).

#### 2.6.2.1 VVIQ 2 groups

There were no overall differences in accuracy between the two VVIQ groups (*p* = 0.3). The typical imagery group was more accurate in the control problems than the visual ones (odds ratio = 1.91, 95% CI = [1.04, 3.5], *p* = 0.034), but showed no difference between control and spatial (*p* = 0.52) or spatial and visual (*p* = 0.32). The aphantasia group did not show any differences in accuracy between the categories (all *p*-values 0.97). There was a trend interaction contrast between group and category for the control/visual difference (odds ratio = 0.57, 95% CI = [0.32, 1], *p* = 0.056), suggesting that this effect of category is markedly different for the two groups.

|  |
| --- |
| Figure 3: Average accuracy on the reasoning task. The opaque coloured dots represent the mean accuracies of each group or cluster on each category and the coloured bars represent the 95% CI of the mean. The semi-transparent dots represent the mean accuracies of each participant on each category to provide an overview of the sample distributions. In the leftmost plot, the “Aphantasia” group includes all participants with a VVIQ 32. In the middle plot, the “Aphantasia” group is restricted to VVIQ = 16 and participants with VVIQ [17, 32] constitute the “Hypophantasia” group. Black symbols indicate statistical significance. °: trend effect; \*: p .05. |

#### 2.6.2.2 VVIQ 3 groups

There were no overall differences in accuracy between the three VVIQ sub-groups (all *p*-values 0.65). The typical imagery group being the same as in the 2-group classification above, the difference between control and visual for this group is the same in the 3-group model. There were no other significant differences in accuracy between categories for any of the groups (all *p*-values 0.31). As opposed to the 2-group model, there were no significant interaction contrasts between groups and categories. This difference is likely due to the lack of statistical power conferred by the smaller sub-groups.

#### 2.6.2.3 OSIVQ 3 clusters

There were no overall differences in accuracy between the three OSIVQ clusters (all *p*-values 0.68). The only trend was a superior accuracy in the control than in the visual category for the visualiser cluster (Odds ratio = 1.86, 95% CI = [0.97, 3.55], *p* = 0.066), which echoes the effect observed in the typical group. There were no other significant contrast between categories for any of the clusters (all *p*-values 0.48) and no significant interaction contrasts (all *p*-values 0.17).

### 2.6.3 Response times

The means and distributions of total response times (RTs) for the two VVIQ groups, the three VVIQ sub-groups and the three OSIVQ clusters are displayed in [Figure 4](#fig-rt).

#### 2.6.3.1 VVIQ 2 groups

There were no overall RT differences between the two VVIQ groups (*p* = 0.6). However, the model revealed a visual impedance effect in the typical group, which is 2.4s slower in the visual category compared to the control or spatial categories ( spatial-visual = -2.45s, 95% CI = [-4.52, -0.38], *p* = 0.015; control-visual = -2.41s, 95% CI = [-4.27, -0.55], *p* = 0.007, spatial-control = 0.037s, 95% CI = [-1.83, 1.91], *p* = 0.99). The aphantasia group was also 1.6s slower in the visual category, although this effect was not statistically significant ( spatial-visual = -1.62s, 95% CI = [-3.97, 0.73], *p* = 0.24; control-visual = -1.69s, 95% CI = [-3.82, 0.45], *p* = 0.15, spatial-control = 0.066s, 95% CI = [-2.22, 2.09], *p* = 0.99). The 0.8s difference in the impedance effect between the groups was not statistically significant[[1]](#footnote-1), as shown by interaction contrasts ( aphantasia-typical for control-visual = 0.73s, 95% CI = [-0.63, 2.08], *p* = 0.29; for spatial-visual = 0.83s, 95% CI = [-0.94, 2.6], *p* = 0.36).

|  |
| --- |
| Figure 4: Total response times (RTs) on the reasoning task (time spent from the beginning of the first premise to the answer on the conclusion screen). The black shapes represent the mean RTs for each group or cluster and category and the black vertical bars represent the 95% CI of the mean. The semi-transparent coloured dots represent the mean RTs of each participant, and the coloured areas on the right represent the distribution of the RTs in each group or cluster. In the leftmost plot, the “Aphantasia” group includes all participants with a VVIQ 32. In the middle plot, the “Aphantasia” group is restricted to VVIQ = 16 and participants with VVIQ [17, 32] constitute the “Hypophantasia” group. Black stars above the horizontal lines indicate statistical significance. \*: p .05; \*: p\* .01; \*\*\*: p .001. |

#### 2.6.3.2 VVIQ 3 groups

There were no overall RT differences between the three VVIQ sub-groups (all *p*-values 0.19). The typical imagery group being the same as in the 2-group classification above, the significant visual impedance effect for this group was the same in the 3-group model. Dividing full aphantasia and hypophantasia revealed different trends of visual impedance effect in the two sub-groups. The hypophantasia sub-group was estimated to be 2.2s slower in the visual category ( spatial-visual = -2.23s, *p* = 0.23; control-visual = -2.29s, *p* = 0.12), whereas the full aphantasia group was estimated to be only 1.3s slower ( spatial-visual = -1.28s, *p* = 0.53; control-visual = -1.35s, *p* = 0.42), although those differences were not significant. As with the 2-group model, the impedance effect differences between sub-groups were not reflected in interaction contrasts, none of which were significant (all *p*-values 0.19).

