Clarifying aphantasia

by

Alexander R. Toftness

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Psychology

Program of Study Committee:
Eric E. Cooper, Co-major Professor
Jonathan W. Kelly, Co-major Professor
Christian A. Meissner
L. Alison Phillips
John G. Grundy

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2022

Copyright © Alexander R. Toftness, 2022. All rights reserved.

DEDICATION

Dedicated to my incredible partner, Vanessa Castillo, and the loudest cat, Sabrina.

Aphantasia (a poem by Alexander R. Toftness)

I can’t count your freckles unless you’re near
And I can’t hear your voice ‘til it’s in my ear
I recognize your face in much the same fashion
that a puppy gets excited when the front door unfastens

And I’m sure that your eyes are a cavalcade of color
But in my mind’s eye they are unfortunately duller
Represented without image but by the lonely word ‘brown’
I know that don’t do them justice, don’t cast them down

I’m not saying you’re not on my mind
I’m thinking about you all of the time
But to see you, see,
I need you
Right here
Next to me

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my committee chairs, Eric Cooper and Jon Kelly, and my committee members, Chris Meissner, Alison Phillips, and John Grundy for their guidance and support throughout the course of this research. I would like to thank Taylor Doty, Alex Toma, and Jack Nichols for commenting on the pilot version of the tasks used in this series of studies. In addition, I would also like to thank my friends, colleagues, and the department faculty and staff at Iowa State University. I also want to express my appreciation to those who participated in my studies, without whom, this dissertation would not have been possible.

I want to especially thank Christina M. Meyer, Logan Toftness, and Vanessa Castillo for listening to me chatter incessantly about this project and supporting me throughout the process. Big shoutout to my parents, family, and friends out in the rolling hills of Wisconsin and beyond. Finally, thank you to the online communities that I call home, especially WeCreateEdu and FunkyTown – y’all are great.

ABSTRACT

In a series of four studies, relationships between subjective and objective measurements of mental imagery vividness and use were examined for participants with developmental aphantasia compared to control participants. Study 1 reports demographic differences and results from subjective measurements of mental imagery. Subjective scores of control participants were impacted by an instructional video manipulation, showing that control participants may be naïve about individual differences in mental imagery. Study 2 examined the relationship between a conventionally used measure of mental imagery, the VVIQ, and several cognitive tasks previously reported by the literature to be related to mental imagery vividness: the backwards spelling task, the snowy pictures task, and the tail length task. Study 2 found that none of those three tasks seem appropriate for research into developmental aphantasia due to no detectable group differences in performance after accounting for both accuracy and response times, and no reliable relationships with VVIQ scores. Study 3 makes use of an established imagery paradigm, mental scanning, and applies the square donut scanning task in a novel way to developmental aphantasia research. Study 3 demonstrates an important interaction between the difficulty of the task and group (aphantasia vs. control) of the participants, revealing an objective cognitive difference between the groups. Study 4 uses a change identification task to also demonstrate that there is a significant difference in the cognitive strategy used by the aphantasia group relative to the controls, made evident by significant interactions between trial complexity and group as measured by trial performance.

CHAPTER 1. GENERAL INTRODUCTION

"Of all the controversies currently raging in philosophy and psychology, none is being conducted with more vigor—if not rigor—than the debate over the nature of, and even the very existence of, mental images"

—Dennett, 1978, p. 174

Mental Imagery

At first, mental imagery seems to have a straightforward definition. Colloquially, when you imagine a sight, sound, smell, or so on, and feel as if you are experiencing it, you are experiencing mental imagery. More technically, a mental image is when you have the subjective perceptual experience of some stimulus when that stimulus is not being objectively perceived (Kosslyn et al., 2006, p. 5). A person who is effectively using mental imagery can voluntarily generate modality-specific representations of various stimuli with previously experienced features or with a novel combination of features such that they subjectively report the perceptual experience of those stimuli (J. Pearson et al., 2015, p. 590). To most people, this experience is familiar enough that explaining it feels unnecessary. And yet, similar to many concepts in psychology, this seemingly simple experience balances atop a complex and not-yet-well-understood collection of mental processes.

To further complicate things, it is becoming increasingly apparent that the use of—or access to—mental processes involved in mental imagery is variable from person to person. The people of most interest for this project are those who claim *not* to experience voluntarily-generated visual mental images—the condition of *aphantasia*. I will attempt to clarify and extend the evidence that aphantasia exists. But, to show that voluntary generation of mental images does not exist in the minds of some people, I first need to go on a journey through the evidence that mental images exist in the first place. Mental imagery is an almost inscrutable subjective experience that is unique to one's own self—attempting to understand the individual differences of how another person's imagination landscape may differ from your own appreciably leads to wonder, if not outright doubt. The aim is to satisfy both the people who doubt that some people lack mental imagery, and to satisfy those who believe the exact opposite—that nobody experiences true mental images.

As will be discussed, individual differences in vividness of visual mental imagery has recently been a hot topic, especially thanks to a paper published in 2015 that gave the name aphantasia to severely impoverished visual mental imagery (Zeman et al., 2015). But despite the new name, this is not a new topic. As the epigraph of this introduction makes clear, the topic of individual differences in mental imagery has been steeped in suspicion across the domains of philosophy and psychology. The fact that mental images are a topic for which the question "do they exist?" lacks a satisfactory answer—at least in the minds of some—makes for an interesting topic of research.

The good news is that there is plenty of fresh ground ripe for the breaking in the world of cognitive measures of mental imagery, especially in the specific case of developmental aphantasia. However, there is much that must be discussed in terms of miscommunications in the existing literature, thanks to the controversial nature of individual differences in mental imagery ability stemming back more than 140 years.

Experimental Psychology and Mental Imagery

The scientific study of mental imagery within psychology began around the time of the birth of psychology (Kosslyn & Jolicoeur, 1980). Indeed, mental imagery refers to a large realm of research, and it "spans the entire history of experimental psychology" (A. Richardson, 1969,

p. 6). Therefore—with apologies especially to Allan Paivio's research into mental imagery as a mnemonic memory aid (e.g., Paivio, 1971)—this historical review will focus on the study of *impoverishment* of mental imagery rather than the entirety of mental imagery history.

Fechner observed that people had different levels of ability when it came to perceiving sensory stimuli: "the same stimulus may be perceived as stronger or weaker by one subject or organ than by another…. we can speak of a greater or lesser intensity of sensation… of the vividness of images of memory and fantasy" (1860/1966, pp. 38, 46). Multiple researchers (e.g., A. Richardson, 1969; Betts, 1909; Kosslyn & Jolicoeur, 1980; Paivio, 1971; Woodworth, 1938) consider Fechner's observations to be the pioneering work in individual differences of mental imagery from an experimental psychology perspective.

The first quasi-experimental approach in individual differences in mental imagery was most likely the work of Galton (1880; 1883). Galton gave surveys to groups of people, including scientists and schoolchildren, and asked them to describe their mental imagery when they tried to recall past experiences—most famously, while they tried to recall their breakfast table as they "sat down to it this morning" (Galton, 1883, p. 84). Galton himself had impoverished mental imagery, and at first doubted people who claimed to have vivid mental imagery—and they doubted his account as well. He wrote:

> Many men and yet a larger number of women, and many boys and girls, declared that they habitually saw mental imagery, and that it was perfectly distinct to them and full of colour. The more I pressed and cross-questioned them, professing myself to be incredulous, the more obvious was the truth of their first assertations. They described their imagery in minute detail, and they spoke in a tone of surprise at my apparent hesitation in accepting what they said. (Galton, 1883, p. 86)

It has been suggested that Galton's surveys were inspired by "his *paucity* of imagery—and his acknowledgement that others had good imagery" (Faw, 2009, p. 11). Galton made a groundbreaking revelation when a subset of those surveyed claimed to have *no* visual mental imagery of which to speak—what we would today call aphantasia. Watkins claimed that "up to six" of the "100 'men of science'" surveyed by Galton were people with aphantasia, although because the responses were qualitative, some interpretation must be used to claim that that those responders experienced aphantasia (2018, p. 43).

Three of the most famous early psychologists—Wilhelm Wundt, Edward B. Titchener, & William James—weighed in on the topic of mental imagery. Wundt acknowledged the role of imagery in imagination, emphasizing "the 'sensory vividness and picturableness of [imagination's] ideas'" (Perky, 1910, p. 424). Titchener himself had vivid mental imagery and believed that images were a crucial building block of the mind, such that *all* introspection made use of them—going so far as to equate "ideas with images" (Faw, 2009, p. 7). Wundt and Titchener presumably would have taken issue with the claim that some people do not experience mental imagery. James, on the other hand, seemed to embrace the idea that some people lack mental images, probably because he himself reported having poor mental imagery (see Faw, 2009). James even wrote that "*some people undoubtedly have no visual images at all worthy of the name*, and instead of *seeing* their breakfast-table, they tell you that they *remember* it or *know* what was on it. This knowing and remembering takes place undoubtedly by means of verbal images" (James, 1890b, pp. 57–58). James wrote about the individual differences in trains of thought—including a figure, see Figure 1 below—and how verbal imagery and visual imagery differ between any two thinkers. He wrote: "it would probably astound each of them beyond measure to be let into his neighbor's mind and to find how different the scenery there was from that in his own" (James, 1890a, pp. 269–270). James, therefore, presumably would have been receptive to the modern concept of aphantasia.

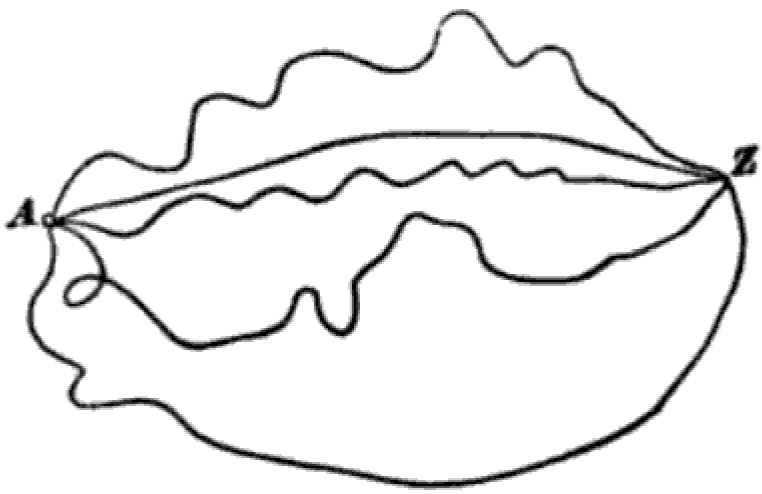


Figure 1. Reproduced from Fig. 28 in James (1890a, p. 269). The original noted: "A diagram may help to accentuate this indifference of the mental means where the end is the same. Let *A* be some experience from which a number of thinkers start. Let *Z* be the practical conclusion rationally inferrible [*sic*] from it. One gets to the conclusion by one line, another by another; one follows a course of English, another of German, verbal imagery. With one, visual images predominate; with another, tactile. Some trains are tinged with emotions, others not; some are very abridged, synthetic and rapid, others, hesitating and broken into many steps. But when the penultimate terms of all the trains, however differing *inter se*, finally shoot into the same conclusion, we say and rightly say, that all the thinkers have had substantially the same thought."

Galton's qualitative attempt to look at individual differences in mental imagery was further developed into a quantitative questionnaire by Betts (1909) who asked people to self-report their own use of mental imagery in seven different modalities. That is, he asked about imagery in the modalities of visual, auditory, cutaneous, "kinaesthetic," gustatory, olfactory, and "organic," with organic being defined as miscellaneous sensations such as hunger, fatigue, drowsiness, and headache. Using a self-reported 1–7 Likert scale, he measured the vividness with which each person could voluntarily call to mind each of the requested examples, including the taste of "your favorite soup," the smell of "an oil lamp blown out," the feeling of "the heat of a burning sun," and the "kinaesthetic" imagery of "rising out of a low chair" (Betts, 1909, pp.

23–24). Crucially, Betts discovered that "there is a moderately high correlation between ratings for different modalities" as part of his work in this area (see Marks, 1973, p. 17). People who report low imagery vividness for one modality tend to report low imagery vividness in other modalities. Betts also provided a rating option for "no image" which allowed for the possibility of aphantasia, although he did not overtly draw attention to that possibility and rather described such instances as "the image failing to appear" rather than attributing it to differences in the brains of the participants (Betts, 1909, p. 28).