#### 2.6.3.3 OSIVQ 3 clusters

There were no overall RT differences between the three OSIVQ clusters (all *p*-values 0.73). The model revealed a large visual impedance effect in the visualiser cluster, which was much slower in the visual category than in the other two categories ( spatial-visual = -2.62s, 95% CI = [-4.8, -0.44], *p* = 0.013; control-visual = -2.89s, 95% CI = [-4.8, -0.98], *p* = 0.001, spatial-control = 0.27s, 95% CI = [-2.22, 1.67], *p* = 0.95). The spatialiser and verbaliser clusters showed weaker, non-significant impedance effects (spatialisers: spatial-visual = -1.57s, control-visual = -1.37s; verbalisers: spatial-visual = -1.7s, control-visual = -1.54s; all *p*-values 0.23). The 1.4s difference in impedance effect between the visualiser cluster and the other two was not large enough to be detected by the model, as no interaction contrast reached statistical significance (all *p*-values 0.09).

### 2.6.4 Strategies

The mean scores on the strategy use questionnaire for the two VVIQ groups, the three VVIQ sub-groups and the three OSIVQ clusters are displayed in [Figure 5](#fig-strat_scores). The proportion of answers for each level of the Likert scale are displayed in [Figure 6](#fig-strat_proportions).

#### 2.6.4.1 VVIQ 2 groups

The only difference in reported strategies between the two VVIQ groups was in the frequency of visual strategy use, the typical group using more visual strategies than the aphantasia group ( aphantasia-typical = -1.69, 95% CI = [-2.18, -1.2], *p* 0.001). There was a trend difference between the two groups in the use of spatial strategies ( aphantasia-typical = -0.39, 95% CI = [-0.82, 0.03], *p* = 0.068), but no differences in verbal (*p* = 0.203), semantic (*p* = 0.379) or sensorimotor (*p* = 0.531) strategies.

|  |
| --- |
| Figure 5: Average scores on the strategy questionnaire per strategy and group or cluster. The opaque coloured dots represent the mean scores on the Likert scales and the coloured bars represent the 95% CI of the means. The horizontal curves connect the group means for each strategy to highlight trends in the differences (or lack thereof) between groups. The semi-transparent dots represent the score of each participant for each strategy to provide an overview of the distributions. In the leftmost plot, the “Aphantasia” group includes all participants with a VVIQ 32. In the middle plot, the “Aphantasia” group is restricted to VVIQ = 16 and participants with VVIQ [17, 32] constitute the “Hypophantasia” group. Stars indicate statistical significance. \*: p .05; \*: p\* .01; \*\*\*: p .001. For ease of understanding, the stars and lines have a different appearance for each strategy. Blue stars above solid lines represent differences in visual strategy scores. The green star above a dashed line represent differences in spatial strategy scores. There were no group differences for other strategies. |

#### 2.6.4.2 VVIQ 3 groups

Likewise, the major differences in reported strategies between the three VVIQ sub-groups was in the visual strategy: the full aphantasia group used fewer visual strategies than the hypophantasia group ( aphantasia-hypo = -1.6, 95% CI = [-2.84, -0.37], *p* = 0.006) and the typical group ( aphantasia-typical = -2.64, 95% CI = [-3.74, -1.54], *p* 0.001), while the hypophantasia group used fewer visual strategies than the typical group ( hypo-typical = -1.04, 95% CI = [-1.77, -0.3], *p* = 0.003). The hypophantasia group also used fewer spatial strategies than the typical group ( hypo-typical = -0.79, 95% CI = [-1.55, -0.04], *p* = 0.036). There were no significant differences between the three sub-groups on the other strategies (all *p*-values 0.2).

|  |
| --- |
| Figure 6: Detail by strategy of the proportion of participants from each group or cluster who chose each of the five options on the Likert scales. In the leftmost column, the “Aphantasia” group includes all participants with a VVIQ 32. In the middle column, the “Aphantasia” group is restricted to VVIQ = 16 and participants with VVIQ [17, 32] constitute the “Hypophantasia” group. |

#### 2.6.4.3 OSIVQ 3 clusters

In line with the definition of the visualiser cluster, the only difference in strategy use between the three OSIVQ clusters was a higher use of visual strategies by the visualiser group compared to the other two ( visualiser-spatialiser = 0.81, 95% CI = [0.09, 1.53], *p* = 0.024, visualiser-verbaliser = 1.52, 95% CI = [0.91, 2.13], *p* 0.001, spatialiser-verbaliser = 0.71, 95% CI = [-0.07, 1.49], *p* = 0.083). There were no significant differences between the three clusters on the other strategies (all *p*-values 0.19).