Several other tests were developed in the first decade or so of the twentieth century, and these other tests claimed to be "objective" rather than subjective like Betts' survey. However, as pointed out by Kosslyn & Jolicoeur, "none of the early work is compellingly 'face valid'" —that is, many other explanations besides differences in mental imagery could explain the findings (1980, p. 144). The most notable early researchers involved with this "objective" measurement work were Angell (1910) and Fernald (1912). Some of these early measurements included frequency of word "types" used when writing (e.g., how frequently people used sight-related or sound-related words when writing prose), ability to mentally arrange letters into rows and columns and "read off" the image, ability to spell backwards, and finding a person's optimal presentation modality in list learning (i.e., learning by eye or by ear). In general, these former methods of assessment have been criticized, especially because the generation of mental images is not necessarily involved in these so-called objective tasks (see Woodworth, 1938, pp. 39-43; Kosslyn & Jolicoeur, 1980, pp. 141–144). I will return to this problem of "objective" measurement of mental imagery ability later on (see "The Struggle of Measurement").

The energy invested into discovering individual differences in mental imagery ability was cut short by psychology's movement towards behaviorism. This period in which mental imagery "began to fade as a serious subject for investigation" from the 1920s to the 1940s has been called the "great eclipse" of behaviorism (A. Richardson, 1969, p. 6). Most notably, John Watson—who might have had impoverished mental imagery—suggested that inner speech sourced from laryngeal movement ("subvocal talking") is the foundation of thoughts, including what people call visual imagery (see Faw, 2009, pp. 8–10). That is, Watson "rejected the idea that picturelike mental images exist" and "claimed that subtle movements of the larynx accompany imagery and that these movements are all that is important" (Kosslyn & Thompson, 2003, p. 723).

One exception to the eclipse of behaviorism was the work of T. H. Pear who continued to write about mental imagery. In one intriguing study, the research group looked at the individual differences in mental imagery use while listening to a radio drama delivered by gramophone. Based on questionnaires, it was determined that "people who like radio drama have more vivid visual imagery than those who dislike it" (M. Kerr & Pear, 1931). This was one of the earliest forays into how a person's mental imagery abilities shape that person's interactions with the world.

Interest in mental imagery re-emerged following the great eclipse of behaviorism, but it took several decades of investigation before reaching the modern discussion of aphantasia.

The Great Imagery Debate

One of the hot topics in cognitive psychology back in the 1970s (and running into the 2000s, see Box 1 from J. Pearson, 2019) was whether visual mental imagery was in the form of truly *depictive* representations in the brain, or if visual imagery was merely epiphenomenal following the activation of *propositional* encoding. The debate between depictive accounts and propositional accounts was known as the "imagery debate."

A depictive representation is one in which *distance* itself is part of the mental representation. More specifically, "each part of the representation corresponds to a part of the represented object such that the distances among the parts in the representation correspond to the actual distances among the parts. Thus, a depiction requires a functional space (e.g., an actual page or XY coordinate space)" (J. Pearson & Kosslyn, 2015, p. 10089). In other words, according to the depictive account, when a person generates visual imagery, the brain makes use of points that are placed in locations in a space. The depictive account was championed by Kosslyn and several collaborators (e.g., Kosslyn, 1983, 1994; Kosslyn et al., 1978, 2006).

The alternative explanation is the propositional account, which maintains that visual mental images are represented with propositional logic such that they are conceptual and not depictive (e.g., Pylyshyn, 1973, 2002, 2003). It is assumed that propositional representations may include entities (e.g., BALL, BOX), relationships between objects or parts of objects (e.g., ON, NEAR, FAR), properties (e.g., RED), and logical relationships (e.g., NOT, ALL), in different combinations, but would *not* include specific depictive/metric information in the form of points placed in locations in space (Kosslyn et al., 2006, p. 14). In other words, it is a null hypothesis that claims that "the process of imagistic reasoning involves the same mechanisms and the same forms of representation as are involved in general reasoning" (Pylyshyn, 2002, p. 158). Because this is a null hypothesis, no specific propositional coding exists in the form of a testable theory. Indeed, "no detailed versions of propositional theories of imagery have been developed" (Kosslyn et al., 2006, p. 122).

It is important to note that the depictive account does not assume that the brain *always* encodes visual information in a depictive form, or that all representations of visual information make use of depiction—it merely asserts that depiction is one possible format of representation that the brain can use, in *addition* to propositional representations like those used for language. For example, according to the depictive account, a person could encode a visual scene of a ball sitting on top of a box as a series of points places in relative locations in a space (a depictive format) *and/or* as propositional language such as ON(BALL, BOX). According to the propositional account, the person could *only* encode and represent that scene using propositional language such as ON(BALL, BOX).

Importantly, propositional accounts consider the experience of mental imagery to be *epiphenomenal*. An epiphenomenal experience is an effect that occurs as a byproduct of a cognitive process but does not influence the outcome of that process. That is, propositional accounts posit that visual mental images are not used in the brain's calculations to complete tasks, and are instead an epiphenomenon—illusory experiences without functional impact on the underlying mathematical processing. This idea of imagery as an epiphenomenon is complicated, so an example from Kosslyn et al. (1999) will help to illuminate. When a person reads a book using a lightbulb, the lightbulb produces two things: light and heat. However, only the light plays a functional role in the person's ability to read the book. Removing the heat would not affect the outcome of reading. Applying this to mental imagery, the depictive account argues that mental imagery is like *light*—causally linked to performance on tasks that make use of mental imagery, such as impacting accuracy or response time. However, propositional accounts argue that mental images are like *heat*—there is no causal link between the imagery itself and performance on such tasks, and that the tasks are influenced only by underlying propositional (non-depictive) representations. If propositional accounts are correct that conclusions are not drawn from depictive representations, then accuracy and response time should not depend on imagery ability.

As Kosslyn et al. (2006) point out, “there is no dispute that virtually any information can be represented in terms of propositions” (p. 52). Both accounts agree that propositional encoding is used in the brain. However, for the purposes of the present studies, it is important to determine whether the brain is also capable of using depictive representations that preserve spatial relationships and whether it uses them to make calculations that affect responses on tasks. I will now briefly review the evidence that the brain does make use of depictive representations.

Researchers have found strong evidence for depictive representations using a variety of methodologies. Initial evidence for the depictive nature of mental images came from classic studies of *mental rotation* and *mental scanning*.

Examples of classic mental rotation tasks can be found in Shepard & Cooper (1982). In one experiment originally reported in Cooper & Shepard (1973), they demonstrated that it takes a person longer to mentally rotate a visually presented stimulus the further that it is rotated away from its upright position. They used letters with one correct orientation (e.g., the letter ‘R’) and presented them at various orientations (i.e., rotated in degrees clockwise from upright: 0, 60, 120, 180, 240, or 300). It took participants longer to determine whether the letters were correctly written or mirrored depending on how many degrees from upright they had been rotated (see Figure 2). Response times were slowest when letters were upside-down (180 degrees), followed by 120/240, then 60/300, and responses were fastest when they were not rotated (0 degrees).

The steeple-shaped function for reaction time by rotation was preserved even when participants were cued about the identity of the upcoming letter, or when they were cued with an arrow about the orientation of the upcoming letter. However, when both the identity of the letter and the orientation of the rotation were cued ahead of time—either cued separately or cued together—the function flattened into non-significant differences between orientation, and overall reaction times decreased significantly. This result suggests that participants were able to mentally generate and rotate anticipated letters before presentation—but only if both object identity and orientation were known. When asked, participants explained that they created an "internal representation as a template against which they could rapidly match the visually presented test stimulus when it then appeared in that same orientation" (Shepard & Cooper, 1982, p. 95).

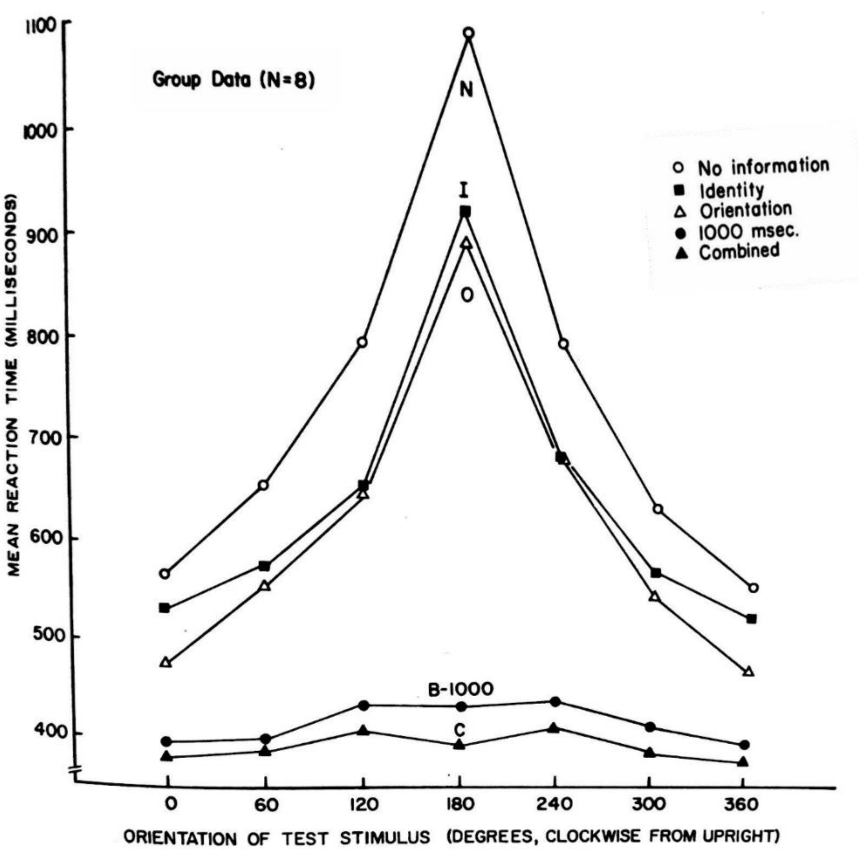


Figure 2. Reproduced from Figure 4.5 in Shepard & Cooper (1982). Reaction time as a function of orientation of the test stimulus. Advance information conditions are denoted by the symbols.

These findings suggest that mental rotation resembles genuine object rotation such that it takes longer to rotate an object when the rotation is larger—therefore, the brain is likely representing this rotational transformation in a depictive way that preserves this physical relationship. And, importantly, this depiction appears to influence a person's response time such that the visual mental imagery could not be described as unimportant to the mental calculation (i.e., the mental imagery was not epiphenomenal).

Another classic demonstration that mental imagery can be depictive and causally linked to performance is mental scanning. One interesting scanning task comes from Dror et al. (1993). In this task, a square donut shape was created out of several small squares (see Figure 3). During presentation, some of the squares were black while the rest were white. After studying the square donut stimulus, the participant hit a button to make an arrow appear in the ring for 50 ms, and then the entire stimulus disappeared. The participant was then asked to determine whether the arrow had been pointing at a black square. Participants had slower reaction times when the displayed arrows were located further away from the square that required inspection. This suggests that inspecting the mental image of the square donut relied on a depictive representation of the space between the arrow and the square.

Other mental scanning experiments have been used in a wide variety of forms (e.g., Borst & Kosslyn, 2010; Finke & Pinker, 1983; Kosslyn et al., 1978). Regardless of the exact nature of the task, these mental scanning experiments show a similar pattern of results—reaction times increase when required scanning distance increases. This positive correlation between scanning distance and reaction time is robust (see Reisberg, 2013).

The key takeaway from these experiments is that there is a relationship between the manipulation performed on the mental image (amount of rotation, distance to be scanned, etc.) and the amount of time that it takes to complete the manipulation. Other imagery manipulations have also been examined, including mental zooming, which reveal similar straightforward relationships between the features of the mental images and the amount of time needed to complete a feature-dependent task—and the relationship is often “a strikingly linear function” in studies using rotation, scanning, or zooming (see Reisberg, 2013, p. 375). All of this is strong evidence for depictive mental images, but the evidence does not stop there.