### 2.6.5 Response times per trial phase (non-linear modelling)

The mean RTs per category and trial phase and a representation of the non-linear models for the two VVIQ groups, the three VVIQ sub-groups and the three OSIVQ clusters are displayed in [Figure 7](#fig-nl).

#### 2.6.5.1 VVIQ 2 groups

The non-linear model fitted with the two VVIQ groups revealed similar dynamics between the groups. In the second premise, participants were faster in the control than in the spatial category, significantly for the typical group ( = -1.18s, *t* = -3.07, *p* = 0.002) and trendily for the aphantasia group ( = -0.68s, *t* = -1.71, *p* = 0.088). In the third premise, the aphantasia group showed a trend toward slower RTs in the visual than in the control category ( = -0.77s, *t* = -1.71, *p* = 0.088) and significantly slower RTs in the visual than in the spatial category ( = -1.33s, *t* = -3.01, *p* = 0.003), while the typical group showed slower RTs in the control than in the spatial category ( = 0.85s, *t* = 2.1, *p* = 0.035) and slower RTs in the visual than in the spatial category ( = -1.44s, *t* = -3.4, p = 0.001). No other differences between categories were significant for either group, at any phase of the trial (all *p* 0.11).

#### 2.6.5.2 VVIQ 3 groups

As the typical group was the same in the 3-group VVIQ model, the effects for this group were identical to the 2-group VVIQ model. Dividing the full aphantasia and hypophantasia sub-groups showed that the difference between control and spatial in the second premise was specific to the full aphantasia group ( control-spatial = -0.98s, *t* = -2.02, *p* = 0.04) and was absent in the hypophantasia group ( control-spatial = -0.33s, *t* = -0.55, *p* = 0.58). The slower RTs in the visual than in the spatial category were observed in both groups (full aphantasia: = -1.27s, *t* = -2.31, *p* = 0.021, hypophantasia: = -1.43s, *t* = -2.2, *p* = 0.028). No other differences between categories were significant for any of the three groups, at any phase of the trial (all *p* 0.14).

#### 2.6.5.3 OSIVQ 3 clusters

The non-linear model fitted with the three OSIVQ clusters showed trendily slower RTs in the visual than in the control category in the first premise, only for the visualiser cluster ( control-visual = -0.99s, *t* = -1.77, *p* = 0.077). In the second premise, visualisers and verbalisers showed slower RTs in the spatial than in the control category (visualisers: = -1.29s, *t* = -3, *p* = 0.003, verbalisers: = -0.79s, *t* = -1.86, *p* = 0.063). In the third premise, visualisers and verbalisers were slower in the visual than in the spatial category (visualisers: = -1.7s, *t* = -3.57, *p* 0.001, verbalisers: = -1.32s, *t* = -2.83, *p* = 0.005), while visualisers were also faster in the spatial than in the control category ( = -1s, *t* = -2.21, *p* = 0.027). No differences between categories were significant for the spatialiser cluster, and no other differences were significant at any phase of the trial for any of the three clusters (all *p* 0.1).

|  |
| --- |
| Figure 7: Average response times (RTs) per trial phase. The coloured dots represent the mean RTs for each group, category and trial phase, and the coloured bars represent the 95% CI of the mean. The smooth coloured curves represent the non-linear models fitted on the RT data across trial phases. In the top row, the “Aphantasia” group includes all participants with a VVIQ 32. In the middle row, the “Aphantasia” group is restricted to VVIQ = 16 and participants with VVIQ [17, 32] constitute the “Hypophantasia” group. Black symbols above horizontal lines indicate statistical significance. °: trend effect; \*: p .05; \*: p\* .01; \*\*\*: p .001. |

Besides statistical tests, the most interesting insights from the non-linear models appears upon visual examination of the smooth term dynamics. [Figure 7](#fig-nl) shows what underlies the contrasts above: the most striking difference between problem categories lies less in the RT differences than in the distribution of the time spent on the different phases of a problem. People consistently spent similar amounts of time on each premise and conclusion for spatial problems, while they showed a “wave-like” pattern for the visual and control categories, being slow on premise 1, fast on 2, slow on 3, and fast for answering on the conclusion screen. These exploratory analyses highlight that there may be fundamental differences in the dynamics of reasoning processes depending on the category of relationships, differences that are hidden when analysing total response times. However, the VIIE is scattered throughout the trials when dissecting their different phases, and therefore seems to be an overall slowing effect that only appears when considering the total time spent on the problems.

## 2.7 Discussion

In this study, we sought to test the VIIH by administering a reasoning task to aphantasics, who, according to Knauff and Johnson-Laird’s hypothesis (2002), should be immune to the negative impact of visual mental imagery. Using an initial classification based on the VVIQ, we confirmed the presence of the VIIE in control participants, with visual problems eliciting significantly longer RTs, compared to spatial and control problems. In aphantasics, this effect was less marked, but the difference between the two groups is not substantial enough to be conclusively established. In terms of accuracy, no significant differences were observed between groups.

In recent studies, it has been suggested to distinguish between “complete aphantasics” (VVIQ = 16) and “hypophantasics” (VVIQ [17, 32]), as they appear to function differently Reeder et al. (2024). To provide clarification on these points from our own data, we performed an in-depth analysis based on finer VVIQ groups, dissociating these two subgroups. These analyses revealed different trends of VIIE between aphantasics and hypophantasics, with a slowdown in visual problems of respectively 1.3s and 2.2s (while typical imagers show a VIIE of 2.4s). Although no interaction effects appeared significant through our analyses, the estimates of our models remain reliable and enable us to suggest that differences between these subgroups indeed exist. Thus, these results suggest that the slowdown initially observed in the broader aphantasic group (obtained from the first VVIQ classification) is mainly caused by hypophantasics. These results support the findings of Purkart et al. (2025), which suggest that hypophantasics may experience some unconscious imagery, with a reduced ability to voluntarily generate conscious mental images, while complete aphantasics are characterised by a complete absence of visual images. Consequently, our results support the need to make a clear distinction between these two groups and to admit the heterogeneity of aphantasia (Schwarzkopf, 2024). Although weak, the imagery experienced by hypophantasics seems to influence their performance - at least in such a task - and represents a factor that deserves to be recognized when studying this condition.

Concerning OSIVQ classification analyses, no significant differences were observed between clusters in accuracy. Contrast analyzes on RTs, for their part, revealed that only the visualiser cluster demonstrated the VIIE. In other words, for participants with a visual cognitive style, visual problems caused longer response times than spatial or control problems. Furthermore, as in Delem et al. (2025), the clustering method used for the OSIVQ classification demonstrates the heterogeneity existing among aphantasics - who are divided into two subgroups (verbalisers and spatialisers) - while also indicating that, even in typical imagers, differences in information processing styles can lead to differences in performances.

Our results add nuances to the conclusions of Gazzo Castaneda & Knauff (2013). Contrary to their results, visualisers did not show VIIE for all problems, but only for visual problems. Moreover, verbalisers do not seem immune to this effect. In our study, spatialisers and verbalisers show a pattern of results that resembles the VIIE. Our statistical models estimate a VIIE of 1.4s in spatialisers and of 1.6s in verbalisers. However, due to an insufficient number of participants (for spatialisers) or too small an effect size (for verbalisers), these effects are not detectable, as our power analyses demonstrate (see Figure 2). Therefore, our study may suggest that visual problems negatively impact all individuals, but the magnitude of this effect varies depending on their cognitive style. Indeed, it seems significantly easier to detect a VIIE among visualisers, hence the importance of considering the influence of cognitive styles in the study of this effect.

Focusing now on a different aspect, one of the major contributions of our study is the addition of a temporal dimension to the investigation of the VIIE. Each problem has been divided by term (P1, P2, P3 and conclusion), in order to better understand the dynamic of the VIIE. This exploratory analysis provides new information and enables us to go beyond the analysis of total RTs, by showing that the slowdown observed for visual problems is not present at every term but seems mainly induced by processes occurring during the reading of the third premise. Moreover, spatial problems appear relatively stable across the different terms, whereas visual problems are inconsistent, with relatively fast processing in P2 followed by a significant slowdown in P3. Although spatial problems are the longest to be processed in P2, the slowdown in visual problems in P3 is greater and causes the difference observed in total RTs.

According to mental model theory, integrated mental representations are constructed from the information given in the premises of a reasoning problem Krumnack et al. (2011). During P1, participants might focus on constructing a representation that helps them memorize this premise, followed by the relatively simple and rapid addition of the new information presented in P2. Our analyses suggest little influence of the category during these phases. During P3, the participants’ aim is to integrate the new information contained in this premise into the mental representation constructed so far. The slowdown observed in P3 for visual problems could correspond precisely to the recall of this representation, during which these irrelevant visual details appear until the correct integration of this new information. As proposed by Fangmeier et al. (2006) and their results from fMRI studies, visual mental imagery would therefore be more a strategy for memory retention than a necessary process for reasoning. In their study as well as in ours, this slowdown is not observed during the conclusion, the precise stage during which reasoning processes occur. These authors argue that the latter rely on more abstract spatial representations, in accordance with mental model theory (Knauff & May, 2006). Our analyses demonstrate the limitation of the total RT analyses, which prevent us from considering the dynamic of the VIIE. In addition, although studies have limited themselves to the analysis of the conclusion Gazzo Castaneda & Knauff (2013), our results clearly establish the inadequacy of this method, as the differences between categories are very small during the processing of this final term. These term-based analyses appear to be the most appropriate for the VIIE study, as they account for the temporal dynamics of the effect.