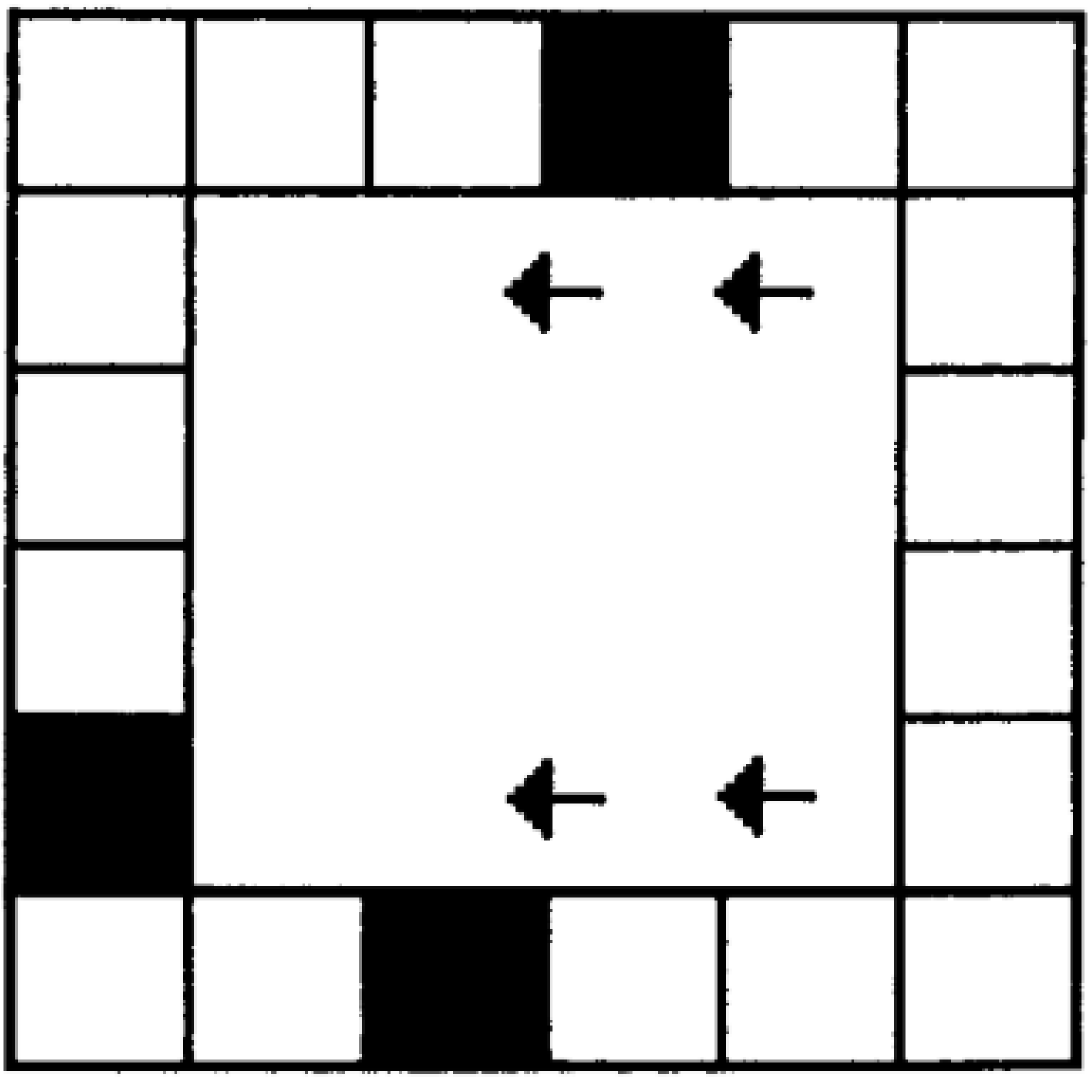


Figure 3. Adapted from Figure 5 in Dror & Kosslyn (1994). Three squares were colored black in each trial, but the locations varied. Only one arrow was displayed during each trial of this experiment. Not all possible arrows are displayed in this figure.

Further evidence for the depictive nature of mental images comes from an experiment that made use of structured visual noise to disrupt depictive representation of mental images. In a series of experiments, Borst et al. (2012) had participants answer questions about letters using mental images of those letters (e.g., does a given letter contain an enclosed space; does a given letter have a symmetrical form). While completing the task, the participants passively viewed a dynamic visual noise mask, or a control mask. The masks came in three forms (see Figure 4). The control mask was an unchanging gray field, the unstructured visual noise mask showed a randomly changing field of black and white dots, and the structured visual noise mask showed a changing field made up of parts of letters. As predicted by the depictive account, the structured visual noise mask was most disruptive to the processing of mental images of letters, such that more errors in questions about features of letters were committed when a structured visual mask was displayed as opposed to an unstructured or control mask. In theory, this result is due to the mental image of letters sharing depictive space with visual perception, and the similarity between the structured mask's letter-like features and the mental image's features led to more errors during mental inspection.

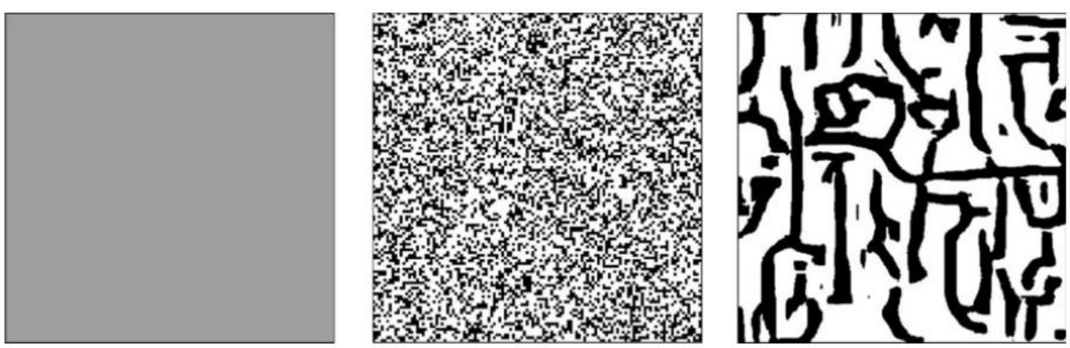


Figure 4. Adapted from Figure 1 in Borst et al. (2012). On the left is the blank gray visual field used as a control, in the middle is the unstructured visual noise mask, and on the right is the structured visual noise mask. Both the unstructured and structured visual noise masks were animated such that the pattern appeared to be moving and changing.

One important task that compared the similarity of mental imagery to visual perception was the image generation task from Podgorny & Shepard (1978; see also Task 1 from Dror & Kosslyn, 1994). Using a 5 x 5 grid of tiles, participants were asked to determine whether a probe that appeared in one tile overlapped with a form that was either imagined or perceived within the grid. Regardless of whether participants generated an image themselves from memory or perceived the stimulus directly, their ability to inspect and react to differences in the image followed predictable patterns such that more complicated stimuli required more time to inspect—for example, the letter F took longer to inspect than the less complicated letter L for both the mental imagery and direct perception conditions. Additionally, features of perception were preserved in mental imagery. Participants were able to react faster when the probe was on-target than when it was off-target. Additionally, people could react faster to on-target probes when the target fell on a vertical-horizontal junction (e.g., where the two lines meet in the letter 'L'). Overall, response times differed based on the forms perceived and the prompts used, but did not vary depending on whether imagery or perception was utilized. This finding supports the depictive account of mental image representations because it provides evidence that visual perception and visual imagery tap similar underlying processes.

One unusual piece of evidence for the depictive nature of mental images comes from the work of Mary Cheves West Perky (see Bartolomeo, 2002, p. 359). Perky (1910) built an apparatus that subtly displayed colored images of objects, ranging from a completely subliminal level to a barely supraliminal level. Participants stared at a blank glass window—with the apparatus cleverly concealed behind it—while imagining a series of requested objects. As the participants imagined the requested object (e.g., a banana), experimenters behind the window manipulated the apparatus such that the colored object just barely began to show up as noticeably perceivable by the participant. Over the course of three experiments, dozens of participants continued to believe that they were imagining the stimulus even after it was supraliminal—"at the end of the series, after all the introspections had been recorded, the observer was asked whether he was ‘quite sure that he had imagined all these things.’ The question almost always aroused surprise, and at times indignation” (Perky, 1910, p. 431). Intriguingly, the banana stimulus was displayed from an unusual visual angle—standing on end rather than lying down—but participants dismissed this as poor control over their own imaginations rather than discovering the truth that they were indeed perceiving a banana! One participant remarked “the banana is up on end; I must have been thinking of it growing” (Perky, 1910, p. 431). Perky’s finding was replicated over 50 years after its original debut, with these additional investigators theorizing that the confusion between genuine perception and visual imagery seems to occur because forming a mental image interferes with the normal use of perceptual processes such that perceptual detection thresholds rise (Segal & Nathan, 1964; Segal, 1971).

Conversely, people have also confused a true event of visualizing a stimulus with a false memory of having perceived it in a number of studies (e.g., Finke et al., 1988; Intraub & Hoffman, 1992; Johnson et al., 1979). For example, Johnson et al. (1979) found that a person’s estimate for the number of times that they had visually perceived a stimulus increased with the number of times that they had previously been asked to imagine that stimulus.

Together, these phenomena of mistakenly experiencing perception as mental imagery and mistakenly remembering mental imagery as perception are evidence for truly depictive representations of visual mental imagery. Visual scenes are perceived in a depictive way, and in order to confuse such perceptions with mental imagery, mental imagery most likely has similar depictive traits. Apparently, imagery depends on shared depictive space with visual perception.

Additional evidence comes from responses in the eye during or after mental imaging. One intriguing example is pupillary response to imagined light: imagining brighter lights leads to measurable constriction of the pupil, just as the pupil would respond to perceived light (Laeng &

Suluvedt, 2014). Similarly, vivid visual imagers show electroretinogram responses while *visualizing* flashes of light—these electroretinogram responses resemble responses at the retina during *perceived* flashes (Kunzendorf, 1984). Additionally, imagining bright light for long periods of time fatigues the retina more so than imagining bright light for shorter periods of time (Kunzendorf & Hall, 2001). There are even reports of vivid visual imagers naïve to opponent color theory experiencing opposite-color after-imagery following imagining of a stimulus (see Faw, 1997, p. 283). These surprising effects are plausible because the retina receives input from the brain via the optic nerve (e.g., Gastinger et al., 2006), ostensibly following a path from V1 to the lateral geniculate nucleus to the retina (see Kunzendorf & Hall, 2001).

It has been suggested that V1 in the primary visual cortex fulfills the role of depicting visual mental imagery (e.g., Kosslyn et al., 2006). V1 is known to depict visual perception: in the macaque brain, radioactive sugar was taken up by cells in V1 to match the pattern of lights that the monkey was trained to stare at, revealing a literal picture in the brain (Tootell et al., 1988). Strong evidence for depictive representations of visual mental imagery in the visual areas of the brain comes from various neuroscience domains.

Event-related potentials (ERPs) suggest that visual areas are active during mental imagery even in the absence of the perception of visual stimuli (Farah & Peronnet, 1989). Additionally, repetitive transcranial magnetic stimulation (rTMS) applied to V1 can disrupt visual imagery (Kosslyn et al., 1999). When measured with functional magnetic resonance imaging (fMRI), mental rotation tasks engage the extrastriate region called MT which is known for its role in the perception of motion (Cohen et al., 1996). All of this is evidence that mental imagery engages similar neural pathways as direct perception.

Additional evidence that imagery is depicted comes from decoding algorithms that can correctly classify which mental image a person is generating based on activity in the primary visual cortex (see J. Pearson & Kosslyn, 2015). J. Pearson (2019) summarizes this research area:

> …the content of mental imagery can be decoded from early visual areas, including V1 and V2… despite relatively low BOLD amplitude responses. The decoding algorithms used by these approaches can be trained on visual perception, visual working memory content and, of course, imagery, and in all cases can accurately decode the content of mental imagery. (p. 626)

In other words, by looking at the voxels that are most active during fMRI, researchers can determine which object, category of object, or visual feature is being imagined using mental imagery (see also Roldan, 2017). People who were more precise at a mental rotation task showed brain responses to mental imagery that were more accurately classified by such algorithms (see the "Staircase Procedure" section of the supplemental experimental procedures and Figure S2 of Albers et al., 2013). This result suggests that information is depicted in early visual areas of the brain in a way that is useful for performing tasks that benefit from such depictions.

The imagery debate was declared over by J. Pearson and Kosslyn (2015) in an article aptly named "The heterogeneity of mental representation: Ending the imagery debate." In terms of remaining questions in imagery research, Kosslyn et al. wrote: "if… the [imagery] debate is settled: depictive representations are used in at least some circumstances. The focus would then shift to when, precisely, depictive representations are used" (2006, p. 69). The present series of studies will attempt to further our understanding of when depictive representations are used by testing whether depictive representations are *not* used by people with aphantasia.

Before I can discuss cognitive aphantasia studies, we must examine the evidence from neuropsychology. Neuropsychology provides strong—albeit subjective—evidence that mental images use depictive representations. This evidence comes from neuropsychological cases studies in which a person had the ability to visualize mental images depictively (e.g., with form, color, etc.), but then lost or partially lost that ability.

Neuropsychological Discoveries of Impoverished Mental Imagery

Neuropsychological researchers were no strangers to the idea that there were individual differences in mental imagery ability. It has been known that brain damage can cause a person to lose imagery abilities since at least 1883 (Charcot & Bernard, 1883).

Indeed, there is no shortage of case studies of people losing mental imagery abilities following brain damage (e.g., Basso et al., 1980; 2 cases in Brain, 1951; Charcot & Bernard, 1883; Chatterjee & Southwood, 1995; Farah, Levine, & Calvanio, 1988; Goldenberg, 1992; Humphrey & Zangwill, 1951; Luria, 1972; Ogden, 1993; Sacks, 2010, p. 203; Spalding & Zangwill, 1950; Thorudottir et al., 2020; Zeman et al., 2010). A review from nearly 40 years ago found 37 case descriptions of "loss of visual imagery due to brain damage" published in English (Farah, 1984, p. 245), and the number has only increased. As will be discussed later, there was not one agreed-upon way to refer to the symptom of the loss of mental imagery, but it is now called *acquired* aphantasia (see "The Long Road to Aphantasia"). The symptom of the loss of mental imagery was considered important enough to neuropsychology that the cases of Charcot & Bernard (1883) and Wilbrand (1887) and related discussions of mental imagery ability were both included in the first volume of the seminal collection *Classic Cases in Neuropsychology* (respectively discussed by Young & van de Wal, 1996; Solms et al., 1996).

Even Luria's famous biography of Lieutenant Zasetsky—the long-form brain damage case study published in 1972 as *The Man with a Shattered World*—specifically mentioned a loss of imagery following his severe wounding. Zasetsky described this to Luria:

> 'My therapist would mention the word cat or dog and say: "Try to picture to yourself what a dog looks like, what kind of eyes and ears it has. Can you see it?" But I couldn't visualize a cat, dog, or any other creature after I'd been wounded…. I tried to remember my mother's and sister's faces but couldn't form any image of them.' (Kaczmarek et al., pp. 139–140)

What arises from these neuropsychological cases is evidence that there is a double dissociation for the cognitive abilities of visual perception and visual imagery. Some people lose visual perception but not visual imagery (e.g., Bartolomeo et al., 1998; Behrmann et al., 1994; Zago et al., 2010), some people lack visual imagery but have intact visual perception (e.g., Brain, 1954; Riddoch, 1990), and some people lose both visual perception and visual imagery (e.g., Brown, 1972; Farah, Levine, & Calvanio, 1988; Policardi et al., 1996). However, it is more common to lose both visual perception and visual imagery than it is to lose one but not the other (Farah, 1984, 2000; Trojano & Grossi, 1994). This finding bolsters the argument that mental imagery and visual perception inhabit similar brain regions.

Relevant to the discussion of whether mental imagery is depictive is the fact that damage to V1 is neither necessary nor sufficient for inducing imagery deficits. People without V1 sometimes report mental images (e.g., Chatterjee & Southwood, 1995; de Gelder et al., 2015). In fact, "reports of cortically blind patients with spared visual imagery are not rare" (Bartolomeo, 2002, p. 363). Additionally, some people with an intact V1 region have shown severe impairments in mental imagery (e.g., Thorudottir et al., 2020). A review of neuropsychological evidence put forth the theory that "occipital damage is neither necessary nor sufficient to produce imagery deficits…. visual mental imagery abilities might require the integrity of brain areas related to vision, but at a higher level of integration than previously proposed," which suggests that visual areas supporting the processes of identification such as areas found in the inferior temporal lobe may be sufficient for mental images to be generated and experienced (Bartolomeo, 2002, p. 357). Therefore, while visual imagery has been shown by a large volume of evidence to be capable of being depictive in nature, it is not settled whether it must always be depicted in V1. The present studies are agnostic as to the location of mental imagery mechanism(s) in the brain, but future aphantasia research will need to grapple with such questions.

There does, at times, appear to be a relationship between specific deficits in perception and specific deficits in mental imagery ability (e.g., Bartolomeo, 2002; Farah, 1984; Ganis et al., 2003, p. 637; Goldenberg, 1992; Levine et al., 1985). People who acquire prosopagnosia are likely to also lose the ability to imagine faces (e.g., Levine et al., 1985; Shuttleworth et al., 1982). People who lose the ability to perceive color may lose the ability to imagine color (e.g., Ogden, 1993; Riddoch & Humphreys, 1987; Sacks & Wasserman, 1987). Visual form agnosia and losing the ability to imagine object form seem to be related (Farah, 1984; Trojano & Grossi, 1994). In some cases of unilateral neglect, there is evidence that a person loses the ability to use mental imagery for that side of space as well (Bisiach & Luzzatti, 1978). Visual disorientation in perceptual tasks is associated with impaired spatial imagery (e.g., Levine et al., 1985).

However, it is important to note that "the relation between the site of damage and the type of deficit is not simple or straightforward" (Ganis et al., 2003, p. 643). While reviewing the literature, Farah wrote that "parietal lobe, occipital lobe and temporal lobe damage have all been associated with loss or severe deficit of visual imagery ability, and neither hemisphere can be excluded" (1984, p. 246). In fact, the more that cases of acquired aphantasia are studied, the more apparent it becomes that there are multiple levels of dissociation in the human brain when it comes to mental imagery. These levels of dissociation will be briefly explored in order to better pinpoint operational definitions for the present studies.

Dissociations of Mental Imagery Component Dimensions

As research into mental imagery deepened, it became apparent that many component dimensions demanded consideration. Prominently among them are imagery modality, the dissociation between spatial and object imagery, the dissociation between imagery vividness and accuracy, imagery voluntariness and controllability, the state of consciousness during imagery occurrence (i.e., dreaming vs. awake), and implicit versus explicit mental imagery.

Imagery Modality

Dissociations of modality continue the work of Betts (1909) by splitting mental images into categories of the sensations/perceptions that they resemble. For example, "auditory mental imagery is accompanied by the experience of 'hearing with the mind's ear,' and tactile imagery accompanied by the experience of 'feeling with the mind's skin,' and so forth" (Kosslyn et al., 2006, p. 5). The modality of concern for this project is vision because the definition of aphantasia depends upon impoverished visual mental imagery, while it is agnostic towards other modalities. In fact, many people with aphantasia report experiencing imagery in other modalities, most commonly auditory (Zeman et al., 2020). Thus, the proposed experiments will target imagery that makes use of the sense of vision.

Another dissociation for mental imagery concerning modality is between spatial imagery and object imagery. The object/spatial vision dissociation—also known as the ventral/dorsal vision dissociation—has been well established in the literature for several decades (e.g., Mishkin

& Ungerleider, 1982). In bottom-up perception, information is sent from the primary visual cortex through a ventral path to the inferior temporal lobe where object features and identities are further processed—a separate path allows the primary visual cortex to send information towards the posterior parietal lobe, where information about location and motion are further processed (see Ganis et al., 2003, p. 633). Some other evidence for this dissociation between space and object imagery includes the fact that congenitally blind people can still complete spatial imagery tasks like mental scanning when the task is adapted to a non-visual methodology (N. H. Kerr, 1983).

The dissociation between object and spatial imagery is supported explicitly by a meta-analysis by Mazard et al. (2004) that found evidence that the superior parietal cortex was more strongly activated by spatial imagery (e.g., mentally scanning across a mentally generated image of a previously-learned map) while the fusiform gyrus and other anterior parts of the ventral pathway of vision were specifically engaged by visual mental imagery (e.g., generating mental images of concrete objects). However, this meta-analysis also noted that "even when the tasks are designed specifically to draw on spatial processing, they require some object processing and vice versa"—so while one can be prioritized, they remain linked (p. 688).

This linkage is important because—as will be discussed—self-reported mental imagery ability usually focuses on object imagery but not necessarily on the spatial dimension of imagery (Reisberg et al., 1986; Farah, Hammond, et al., 1988). Aphantasia as a label does not explicitly consider spatial imagery, and in some cases a deficit of spatial imagery is considered separately from object imagery aphantasia (e.g., Blazhenkova & Pechenkova, 2019; Crowder, 2018).

Vividness

Vividness refers to how much the imagery driven by top-down processes resembles perception driven by bottom-up processes. The more vivid a mental image, "the closer it approximates an actual percept" (Marks, 1999, p. 570). For example, when the modality is visual, "*high vividness imagers* report images rich in visual detail, so that the imagined object is described as viewed from a particular vantage point, as having a particular position, colour and size, and so forth. *Low vividness imagers* report experiences which leave all these details unspecified" (Reisberg & Leak, 1987, p. 521).

Subjective self-reports from people with extremely vivid visual mental imagery describe "the experience of perceiving, but in the absence of the immediate appropriate sensory input" (Kosslyn et al., 2006, p. 4). Shakespeare's Hamlet used the term "mind's eye" to describe this quality of mental images in which they can be "read" like a printed picture or page of a book, and "this expression remains in common use 400 years after Shakespeare coined it" (Reisberg, 2013, p. 375). Visual imagery does strongly resemble visual perception in some people, not only in subjective self-reports, but also in more objective patterns of neural activity.

Objective measurements from brain scans such as fMRI show brain activity common to both visual imagery and visual perception, with relatively low similarity in the "early" visual areas (e.g., V1) and relatively high similarity in higher-level visual areas (e.g., V3, FFA, etc.), overall "becoming increasingly similar with ascension up the visual processing hierarchy" (J. Pearson, 2019, p. 626). There may also be "systematically related yet separate cortical areas for perception and imagery… the neural representation during recall may lack much of the richer, elaborative processing of the initial perceptual trace" (Bainbridge et al., 2021, ln. 133). When the neural territory for both imagery and perception is considered, it has been estimated that 92% is common to both, meaning that the physical overlap between the mechanisms is enormous (Ganis et al., 2004). As previously mentioned, brain damage is more likely to disrupt both perceptual processes and imagery processes, as opposed to disrupting one but not the other (e.g., Farah, 1984, 2000; Trojano & Grossi, 1994). This large degree of overlap may help explain why it is possible for mental imagery to closely resemble perception in people with vivid mental imagery.

The activation of several posterior regions (e.g., the fusiform gyrus) show positive correlations with reported vividness, while the activation of frontal areas (e.g., the anterior cingulate cortex) and other areas such as the auditory cortex show negative correlations with vividness (Fulford et al., 2018). Resting state fMRI shows that people with more vivid mental imagery have stronger connectivity between pre-frontal cortices and visual areas (Milton et al., 2020). I will revisit these neural findings later when constructing my theory of brain differences in people with aphantasia (see “Theories of Individual Differences in Visual Imagery”).

An important note about imagery vividness is that it is not the same as imagery accuracy. The dimensions of vividness and accuracy are sliding scales that can vary somewhat independently from one another. A highly vivid mental image is not necessarily a highly accurate mental image. A. Richardson writes: “whatever it is that we measure with our vividness questionnaires, it is not something that necessarily predicts high and low accuracy of recall” (1994, p. 7). Supporting this are studies showing that people with highly vivid mental imagery may have inferior accuracy for visual details when compared to people with less vivid mental imagery (e.g., Reisberg & Leak, 1987). People who report vivid mental imagery are also more likely to make source monitoring errors between information they read in a paragraph and information received from pictures (Dobson & Markham, 1993). Similarly, people with highly vivid mental imagery have been shown to be more susceptible to misinformation, possibly because they can convince themselves of its reality using vivid imagery (e.g., post-event misinformation presented following a film in Tomes & Katz, 1997; suggestibility for words *not* presented in the Deese-Roediger-McDermott paradigm in Winograd et al., 1998). However, the correlation can also run the other way such that higher imagery vividness is associated with correct recall (e.g., memory for scenic details in Marks 1973; identifying the perpetrator of a mock crime in Riske et al., 2000). Therefore, while mental imagery vividness is often considered to be positively correlated with accuracy on tasks that make use of mental imagery, this is not necessarily always the case. Statistical analyses of the effects of subjective mental imagery abilities on objective measurements of behavior tasks should allow for effects in either direction.