Furthermore, one of the main goals of this study was to investigate the strategies used to solve this reasoning task, specifically taking inspiration from the work of Reeder et al. (2024). From the second VVIQ classification, which distinguishes complete aphantasia and hypophantasia, analyses of participants’ responses to the questionnaire evaluating the use of five resolution strategies revealed differences in the use of the visual strategy. Complete aphantasics used it less than hypophantasics, who themselves used it less than typical imagers; which is consistent with the different levels of visual imagery across groups. Moreover, all individuals show the same pattern for the remaining strategies: the verbal strategy is the most employed, followed by spatial and sensorimotor strategies, while the semantic strategy is scarcely used. This pattern is probably due to the verbal nature of the task. Concerning OSIVQ classification, a single difference was observed between clusters: only visualisers rely on the visual strategy. Verbalisers and spatialisers shared similar patterns, with a dominant use of verbal strategy followed by a combination of spatial and sensorimotor strategies; whereas visualisers used a combination of verbal, visual and spatial strategies followed by the sensorimotor strategy. Thus, considering only the responses to this questionnaire, only the use of visual strategies explain the differences in performance in the reasoning task between the clusters, which is consistent with the VIIH.

Moreover, all participants - regardless of their visual imagery capacity or cognitive style - used the verbal strategy more frequently. This result is consistent with studies using memorization tasks in which aphantasics indicated preferences for propositional strategies Monzel et al. (2024), and further corroborates findings suggesting that non-visual individuals rely more frequently on such strategies Pearson & Keogh (2019). However, our overall results also show that aphantasics and non-aphantasics use spatial and sensorimotor strategies for reasoning, which underline the importance of varying the nature of the strategies investigated in this type of questionnaire. These small adjustments allow us to go beyond the classic conception opposing visualisers to verbalisers (Richardson, 1977), and bring out a more accurate representation of inner experiences.

Finally, this preference for verbal strategy questions the postulate of mental model theory, according to which reasoning is underpinned by spatial processes. However, our questionnaire did not distinguish the phase during which each strategy was used. It is conceivable that participants, according to their cognitive style, use verbal, visual or other strategies to retain information in working memory, whereas it is in fact spatial processes that are involved during the actual reasoning phase. It is also plausible that combinations of these representations coexist and underlie the reasoning. In addition, the limits of such questionnaire are fully recognized, as they remain self-reported: participants may report the use of certain strategies while their information processing may be completely distinct. Furthermore, it has been suggested that verbal processes, by participating in encoding, work to construct a reliable mental representation and are therefore crucial for the outcome of reasoning mechanisms (Krumnack et al., 2011). In our study, the fact that aphantasics did not rely on visual strategies may suggest a distinct reasoning style, potentially offering an advantage in this type of task. These results could partially explain the over representation of aphantasics in STEM observed by Zeman et al. (2020). Future studies will need to be conducted to determine whether aphantastics could actually benefit from reasoning processes, which would lead them to pursue studies and professions that require these processes.

Although our study has shown promising results, several limitations must be acknowledged. While a VIIE is observed in visual participants, our statistical models also estimated a slowdown for visual problems in remaining participants, albeit not significant ( 1.6s in the verbaliser cluster; and 1.3s within the “complete aphantasia” group and the spatialiser cluster). If such findings are confirmed, it could suggest that visual problems negatively affect all individuals to varying degrees, depending on cognitive style; or that even complete aphantasics experiment a type of visual imagery (or at least try to do so), perhaps unconscious, that affects their performance. Moreover, regardless of the classification used, the estimates of our statistical models agree that, if a difference in VIIE exists between the groups/clusters, this difference is 1s (VVIQ’s VIIE : 2.4s for typical imagers and 2.2s for hypophantasics versus 1.3s for aphantasics ; OSIVQ’s VIIE : 2.7s for visualisers versus 1.6s for verbalisers and 1.4s for spatialisers). Our power analysis clearly demonstrate why these differences are not substantial enough to lead to significant interaction effects: an effect of 1s requires more than 160 participants per group/cluster to be detected (see Figure 2). Thus, to better capture differences between such groups/clusters, the recruitment of our study could be extended to an English-speaking population to provide solid evidence on the VIIH. Furthermore, an alternative explanation for the VIIE observed in complete aphantasics may be that some individuals who experience visual imagery, even in its weakest form, have been misclassified as complete aphantasic rather than hypophantasic. This last point raises a criticism commonly reported in mental imagery studies, namely the difficulty of forming groups based on questionnaires assessing participants’ conscious processes alone. Future research should focus on developing tools that measure visual imagery capabilities without relying on individual subjectivity, as exemplified by the work of Purkart et al. (2025).