Voluntary/Involuntary and Controllability Dimensions

Another dimension of imagery is whether an instance of mental imagery is voluntary or involuntary. The key difference here is whether the imagery was intentionally called to mind, or if it began without intentional generation. Three important examples of involuntary mental images are those that occur during hypnogogic hallucinations and REM-stage dreams, those that occur during Post Traumatic Stress Disorder (PTSD), and those that occur during the use of hallucinogens.

Despite claiming no experience of *voluntary* mental imagery, people with aphantasia do often claim to experience *involuntary* mental imagery. Interestingly, one large survey of 2000 people with aphantasia found that 63.4% reported dreaming visually (Zeman et al., 2020). Other survey work with 267 participants found that, when compared to controls, people with aphantasia have fewer dreams at night, less control during their dreams, and fewer sensory qualities in their dreams in favor of more semantic content (Dawes et al., 2020). Because people with aphantasia are a heterogeneous group when it comes to involuntary visual imagery, it is primarily voluntary mental imagery that is under investigation in this series of studies. However, also of interest is the frequency with which people with aphantasia experience visual imagery during dreams, because this is informative for whether a particular person's brain is *capable* of generating mental images.

Similar to the voluntary/involuntary dimension is the dimension of controllability. When mental imagery is controllable, a person can manipulate images with ease. When mental imagery is uncontrollable, the person is less able to manipulate the images using conscious effort. Images that start out as involuntary can become controllable, such as in the case of lucid dreams or during cognitive behavioral therapy in which a person learns to replace traumatic images with images that are more positive in the hopes of reducing symptoms of PTSD. Images that are voluntarily generated may have different degrees of controllability, and some imagery questionnaires purport to measure individual differences in ability to control voluntary mental imagery (e.g., A. Richardson, 1994; Gordon, 1949). For example, participants may be asked to imagine a car in different colors, lying on its back, or moving along a road, requiring the participants to update their mental image with these new details and report their ability to do so (Gordon, 1949). While the dimension of controllability has been of major historical interest (see A. Richardson, 1969, p. 50), it is beyond the scope of the current aphantasia research because most people with aphantasia claim not to experience voluntary mental imagery in the first place, and therefore controllability does not appear to be a valid dimension for aphantasia research.

Implicit versus Explicit

Some theorists distinguish between implicit mental imagery and explicit mental imagery, but with various terminology. Phillips (2014) wrote of the difference between *representational* imagery and *experiential* imagery. Representational imagery is depictive imagery used by the brain for cognitive calculations but without a conscious experience of perceived mental images. In contrast, experiential imagery is consciously experienced as perception. Holt believed that mental images could sometimes be unconscious; he suggested that "a brain process without awareness" should be called a "*presentation*" while "phenomenal content of a sensory or quasisensory" nature (explicit images) should be referred to as an "*image"* (1972, p. 10). The distinction between representational/unconscious/presentations and experiential/conscious/images is more concisely referred to as the implicit/explicit theory of mental imagery. Implicit imagery differs from the propositional representations discussed earlier (see "The Great Imagery Debate") because implicit imagery is said to be depictive in nature, even though it is not experienced as depictive.

Some definitions of mental imagery *include* the fact that the imagery should be *explicitly experienced*, but there are also definitions in the literature that do not have this requirement. For example, Paivio (1971) wrote that "imagery" as a term should be "used to refer to a memory code or associative mediator that provides spatially parallel information that can mediate overt responses without necessarily being consciously experienced as a visual image" (pp. 135–136). This definition allows for both explicit and implicit imagery. Similarly, Nanay (2021) proposed an inclusive definition: "mental imagery is perceptual processing (that is, processing in the early sensory cortices, V1, V2,V4/V8, MT) that is not triggered directly by sensory input" (p. 1).

The important takeaway here is that many mental imagery researchers have decided that vividness—defined earlier as resemblance to perception—is not a necessary component of a depictive representation in the brain, such as in the case of implicit mental imagery. For example, the great imagery debate left room for both explicit and implicit images. The conclusion of the debate was that people both *experience* mental imagery and *use* mental imagery in order to perform cognitive calculations, but the evidence did not suggest that both are always equivalent such that the experienced image is also the basis of the cognitive calculation itself (e.g., the experienced "explicit" image may be partly or entirely epiphenomenal while an "implicit" image is responsible for the cognitive calculation; see Phillips, 2014).

If there truly is an implicit/explicit distinction, then perhaps implicit mental imagery is used to solve tasks like mental rotation and mental scanning, while explicit mental imagery is epiphenomenal but reported on subjective questionnaires about vividness (Phillips, 2014). In theory, implicit mental images should be measurable by cognitive tasks (e.g., response time to image rotation, image scanning, etc.), but the measurement of—and very existence of—explicit mental imagery *depends* upon subjective reporting from individuals because due to its epiphenomenal nature, explicit imagery is not measurable by cognitive tasks. This could lead to a mismatch between two ways of measuring mental imagery: "objective" cognitive tasks that measure implicit imagery versus "subjective" self-reported vividness which taps explicit imagery. Indeed, the literature reflects a mismatch between these ways of measuring mental imagery, but whether the mismatch is due to implicit/explicit imagery in unclear. This interpretation of the difference in explicit (objective) and implicit (subjective) measurement will be discussed in greater detail later (see "Implicit/Explicit Theory of Aphantasia") and will be important later when considering theories of aphantasia.

Now that we have some working understanding of what mental imagery consists of from the perspective of a person experiencing it (i.e., it has modality, vividness, controllability, etc.), it is now time to more specifically discuss the people who claim *not to experience* mental imagery.

The Long Road to Aphantasia

From the very first investigations, researchers encountered people who claimed to have no mental imagery. Unlike the neuropsychological cases described above, these people did not have apparent brain injuries. In reference to those who reported no mental images, Fechner described people with "bare thought," Galton wrote of people who "reported an *absence* of images," and Betts allowed for ratings of "no image" when evaluating the seven sense modalities on his survey (Roeckelein, 2004, pp. 158–161).

It is not a new idea that mental imagery ability is a spectrum in neurotypical people from high-ability imagers to low-ability imagers. Individual differences between neurotypical people leading to differences in imagery ability—especially gender, age, and intelligence—have been proposed since at least the days of Galton (1883). What *is* new is the specific investigation into the absolute zero point of the spectrum where a person reports no visual imagery at all despite having a seemingly intact brain: *developmental aphantasia*.

It is important to distinguish the difference between acquired aphantasia and developmental aphantasia. Acquired aphantasia is when a person had the ability to use mental imagery but then lost that ability due to changes in their brain, often because of a lesion. Examples of this are explored in a previous section called "Neuropsychological Discoveries of Impoverished Mental Imagery." In contrast are the people with developmental aphantasia. These people reportedly grow up without the use of mental imagery. Most people with developmental aphantasia spend a good portion of their lives unaware that the way that they imagine things is different from the typical person. The majority learn of their mental imagery deficit after the age of 20, and while there are a variety of different discovery methods, the most typical is a discussion with someone who has different imagery abilities (see Fig. 5 of Zeman et al., 2020).

The emergence of documented developmental cases of aphantasia probably began with a single case study from Botez et al. (1985) wherein a person went to the hospital because of an inability to visualize pictures in his imagination—and claimed that he had never been able to do so. This deficit was described as "pure" in the case study, in reference to the fact that this developmental type did not seem to be co-morbid with other visual symptoms as is common in acquired cases.

However, the most famous paper discussing developmental aphantasia did not arrive until 2015. A research group published a typical acquired aphantasia case study in 2010 (Zeman et al., 2010) which then attracted the attention of a science communicator who promoted the piece in a popular science magazine (Zimmer, 2010). This event prompted many readers who recognized the symptoms in themselves to contact the authors, resulting in the identification of dozens of people with developmental aphantasia. As a result, the disorder was officially named: "We propose the use of the term 'aphantasia' to refer to a condition of reduced or absent voluntary imagery" (Zeman et al., 2015, p. 379). Modern self-report data indicates that 2–5% of people are "very poor- or non-visual imagers" despite having no reported deficiencies in visual recognition abilities or brain damage (Faw, 2009, p. 1). Recently, some have defined aphantasia specifically as a person possessing *zero* mental imagery abilities rather than reduced abilities. For example, Dawes et al. (2020) wrote "...aphantasia appears to represent a veridical absence of voluntarily generated internal visual representations" (p. 1). As will be discussed later, categorical distinction of aphantasia severity is still in development (see "Aphantasia as VVIQ Score").

The term aphantasia was not the first term coined to refer to people with zero mental imagery. "Charcot-Wilbrand Syndrome" was used to refer to a loss of visual imagery combined with visual agnosia and disturbed visual dream content, based on publications by Charcot in

1883 and Wilbrand in 1887 (although, curiously, Wilbrand's patient did not show a loss of mental imagery: see Solms et al., 1996). Other historical terms included "visual irreminiscence" (Nielsen, 1946, p. 270), "disturbed revisualization" (Nielsen, 1946, p. 81), "defective revisualization" (Botez et al., 1985), "non-conscious imagers" (Faw, 1997), and "blind imagination" (Zeman et al., 2010). Importantly, these terms were mostly applied to acquired cases. Aphantasia can refer to either acquired or developmental cases, giving the term a broader range and therefore a more secure purchase in humanity's collective lexicon, although the debate about the name for this disorder is not finished (see Monzel et al., 2022).

Why did it take well over 100 years to give an official name and line of investigation to people claiming no mental imagery? One issue is that personal introspective experiences of mental imagery have been shown to influence researchers' stances on whether mental imagery takes a propositional or depictive format. Reisberg et al. (2003) found a correlation between researchers' VVIQ scores—a common measurement for mental imagery vividness, discussed in more detail later (see "Subjective Measurement of Imagery Ability")—and what their beliefs had been about the format of mental imagery in the year 1980. This result revealed that researchers with less vivid mental imagery more favored propositional accounts. Interestingly, opinions shifted over time such that researchers became more in favor of depictive accounts, suggesting "that scientists really do pay attention to data and let the data take precedence over their introspections or intuitions" (Kosslyn et al., 2006, p. 181). Although, importantly, "some of the extreme cases—who reported no imagery—persisted in denying the existence of depictive representations" (Kosslyn et al., 2006, p. 181). This personal bias may have influenced historical researchers as well, including Titchener, James, and possibly Watson (see "Experimental Psychology and Mental Imagery").

Even though a potential solution here is obvious, researchers have been hesitant to agree upon it until recently. As pointed out by the philosopher Dennett: "...one breathtakingly simple explanation of the phenomenon, and one that is often proposed, is that in fact some people do have mental images and others don't" (1978, p. 175). In other words, that there are extreme individual differences in how people experience or do not experience mental imagery. Dennett was not the first to point out this possibility, but he did memorably name the sides of the debate: "a war between the believers and the skeptics, the lovers of mental images—let us call them iconophiles—and those who decry or deny them—the iconophobes" (1978, pp. 174–175). These iconophile/iconophobe camps of thought do not perfectly align with the depictive/propositional imagery debate (e.g., someone arguing for propositional mental images could believe that mental images are epiphenomenal while being either an iconophile or an iconophobe). However, as Reisberg et al. (2003) found, people who experience stronger mental imagery are more likely to believe that it is an important mental mechanism, and vice versa. This bias is a problem for the investigation of mental imagery because every person—with the possible exception of people who acquire aphantasia suddenly—has access only to one type of mental imagery experience and cannot understand the qualia that the other group is describing.

Just as iconophiles have asked iconophobes for evidence that the iconophobes do not experience mental imagery, so too have iconophobes asked for evidence that iconophiles truly "see images" in their minds. Both sides have doubts that the other side is engaging honestly in the discussion. Are many of us failing to accurately describe how we "visualize" information? As discussed in "The Participants Are Wrong: The Untrustworthiness of Self-Reporting Mental Imagery," are some people "wrong" about their mental imagery capabilities due to incorrect introspection or incorrect comparison to the abilities of others?

Having one's own experience of visual mental imagery may cloud a person's ability to understand another person's experience. As Reisberg et al. wrote: "if the imager and non-imager are witnesses to each other's report, each will give all signs of being mystified by the other, and will indicated [*sic*] that the experience described by the other is puzzlingly foreign" (Reisberg et al., 1986, p. 52). This notion of being puzzled by another person's thought process harkens back to James' (1890b) comment about different people using different trains of thought to arrive at the same destination (see Figure 1). Because there is doubt about aphantasia, or at least an inability to understand it completely without experiencing it, it is important to consider the evidence for the very existence of aphantasia.

Evidence for Non-Depictive Representations in Aphantasia

The first major piece of evidence for non-depictive representations of mental imagery in people with aphantasia comes from the previously-mentioned clinical cases, in which people had the ability to use mental imagery but then lost it due to brain damage. However, by definition that evidence only applies to the existence of acquired aphantasia.

There is also neurological evidence for developmental aphantasia. People who self-reported low mental imagery vividness showed more widespread cortical activation than people who reported high mental imagery vividness, such that activity in the visual areas associated with mental imagery (e.g., V1) was lower for people with low vividness, demonstrating at least some metacognitive accuracy in a person's judgement of their own mental imagery abilities (Fulford et al., 2018). Resting state fMRI revealed that people with aphantasia have significantly less connectivity between prefrontal cortices and the visual network relative to people with *hyperphantasia*—people with extremely vivid mental imagery (Milton et al., 2020). Additionally, Milton et al. reported that during active fMRI, there was evidence that people with aphantasia have less activation in the anterior parietal regions relative to people with hyperphantasia when considering the difference between brain activation during perception and brain activation during imagination. Wicken et al. (2021) showed that people with aphantasia had a smaller galvanic skin response to written frightening paragraphs when compared to a control group, even though people with aphantasia did not significantly differ from controls in their galvanic skin response when reacting to scenarios experienced via direct visual perception.

One piece of evidence that is frequently cited within the aphantasia literature as objective evidence of developmental aphantasia—even though it *is not* objective evidence—makes use of the binocular rivalry technique. Keogh & Pearson (2018, 2021) published multiple sets of findings that, at face value, seem to demonstrate that people with aphantasia are not as affected by imagery-based priming in the binocular rivalry paradigm when compared to control participants. In the 2018 study and in a portion of the 2021 study, people who self-identified as having aphantasia and a group of controls were asked to imagine either a red-horizontal or a green-vertical Gabor patch for six seconds before actually being presented with both patches (one presented to each eye) for 750 ms, and then they were asked which color they saw during the presentation. The responses of the aphantasia group were not significantly different from random responding as would be expected if they did not successfully generate mental imagery, and they were less "primed" than a control group. That is, control participants were approximately 59% primed by the imagination phase such that the color imagined for six seconds was the same color that they reported perceiving during the 750 ms presentation 59% of the time, but the participants with aphantasia were only around 51% primed (means were not reported in the 2018 publication, so I calculated them from Fig. 2E. I requested the original data from the first author but did not receive the data. The lack of means and standard deviations in both publications obfuscated interpretation).

One interpretation of these data is that the participants with aphantasia truly lack mental imagery. That is, this paradigm *requires* successful generation and maintenance of mental images in order for priming to occur. However, closer consideration is warranted because, to produce these findings, all the aphantasia group needed to do was *randomly respond throughout the studies*. The people with aphantasia in these studies were not naïve to the hypothesis and could have easily malingered a mental imagery deficit. Furthermore, the claim that there is evidence for people with aphantasia lacking mental imagery ability because their responses to priming were *not* significantly different from random responding is invalid—this is not the detection of an effect, it is merely the absence of an effect. Such null results could also have been aided by having a small sample size. In the 2018 study, there were only 15 participants with aphantasia compared to 209 control participants. Even though 15-participant bootstrapping applied to controls also produced an equal or higher mean priming for the control group than the aphantasia group 999 times out of 1000 simulations, this does not change the fact that because of such a small aphantasia sample, even just a handful of malingerers (i.e., random responders) could have resulted in the aphantasia average being not significantly different from random responding. In the 2021 study, there were only 10 participants with aphantasia. Mock priming was used in an attempt to cull malingering (e.g., mock rivalry stimuli were presented as “catch trials” to see if the participant would respond with the primed color anyway). However, the method of calculating the mock priming was problematic such that a person who was randomly responding to all stimuli would not be distinguishable from the group using this method (R. Keogh, personal communication, March 3, 2022). That is, mock priming for any given participant in the aphantasia condition was not significantly different from 50%, but that is true for any person who did not show response bias (e.g., did not fall for the catch trials) and for any person who randomly responded. Therefore, malingerers who randomly responded would *not* have been thrown out from the sample using this method. As such, the binocular rivalry paradigm—which is touted in the aphantasia literature as “objective perceptual measurement” of visualization ability (J. Pearson, 2019, p. 627)—is on shaky footing due to being interpretable as an example of malingering on the part of the participants with aphantasia. For the record, I do not believe that the aphantasia participants were malingering in these studies, I just find it problematic that the statistics used in these studies cannot distinguish between “real” effects of aphantasia and malingered ones.

Overall, little conclusive work has been done in the age of aphantasia (post-2015) to support the idea that this group of people is categorically different from the average person with typical mental imagery ability. This series of studies continues the investigation into whether there is good evidence for people having zero mental imagery ability through the use of cognitive measures. But first, I must discuss the underlying theories as to why aphantasia arises.

Theories of Individual Differences in Visual Imagery

Attention must now be turned to theories of individual differences in visual imagery. As was established earlier, visual mental imagery can at times be truly depictive in nature, and can at other times be propositional (see “The Great Imagery Debate”). It is certainly possible that some people are more likely to use depictive imagery than others—and perhaps that some people do not use depictive mental imagery at all—but what can account for this difference?

Neurotypical Visual Imagery

Before I focus on what could make people with developmental aphantasia unable to use visual imagery, I will briefly review how mental images are generated in neurotypical brains.

A *cognitive* theory of how mental images are brought forth from long-term memory is perceptual anticipation theory (Kosslyn & Thompson, 2003). Essentially, this theory states that mental images are experienced when a person anticipates perceiving an object so strongly that a depictive representation of the object is created in early visual cortex. This explanation provides the piece of the puzzle as to what a person *cognitively does* in order to initiate mental imagery. To my knowledge, perceptual anticipation theory has not been applied to aphantasia.

However, such a theory leaves out what is physically happening at the level of the brain. This omission is solved by applying another theory. The most influential *neurological* theory of mental imagery generation comes from Kosslyn's model (1994, p. 383). He divides the imagery process into four parts: generation, inspection, maintenance, and transformation. These four parts rely on several processing subsystems, which are characterized as mechanical input/output "black boxes" (Kosslyn, 1994, p. 25). These subsystems include a visual buffer, an attention window, separate paths for spatial properties processing and object properties processing, links to associative memory, an information shunting mechanism, and an attention shifting mechanism. Originally, according to Kosslyn, the subsystems "need not always correspond to a group of neurons that are in contiguous parts of the brain" (Kosslyn, 1994, p. 30). However, it is useful to think of these subsystems as mapped to cortex, as was done by Kosslyn (1994), and expanded on by Ganis et al. (2004). An adapted version of the model is presented in Figure 5.

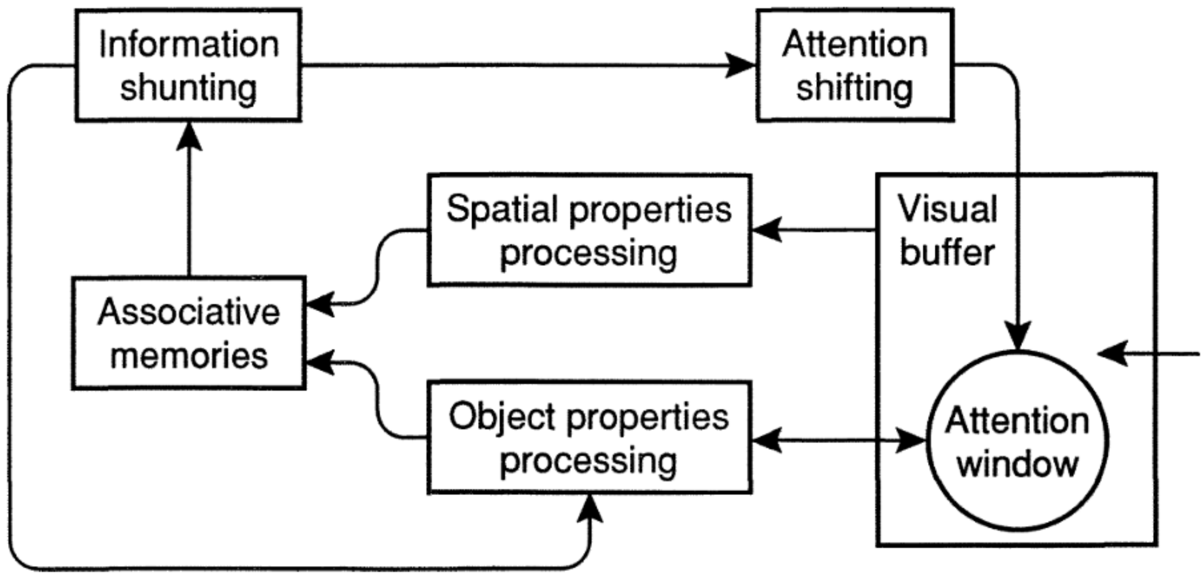


Figure 5. Model from Ganis et al. (2003) Figure 1, which in turn was simplified from Kosslyn (1994) Figure 11.1. Kosslyn's original caption: "The architecture of visual mental imagery."

Importantly, these subprocesses have been mapped to brain areas—although there is still some debate about the specifics. The visual buffer has been mapped to V1 and possibly other retinotopically-organized occipital areas. The object properties subsystem is mapped to the ventral pathway from the occipital lobe to the inferior temporal lobe. The spatial properties subsystem is mapped to the dorsal pathway running from the occipital lobe to the posterior parietal lobe. Associative memory maps to association cortex (e.g., occipital/temporal/parietal junction). Information shunting depends on prefrontal cortex, and so on.

Evidence for these somewhat-independent imagery processing parts comes from empirical studies that found differences in regional cerebral blood flow between tasks thought to tap each of these processes (Kosslyn et al., 2004) and predictable correlations between processes (e.g., Kosslyn et al., 1984). Overall, the picture that emerges is that imagery is not a general undifferentiated ability, but rather a collection of "relatively independent subabilities" (Kosslyn et al., 1984, p. 240). This fact contributes to the heterogeneity of aphantasia.

Aphantasia is Heterogeneous

Because mental imagery is a collection of cognitive abilities, it would stand to reason that aphantasia is a heterogeneous disorder, in which a person's specific deficits depend upon which processes are deficient in each individual case. Evidence for this possibility comes from both acquired aphantasia and developmental aphantasia.

In acquired aphantasia, a variety of brain lesions can lead to imagery deficits of varied types (see "Neuropsychological Discoveries of Impoverished Mental Imagery"). Most people with acquired aphantasia have other associated neuropsychological disorders, especially disorders of visuospatial perception and recognition (discussed in Trojano & Grossi, 1994). Because of these comorbidities, and because of the relative rarity of this disorder, these are not the people with aphantasia that I will be using in my studies.

In developmental aphantasia, surveys of people with aphantasia reveal that they also report differences in their experiences. One difference is whether people with aphantasia experience a deficit of multiple modalities of mental imagery, or if they experience a deficit in only the visual modality—26% of people with aphantasia claimed that they lacked all modalities of imagery in a recent survey study (Dawes et al., 2020). A different survey with a differently worded modality question worded in a different way found that around 54% of people with aphantasia described "faint" or non-existent imagery in all modalities (Zeman et al., 2020). Similarly, around 63% of people with aphantasia reported experiencing visual imagery during involuntary hypnogogic dreams (Zeman et al., 2020). These differences in modalities affected and the presence or absence of involuntary imagery could indicate different underlying subsystems have malfunctioned in the brains of people with aphantasia, suggesting a heterogeneous disorder.

According to the most prominent model of the subsystems used during mental imagery (Kosslyn, 1994, p. 383), there are multiple ways that the interaction of the imagery subsystems could break down such that deficits to mental image generation occur. According to Ganis et al., "this can result from damage to the information in long-term associative memory, to the information shunting system in the frontal lobes, to the object-properties processing system [in the inferior temporal lobe], and to the visual buffer [in the primary visual cortex]" (2003, p. 634). These explanations are more useful when investigating acquired rather than developmental aphantasia because they make references to a disruption in connectivity due to brain damage.