In conclusion, our study is the first to test the VIIE in aphantasia. While the differences between typical imagers and our initial group of aphantasics were inconclusive, more in-depth analyses revealed differences between “complete aphantasics” and hypophantasics. Although weak, this form of visual imagery negatively influences participants’ performances, which reinforces the need to differentiate these two subgroups and consider aphantasia as a heterogeneous condition. In addition, our OSIVQ classification-based analysis, which notably replicates the findings of Delem et al. (2025), along with the investigation of strategies employed to solve these problems, highlights the need to acknowledge the influence of cognitive styles on such reasoning tasks. This inter individual variability must be accounted for in order to formulate meaningful conclusions on human reasoning, without omitting a significant proportion of individuals. Finally, our study is also the first to describe the temporal dynamics of the VIIE. This method seems to be the most appropriate for understanding the impact of visual imagery on such reasoning tasks. Aphantasia reflects the incredible diversity of information processing and representation styles that exist within the human population, recalling that each cognitive style is associated with its own set of strengths and limitations. Thus, the study of this condition should not only focus on its deficits but also on its potential benefits.

## Research Transparency Statement

All the following elements required to reproduce the study and analyses are publicly available on the Open Science Framework (<https://osf.io/hfbcp/?view_only=0ff6ba4fba3e46b281de0c3cc4f94a55>): all online study materials; all anonymised primary data; all analysis code and supplementary information on the analysis process and results. No artificial intelligence assisted technologies were used in this research or the creation of this article.

## Author Contributions

Conceptualisation: MD, DLC, GP. Data curation: MD. Formal analysis: MD. Funding acquisition: GP. Investigation: MD, DLC. Methodology: MD, DLC, GP. Project administration: GP. Resources: MD, DLC, GP. Software: MD. Supervision: GP. Visualisation: MD. Writing - Original Draft Preparation: DLC. Writing - Review & Editing: MD, MM, GP.

## Acknowledgements

This work was supported by a grant (ANR-24-CE28-5187; APHANTASIA) from the French National Research Agency (ANR).

## Declaration of Interests

None.

# References

Albrecht, R., Schultheis, H., & Wai-Tat, F. (2015). (PDF) Visuo-Spatial Memory Processing and the Visual Impedance Effect. In *ResearchGate*. https://www.researchgate.net/publication/323551760\_Visuo-Spatial\_Memory\_Processing\_and\_the\_Visual\_Impedance\_Effect.

Ando, R., Morishita, T., Abe, H., Mineshima, K., & Okada, M. (2023). *Evaluating Large Language Models with NeuBAROCO: Syllogistic Reasoning Ability and Human-like Biases* (arXiv:2306.12567). arXiv. <https://doi.org/10.48550/arXiv.2306.12567>

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; a Journal Devoted to the Study of the Nervous System and Behavior*, *135*, 159–172. <https://doi.org/10.1016/j.cortex.2020.11.014>

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven’s standard progressive matrices test. *Assessment*, *19*(3), 354–369. <https://doi.org/10.1177/1073191112446655>

Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, *23*(5), 638–663. <https://doi.org/10.1002/acp.1473>

Blazhenkova, O., & Kozhevnikov, M. (2010). Visual-object ability: A new dimension of non-verbal intelligence. *Cognition*, *117*(3), 276–301. <https://doi.org/10.1016/j.cognition.2010.08.021>

Chiu, D., & Talhouk, A. (2025). *diceR: Diverse cluster ensemble in r*. <https://github.com/AlineTalhouk/diceR/>

Christensen, R. H. B. (2023). *Ordinal—regression models for ordinal data*. <https://CRAN.R-project.org/package=ordinal>

Cortes, R., Weinberger, A. B., Colaizzi, G., & Porter, G. (2021). (PDF) What Makes Mental Modeling Difficult? Normative Data for the Multidimensional Relational Reasoning Task. *ResearchGate*. <https://doi.org/10.3389/fpsyg.2021.668256>

Dance, C. J., Ipser, A., & Simner, J. (2022). The prevalence of aphantasia (imagery weakness) in the general population. *Consciousness and Cognition*, *97*, 103243. <https://doi.org/10.1016/j.concog.2021.103243>

De Soto, C. B., London, M., & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, *2*(4), 513–521. <https://doi.org/10.1037/h0022492>

Delem, M. (2025). *aphantasiaReasoningViie: Data analysis for "The Impact of Mental Images on Reasoning: A Study on Aphantasia"*. <https://m-delem.github.io/aphantasiaReasoningViie/>

Delem, M., Turkben, S., Cousineau, D., Cavalli, E., & Plancher, G. (2025). [*Uncovering spatial and verbal cognitive profiles in aphantasia through unsupervised clustering*](https://www.ncbi.nlm.nih.gov/pubmed/32446532).

Engel, P. (1994). *Introduction à la philosophie de l’esprit*. La Découverte. <https://doi.org/10.3917/dec.engel.1994.01>

Fangmeier, T., Knauff, M., Ruff, C. C., & Sloutsky, V. (2006). fMRI evidence for a three-stage model of deductive reasoning. *Journal of Cognitive Neuroscience*, *18*(3), 320–334. <https://doi.org/10.1162/jocn.2006.18.3.320>

Gazzo Castaneda, L. E., & Knauff, M. (2013). Individual differences, imagery and the visual impedance effect. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*.

Jacobs, C., Schwarzkopf, D. S., & Silvanto, J. (2018). Visual working memory performance in aphantasia. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *105*, 61–73. <https://doi.org/10.1016/j.cortex.2017.10.014>

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, *5*(10), 434–442. <https://doi.org/10.1016/S1364-6613(00)01751-4>

Johnson-Laird, P. N. (2006). *How we reason* (pp. x, 573). Oxford University Press.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, *107*(43), 18243–18250. <https://doi.org/10.1073/pnas.1012933107>

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction* (pp. xii, 243). Lawrence Erlbaum Associates, Inc.

Keogh, R., Wicken, M., & Pearson, J. (2021). Visual working memory in aphantasia: Retained accuracy and capacity with a different strategy. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *143*, 237–253. <https://doi.org/10.1016/j.cortex.2021.07.012>

Knauff, M., Fangmeier, T., Ruff, C. C., & Johnson-Laird, P. N. (2003). Reasoning, models, and images: Behavioral measures and cortical activity. *Journal of Cognitive Neuroscience*, *15*(4), 559–573. <https://doi.org/10.1162/089892903321662949>

Knauff, M., & Johnson-Laird, P. N. (2002). Visual imagery can impede reasoning. *Memory & Cognition*, *30*(3), 363–371. <https://doi.org/10.3758/BF03194937>

Knauff, M., & May, E. (2006). Mental imagery, reasoning, and blindness. *The Quarterly Journal of Experimental Psychology*, *59*(1), 161–177. <https://doi.org/10.1080/17470210500149992>

Knight, K. F., Milton, F., & Zeman, A. Z. J. (2022). Memory without Imagery: No Evidence of Visual Working Memory Impairment in People with Aphantasia. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).

Kosslyn, S. M. (1994). *Image And Brain: The Resolution of the Imagery Debate*. The MIT Press. <https://doi.org/10.7551/mitpress/3653.001.0001>

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

Kozhevnikov, M., Blazhenkova, O., & Becker, M. (2010). Trade-off in object versus spatial visualization abilities: Restriction in the development of visual-processing resources. *Psychonomic Bulletin & Review*, *17*(1), 29–35. <https://doi.org/10.3758/PBR.17.1.29>

Kozhevnikov, M., Kosslyn, S., & Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition*, *33*(4), 710–726. <https://doi.org/10.3758/BF03195337>

Krumnack, A., Bucher, L., Nejasmic, J., Nebel, B., & Knauff, M. (2011). A model for relational reasoning as verbal reasoning. *Cognitive Systems Research*, *12*(3-4), 377–392. <https://doi.org/10.1016/j.cogsys.2010.11.001>

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies” (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLOS ONE*, *10*(6), e0130834. <https://doi.org/10.1371/journal.pone.0130834>

Leeuw, J. R. de, Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351. <https://doi.org/10.21105/joss.05351>

Liu, J., & Bartolomeo, P. (2023). Probing the unimaginable: The impact of aphantasia on distinct domains of visual mental imagery and visual perception. *Cortex*, *166*, 338–347. <https://doi.org/10.1016/j.cortex.2023.06.003>

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*(1), 17–24. <https://doi.org/10.1111/j.2044-8295.1973.tb01322.x>

McGillycuddy, M., Warton, D. I., Popovic, G., & Bolker, B. M. (2025). Parsimoniously fitting large multivariate random effects in glmmTMB. *Journal of Statistical Software*, *112*(1), 1–19. <https://doi.org/10.18637/jss.v112.i01>

McKelvie, S. J. (1995). The VVIQ as a psychometric test of individual differences in visual imagery vividness: A critical quantitative review and plea for direction. *Journal of Mental Imagery*, *19*(3-4), 1–106.