When looking at people with developmental aphantasia, the existing theories about where the breakdown occurs are different. Multiple factors have been theorized, including neurological differences in backpropagation or connectivity, and differences in processing costs and available processing resources. One purpose of the present studies is to make progress towards investigating such breakdowns as a possible cause of aphantasia.

Based on the reported differences in the occurrence of visual dreams, I hypothesize that there may be two major groups of people with aphantasia, each with a different neural mechanism responsible for the lack of visual mental imagery. These two rival explanations for aphantasia are reviewed below.

Faw's Perceptual/Conceptual Theory of Aphantasia

The perceptual/conceptual cognitive theory is the most straightforward way of conceptualizing what the difference is between people with aphantasia and neurotypical people when it comes to the experience of mental imagery.

As Faw put it: "perhaps conscious imagers can utilize both propositional and pictorial imaging processes, while non-imagers (or 'non-conscious imagers') utilize only the propositional model" (1997, p. 287). This perceptual/conceptual theory put forth by Faw suggests that people with aphantasia simply do not use the depictive type of representation. Faw wrote this theory before aphantasia had its modern name, which is why this quote refers to non-imagers instead.

The perceptual/conceptual theory leaves the question open as to what neural difference can account for this non-usage of depictive representation. However, one such neural model fits best with the idea that some people with aphantasia *never* experience mental imagery.

Diminished Backpropagation or Top-Down Processing Model

Information in the brain is often conceptualized as travelling in one direction (e.g., from "early" processing regions to "later" processing regions). However, information also travels in the opposite direction (e.g., from "later" processing regions to "early" processing regions), which is referred to as *backpropagation*. This series of studies is agnostic about how the brain accomplishes this backpropagation, but perhaps it is a combination of dedicated neural connections that flow from later→early regions, as well as additional signals sent towards dendrites rather than down the length of the axons in the connections that flow from early→later. Regardless, at the neural level, there may be differences in backpropagation in the visual perception systems of people with aphantasia such that mental image generation does not occur.

William James was probably the first to suggest that visual mental imagery was visual perception running backwards—the term backpropagation wasn't around back then. He wrote: "in imagination the starting-point of the process must be in the brain. Now we know that currents usually flow one way in the nervous system; and for the peripheral sense-organs to be excited in these cases, the current would have to flow backward" (James, 1890b, p. 70).

Because mental imagery may simply arise from neurons involved in perception but flowing backwards as James and others have suggested (e.g., a "reverse visual hierarchy" as in J.

Pearson, 2019, p. 625), aphantasia may result from the flow of backpropagation not making it as far as V1, the "starting point" within the brain for the flow of depictive representation.

More generally, there may be an absence of "top-down" connections or signals that causes the imagery deficit. A review of neuroimaging studies (e.g., using fMRI, EEG, etc.) found that "neural representations of imagined and perceived stimuli are similar in the visual, parietal, and frontal cortex. Furthermore, perception and imagery seem to rely on similar top-down connectivity. The most prominent difference is the absence of bottom-up processing during imagery" (Dijkstra et al., 2019, p. 423). Adding to this, Bergmann et al. (2016) found that subjective vividness of mental imagery was positively related to the volume of prefrontal cortex, suggesting the importance of top-down signals for mental imagery. Similarly, Milton et al. (2020) reported "stronger connectivity between prefrontal cortices and the visual network among hyperphantasic than aphantasic participants." These findings support the theory that people with aphantasia simply do not activate their primary visual cortex with top-down processing from the frontal cortex during imagery tasks, and therefore do not experience truly depictive imagery.

There is evidence that some regions of the cortex are involved in the generation of mental images but not in visual perception, including regions of the extrastriate cortex (Brodmann's area 19 in Kosslyn et al., 1997). Backpropagation from the inferior temporal lobe via the occipitotemporal junction (Brodmann's areas 19 and 37) to V1 (Brodmann's area 17) is theorized to be what induces image generation in the brain during mental imagery (e.g., Kosslyn et al., 2006, p. 144; Ganis et al., 2003, p. 634). The occipitotemporal junction is a region activated both by mental image generation and visual perception (Kosslyn et al., 1997). I am suggesting that a failure of backpropagation or "top-down processing" from these areas to the primary visual cortex may provide a neural explanation for some cases of aphantasia in which neither voluntary nor involuntary mental images occur. In other words—using perceptual anticipation theory—when a person without these neurotypical channels of backpropagation anticipates perceiving an object, that object is not experienced as mental imagery because their higher-order understanding of objects is not transmitted back to the primary visual cortex where it can be experienced as a depictive mental image. Therefore, not only do they not consciously experience a visual mental image, but there *is no mental image* of which to speak because it is not generated in the first place.

People with backpropagation-deficient aphantasia should report *not* experiencing visual imagery during dreams, because even when their state of consciousness changes, brain connectivity (e.g., availability of backpropagation channels) should not change. If some people with developmental aphantasia do not have the ability to generate visual mental images, then those people should be detectable using cognitive study designs that show that their responses to tasks that make use of mental imagery are different from people who do not have aphantasia by manipulating the complexity of the mental images used during the task such that using a depiction (imagery) strategy will produce one pattern of results while using a propositional (non-imagery) strategy will produce a different pattern. Studies 3 and 4 investigate this.

Implicit/Explicit Theory of Aphantasia

An alternative cognitive theory explains aphantasia as an inability to *consciously* experience mental imagery, while still allowing for the possibility that the mental imagery is occurring in an *unconscious* manner. In perhaps the first case study ever published on a person with developmental aphantasia, Botez et al. (1985) theorized that their patient who claimed not to experience mental imagery was still producing mental images, but "at a subliminal level"—in other words, below the level of the individual's awareness on an *implicit* level. As discussed in the section "Dissociations of Mental Imagery Component Dimensions," this has become known as the implicit/explicit theory of aphantasia.

The difference between *explicit* mental imagery, which a person experiences, and *implicit* mental imagery, which a person does not experience, has been proposed as a possible explanation for why the VVIQ does not seem to correlate well with the seemingly "objective tasks" of mental imagery—because the people with poor mental imagery are relying on *implicit* mental images for the objective tasks, while the VVIQ measures *explicit* mental images (McAvinue & Robertson, 2007, p. 204). This theory, that the distinction between implicit mental images and explicit mental images is what results in differences on the VVIQ (or other subjective measurements such as surveys) while not resulting in differences on objective measurements (such as cognitive measurements of accuracy on cognitive tasks) is just one possible explanation for why subjective measurement and objective measurement of mental imagery ability seem to be misaligned in the literature. For example, one research group recently speculated that they found no differences in mental rotation task accuracy between an aphantasia group and a control group because some people with aphantasia "may retain the ability to generate visual imagery, but lack conscious access to this imagery" (Pounder et al., 2022, p. 10).

However, as McAvinue & Robertson pointed out, the development of the implicit/explicit theory of mental imagery was created *because of* the mismatch between subjective tests and objective tests (2007, p. 204). If the lack of a correlation between subjective and objective measurements of imagery can be better explained by other factors—such as participants misrepresenting their visual imagery vividness (e.g., misunderstanding the questions on the VVIQ or misunderstanding where their mental imagery abilities fall on an interpersonal spectrum), or the conflation of researcher-chosen cognitive tasks that *do not* require visual imagery with tasks that do require imagery—then the theory of implicit/explicit imagery is unneeded. The misalignment of subjective and objective measurement of mental imagery is discussed in further detail in "The Misalignment of Subjective and Objective Measurements."

What the implicit/explicit theory suggests is that people with aphantasia are not *aware* of depictive representations, and yet are making use of them on an implicit level. Alan Richardson made the strong claim that people who claim not to experience mental imagery are merely not noticing it: "the fact that imagery goes unnoticed by some people almost all the time may be attributed to a variety of causes, but its potential for being noticed by anyone under appropriate conditions is unquestionable" (1983, p. 15). In his view, mental images occur implicitly in all people, and can be used to solve problems even without conscious attention paid to the "*imagery channel*" of the brain (p. 14).

Can mental imagery be representational without being experiential? Evidence from neurotypical individuals indicates that, yes, mental imagery can be implicit (e.g., Nanay, 2021; Phillips, 2014). A person does not need to be *aware* of all calculations that are happening in their brain in order to use the results of the calculations in their behavior, see for example the "filling in" of the blind spot of the retina. Perhaps the most dramatic example is that of *blindsight* (see Ajina & Bridge, 2017). People who experience blindsight report having no conscious awareness of regions of their visual field, but "subjects can be prompted to make highly reliable judgements about certain features in their blind field" such as direction of motion, or the presence and location of a stimulus (Phillips, 2014, p. 287). If we allow that aphantasia may be a kind of *blindsight for mental imagery* instead of the visual field itself, then the implicit/explicit theory is a plausible explanation of what is different about cognitive processes in people with aphantasia.

It may seem that this implicit/explicit theory resembles early propositional theories of mental imagery. However, this is not so. Propositional theories posited that explicit mental imagery may exist, but implicit mental imagery does *not* exist because the mental imagery was epiphenomenal and separate from the true underlying calculations necessary for completing cognitive tasks (e.g., Pylyshyn, 1973). As was previously established in "The Great Imagery Debate," there is ample evidence that mental images are *not* epiphenomenal and are in fact used in cognitive calculations (e.g., evidence from reaction times for mental rotation, mental scanning, etc.) and therefore purely propositional accounts of mental imagery have been disproven. On the other hand, the implicit/explicit theory of aphantasia argues that explicit mental images may *not* exist for people with aphantasia, but implicit mental images *do* exist for people with aphantasia. In other words, according to the implicit/explicit explanation of aphantasia, if people with aphantasia complete those same cognitive tasks originally used to prove that mental imagery can be depictive (e.g., rotation, scanning, etc.) then they should show a *similar* pattern of results to neurotypical people because conscious experience of the mental imagery experience is not necessary for the underlying implicit mental imagery representations to be used for cognitive functions. This idea will be tested in Study 3.

If the implicit/explicit model of aphantasia is correct, what would we expect to be happening at the level of the brain? In the following section, I argue that we would expect differences in the use of processing resources in aphantasia compared to control brains.

Diminished Processing Resources Model

It is possible that people with less vivid mental imagery have fewer processing resources than people with highly vivid mental imagery, and therefore "poor imagers, for any given task, use the processing method that requires the least amount of resources" (Riske et al., 2000, p.

141). A resource account may help explain why people with aphantasia claim no mental imagery: their brain is making use of exclusively propositional representations or implicit mental imagery in order to save on processing resources, because propositional representations are more computationally expensive in the brain. It is also possible that in the brain of a person with aphantasia, processing resources are limited to the point that an explicit depictive representation is more difficult due to those processing resources being gobbled up by other processes (e.g., language, frontal executive processes) in the brain. It has been suggested that a possible difference in the brains of people who experience low vividness of visual mental imagery may be due to "a failure to suppress activity that can interfere with vividness, for example in auditory cortex" (Fulford et al., 2018, p. 33).

This "diminished processing resources" model may be helpful in understanding people with aphantasia who experience mental imagery during dreams but not in wakeful states—a shortage of processing resources due to noise (e.g., language processing) in other parts of the brain necessitates propositional representations while awake, but enough resources are available while asleep (e.g., due to fewer active frontal executive processes) to allow for an increase in depictive representations.

In a brain where there is a shortage of processing resources or a struggle for attention, it may make sense to preserve implicit mental imagery but jettison explicit mental imagery. After all, is the conscious experience of explicit mental imagery really adding that much to the usefulness of cognitive calculations that can already be made based only on implicit mental imagery? The implicit/explicit account of aphantasia would argue that, no, consciously *experiencing* the mental imagery is unnecessary, and therefore is an area where the brain can cut processing costs.

It is worth noting that a diminished processing resources model is also potentially compatible with perceptual/conceptual theory, although in a slightly different manner. Perceptual/conceptual theory would argue that because of a person with aphantasia's reduced processing resources, they are utilizing *only* propositional representations instead of more expensive implicit depictive representations. Either way, the diminished processing resources model of aphantasia seems like a promising place for neurological studies to look in the future.

This series of studies is agnostic as to what exactly the processing costs associated with explicit mental imagery are (e.g., glucose expenditure, attention, etc.). But regardless of what the form of the cost is, Study 4 draws upon an existing perceptual theory—categorical/coordinate processing theory—that touches on how processing costs are different depending upon the properties of object perceptual processes (i.e., propositional representations require fewer processing resources than depictive representations). I will now explain how that theory will be useful for testing the perceptual/conceptual and implicit/explicit theories of aphantasia.