Monzel, M., Handlogten, J., & Reuter, M. (2024). No verbal overshadowing in aphantasia: The role of visual imagery for the verbal overshadowing effect. *Cognition*, *245*, 105732. <https://doi.org/10.1016/j.cognition.2024.105732>

Monzel, M., & Reuter, M. (2024). Where’s wanda? The influence of visual imagery vividness on visual search speed measured by means of hidden object pictures. *Attention, Perception, & Psychophysics*, *86*(1), 22–27. <https://doi.org/10.3758/s13414-022-02645-6>

Monzel, M., Vetterlein, A., & Reuter, M. (2023). No general pathological significance of aphantasia: An evaluation based on criteria for mental disorders. *Scandinavian Journal of Psychology*, *64*(3), 314–324. <https://doi.org/10.1111/sjop.12887>

Palermo, L., Boccia, M., Piccardi, L., & Nori, R. (2022). Congenital lack and extraordinary ability in object and spatial imagery: An investigation on sub-types of aphantasia and hyperphantasia. *Consciousness and Cognition*, *103*, 103360. <https://doi.org/10.1016/j.concog.2022.103360>

Palmiero, M., Fusi, G., Crepaldi, M., Borsa, V. M., & Rusconi, M. L. (2022). Divergent thinking and the core executive functions: A state-of-the-art review. *Cognitive Processing*, *23*(3), 341–366. <https://doi.org/10.1007/s10339-022-01091-4>

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews. Neuroscience*, *20*(10), 624–634. <https://doi.org/10.1038/s41583-019-0202-9>

Pearson, J., & Keogh, R. (2019). Redefining visual working memory: A cognitive-strategy, brain-region approach. *Current Directions in Psychological Science*, *28*(3), 266–273. <https://doi.org/10.1177/0963721419835210>

Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods*, *23*(2), 208–225. <https://doi.org/10.1037/met0000126>

Posit team. (2025). *RStudio: Integrated development environment for r*. Posit Software, PBC. <http://www.posit.co/>

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; a Journal Devoted to the Study of the Nervous System and Behavior*, *148*, 180–192. <https://doi.org/10.1016/j.cortex.2021.12.010>

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>

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

Reeder, R. R., Pounder, Z., Figueroa, A., Jüllig, A., & Azañón, E. (2024). Non-visual spatial strategies are effective for maintaining precise information in visual working memory. *Cognition*, *251*, 105907. <https://doi.org/10.1016/j.cognition.2024.105907>

Richardson, A. (1977). Verbalizer-visualizer: A cognitive style dimension. *Journal of Mental Imagery*, *1*(1), 109–125.

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24*(3), 309–338. <https://doi.org/10.1037/met0000184>

Schwarzkopf, D. S. (2024). What is the true range of mental imagery? *Cortex*, *170*, 21–25. <https://doi.org/10.1016/j.cortex.2023.09.013>

Sima, J. F., Schultheis, H., & Barkowsky, T. (2013). Differences between Spatial and Visual Mental Representations. *Frontiers in Psychology*, *4*. <https://doi.org/10.3389/fpsyg.2013.00240>

Thevenot, C., & Perret, P. (2009). Le développement du raisonnement dans la résolution de problèmes : l’apport de la théorie des modèles mentaux. *Développements*, *2*(2), 49–56. <https://doi.org/10.3917/devel.002.0049>

Tse, A., Ragni, M., & Lösch, J. (2017). *Beyond the Visual Impedance Effect*.

Wechsler, D., Coalson, D. L., & Raiford, S. E. (2008). *Wechsler adult intelligence scale (technical and interpretive manual 4th ed.). San antonio: NCS pearson*. Inc.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>

Wright, D. J., Scott, M. W., Kraeutner, S. N., Barhoun, P., Bertollo, M., Campbell, M. J., Waltzing, B. M., Dahm, S. F., Esselaar, M., Frank, C., Hardwick, R. M., Fuelscher, I., Marshall, B., Hodges, N. J., Hyde, C., & Holmes, P. S. (2024). An international estimate of the prevalence of differing visual imagery abilities. *Frontiers in Psychology*, *15*. <https://doi.org/10.3389/fpsyg.2024.1454107>

Zeman, A., Milton, F., Della Sala, S., Dewar, M., Frayling, T., Gaddum, J., Hattersley, A., Heuerman-Williamson, B., Jones, K., MacKisack, M., & Winlove, C. (2020). Phantasia-The psychological significance of lifelong visual imagery vividness extremes. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *130*, 426–440. <https://doi.org/10.1016/j.cortex.2020.04.003>

1. This can also be explained by statistical power: if the “true” interaction between group and category was 0.8s, simulations show that 50-participant groups could only detect it in 20% of cases. To detect such a small interaction reliably, more than 200 participants per group would be required. [↑](#footnote-ref-1)