The Diminished Coordinate Processing Theory of Aphantasia

The depictive account of visual mental imagery, which allows both depictive and propositional representations, agrees with theories of object recognition that include multiple recognition systems, such as categorical and coordinate processing (e.g., Brooks & Cooper, 2006).

Categorical processing closely resembles *propositional* representations, such that the brain makes use of categorical properties of objects (e.g., ROUND, ANGULAR, CONICAL) and relationships between objects or parts of objects (e.g., BELOW, ABOVE, LARGER THAN), but it does not make use of specific metric distances. Frequently, this basic categorical processing is sufficient for an imaging or recognition task to be completed. For example, remembering that your book is on top of the table does not require forming a mental image of the table and book, it only requires a proposition representing that categorical information of location, “ON TABLE,” and not the depictive visual information.

However, if further detail is needed, the *coordinate* system can be used, at a higher cost of processing resources. The coordinate processing system preserves metric information, such as the relative distance between points in space. Use of this system may occur if a higher-resolution *depictive* mental image is required. For example, if the instance of object recognition depends on coordinate (metric distance) information such as when the object closely resembles two exemplars from the same category (e.g., telling the difference between two similar faces), then the brain can make use of more resources to create a depictive representation. For example, you may need to make use of a coordinate representation that preserves metric information in order to remember *where* on the table you have placed your book, or the precise *shape* of the words on the cover of the book.

More abstractly, other categorical/coordinate information includes colors—this is the difference between remembering that a book *is red* (a categorical process) versus remembering the specific shade of red that the book possesses and *experiencing via visualization* the *relative redness* of the book (a coordinate process—in this case, coordinate is referring to the position of red in color-space relative to other tints/shades of red, other colors, etc.). In theory, the categorical version of the process requires fewer processing resources than the coordinate version, because the categorical encoding (the book is red) is a simpler task than coordinate encoding (e.g., the book is a lighter red relative to the darker color of cherries).

According to the categorical/coordinate processing account of object recognition, *which* process the brain uses depends upon the task demands (e.g., Brooks & Cooper, 2006). During perception, when task demands are simple, the categorical system for recognition is used, because less information needs to be accessed and represented. It requires less computing power and is faster. In theory, a propositional representation of mental imagery would be similar—cheaper and faster in the brain. Therefore, if people with aphantasia have diminished processing resources, then during perception, people with aphantasia may *encode* information in a categorical/propositional way, and during imagery (or lack thereof), they may *recall* information in a categorical/propositional way. Such processing would result in what the perceptual/conceptual theory suggests: people with aphantasia do not use depictive representations.

According to this theory, neurotypical people without aphantasia may also use categorical/propositional representations, especially when task demands are low, but unlike people with aphantasia, they are not *limited* to categorical/propositional representations and are better able to use a depictive/coordinate representation when necessary to meet the demands of the task. For example, depictive representations seem more likely to be used when stimuli are incidentally encoded (vs. intentionally encoded) therefore creating a high-demand mental imagery task during later recall (see Reisberg & Leak, 1987). Additionally, depictive representation—as measured by activation in early visual cortex—seems more likely to occur when high-resolution details of mental images are required, or when shape-based details rather than spatial details are required (Kosslyn & Thompson, 2003). Therefore, part of the puzzle of why there seem to be individual differences in the use of depictive representations may be the increased amount of effort/resources required to generate a depictive representation versus the lower cost for propositional representation.

This coordinate processing theory inspires a way of testing the previous theories of aphantasia: that people with aphantasia are either incapable of using depictive representations (perceptual/conceptual theory) or that people with aphantasia are using depictive representations but are not conscious of them (implicit/explicit theory). All people should perform better on categorical tasks than coordinate tasks because categorical tasks are less demanding of cognitive resources. However, if people with aphantasia have a deficit to speed or accuracy when using coordinate information (i.e., when depiction is difficult) relative to categorical information all relative to controls (i.e., the negative slope of decreased performance between categorical and coordinate trials should be steeper—more negative—for people with aphantasia than for people without aphantasia), this would be evidence for perceptual/conceptual theory. In contrast, if people with aphantasia perform no differently from controls (i.e., if the main effect of better performance on categorical trials relative to coordinate trials is significant but there is no significant interaction—difference in slope—between the groups), this would be evidence for depiction occurring in the aphantasia group and therefore evidence for the implicit/explicit theory. Study 4 will test this.

There is one other influential theory of aphantasia that requires discussion.

The Mistaken/Malingering Theory of Aphantasia

The final theory in the list of explanations for aphantasia is the explanation that the people who claim to experience aphantasia are experiencing no such thing. Some researchers go so far as to hypothesize that people with aphantasia are malingering their deficit, or are misunderstanding what mental imagery is, as in the following quote by Smith (1978).

> What are we to say, then, about these people… who seem to be quite normal in every way, and yet simply deny that they ever have images at all? There seem to be three

> possible alternatives: (1) They lead what would seem to be strangely empty mental lives (in the way that we might feel that congenitally blind and deaf people must lead empty mental lives)... (2) They are lying in order to pose as 'superior intellects' or simply to be perverse. Or (3) They have misunderstood the nature of imagery, and their denial that they have images is the result of their misunderstanding, much as the denial by a respectable housewife that she has erotic impulses would be the result of her misunderstanding. (p. 176)

Some evidence against malingering is suggested by Keogh and Pearson (2018) in that people with aphantasia do not simply respond with the lowest possible response option on all subjective measures of mental imagery. When people with aphantasia were given both the VVIQ and a spatial imagery questionnaire (the Spatial OSIQ; Blajenkova et al., 2006), people with aphantasia "reported having spatial imagery in the OSIQ, which one would not expect if participants were merely saying they cannot imagine anything" (p. 59). Therefore, people with aphantasia did not seem to be attempting to maximize their deficit on paper.

Short of calling people with aphantasia liars, other researchers have suggested that it could be a problem of metacognition or memory. Perhaps it is not that people with aphantasia cannot form mental images, but rather they "do not believe they are visited by them, do not make use of them… [or] do not remember them" (Dennett, 1978, p. 176). Katz (1983/2014) made a memorable statement about this:

> Of course, if we are forced to depend on introspective reports we are also forced to face one of the oldest bugaboos of cognitive psychology. Just because people say they are using imagery, it doesn't necessarily follow that they are, in fact, doing so. (p. 45)

Katz went on to express confidence that "subjective imagery tests will ultimately prove to be particularly good predictors of imagery-mediated performance" following more work in this area (1983/2014, p. 45).

This possibility of a metacognition problem will be further discussed in the section on "The Misalignment of Subjective and Objective Measurements." In contrast, some evidence in favor of the existence of aphantasia was discussed in the previous section called "Evidence for Non-Depictive Representations in Aphantasia."

I myself have deficient mental imagery, a fact that I became aware of only when I was explicitly learning about the scientific study of mental imagery. Some of the tasks that have been used to demonstrate the depictive nature of mental imagery, such as mental zooming (i.e., making an image bigger in the mind's eye) confounded me because I could not achieve that seemingly perceptual effect in my imagination. This realization that my mental imagery was weak led me to aphantasia as a topic of study. At first, I could not tell you whether my deficit was the result of poor metacognitive awareness of imagery, misunderstanding of other people's descriptions of what mental imagery is, or poor memory of instances when I use mental images. I felt very suspicious of my own self-reported deficit in mental imagery, and did indeed suspect myself of misunderstanding at first. However, after an extreme effort to evaluate the available evidence (i.e., writing a dissertation about it), I am convinced that individual differences in mental imagery are just as real as individual differences in perceptive acuity such that some people truly are blind in the mind's eye. I do occasionally use simple imagery, something that I am now hyper-aware of when it happens, but I would put myself extremely low on the bell-curve of mental imagery ability. If I may coin a new term, I would consider myself a *dysphantasic*, following the pattern of other naming monikers such as "dyslexia" and "alexia."

Further understanding of why this feeling of mistakenness or malingering exists at the level of the literature on mental imagery requires a deeper look into the measurement of mental imagery, which is where I will move to next. The measurement of mental imagery is cloudy—and therefore suspicious—for multiple reasons.

The Struggle of Measurement

Individual differences research for visual mental imagery relies upon self-report techniques ("subjective" measurement) and cognitive findings ("objective" measurement). As we will see, these techniques frequently lead to mixed findings because they do not agree with one another in a tidy fashion.

Subjective Measurement of Imagery Ability

Self-report measurements of mental imagery dimensions have been vilified by numerous researchers. One widely cited paper went so far as to claim that "introspective reports, as measures of imagery, do not have construct validity" (Di Vesta et al., 1971). Despite this, subjective measurements are still in widespread use today, partly due to convention and ease of administration. I will discuss the shortcomings of subjective measurement of mental imagery, especially metacognitive considerations of the introspection required to analyze a person's own mental imagery abilities, a bit later (see "The Participants are Wrong: The Untrustworthiness of Self-Reporting Mental Imagery"). For now, I will discuss how subjective measurement of mental imagery has been done, focusing on a particular questionnaire in widespread use: the VVIQ.

Many self-report measures for the previously explained dimensions of mental imagery exist, including the Questionnaire on Mental Imagery (QMI) for modality (Sheehan, 1967), the Vividness of Object-Spatial Imagery Questionnaire (VOSIQ) for the object/spatial dissociation (Blazhenkova, 2016), the Vividness of Visual Imagery Questionnaire (VVIQ) for vividness

(Marks, 1973, 1995), the Test of Visual Imagery Control (TVIC) for image controllability (Gordon, 1949), and many others (see D. G. Pearson et al., 2013 for a review of imagery measures).

No subjective measures were specifically developed to ascertain whether a person lacked capabilities of generating mental imagery, as is the case for people with aphantasia. In fact, it has been pointed out that the classic self-report questionnaires "were not developed on the basis of a theoretical analysis of imagery ability" at all (McAvinue & Robertson, 2007, p. 205)! No specific *aphantasia* measure has been created, let alone one that is theoretically motivated.

Instead, the VVIQ (See Appendix A) has been retrofitted to serve the purposes of an aphantasia questionnaire. The VVIQ was originally created by Marks (1973) as a way to measure "individual differences in verbal reports of imagery vividness" (p. 17). The VVIQ asks 16 questions about imagery vividness, including questions about the clarity of mental images of faces, colors, and shapes. The original VVIQ had people rate each of the 16 prompts from 1–5, with 1 representing "perfectly clear and vivid as normal vision" and 5 representing "no image at all, you only 'know' that you are thinking of the object" (Marks, 1973, p. 18).

McKelvie (1995) wrote extensively about the validity and reliability of the VVIQ, which was quite poor by some metrics, and gave several suggestions for improvements to the instrument (see p. 83). It was then revamped by Marks (1995) to become the VVIQ-2, which has 16 additional questions and a flipped scale such that a rating of 1 now represented the lack of mental imagery while 5 represented perfect clarity.

The VVIQ, the VVIQ-2, and a third version called the VVIQ-R (or VVIQRV) that features more of McKelvie's suggested changes, have all been used, although directions given to participants have been inconsistent (see J. T. E. Richardson, 1999). Researchers may ask participants to complete the instrument with their eyes open, with their eyes closed, without explicit instructions for eye closure, or ask them to first complete it with their eyes open and then complete it again with their eyes closed (e.g., see the ‘modified’ versions of the VVIQ used by Amedi et al., 2005; Crowder, 2018). Researchers may use different Likert scales (e.g., from 1–4 as in Dijkstra et al., 2017; from 1–5 as in Crowder, 2018; from 1–6 as in Walczyk & Taylor, 2000; from 0–7 as in Eberman & McKelvie, 2002), multiple versions of the VVIQ (e.g., Campos & Perez-Fabello, 2009), and use the scale in either direction such that vivid mental imagery is either represented by the highest number or by the lowest number. No version of the VVIQ is administered consistently, even though Marks provided explicit instructions in his 1973 and 1995 versions of the instrument. Despite all of these issues with consistent administration, the VVIQ is considered to have acceptable psychometric properties when used properly, including a single underlying factor and reasonably high reliability (i.e., test-retest, Cronbach’s Alpha, and split-half are acceptable, see McAvinue & Robertson, 2007, p. 193; McKelvie, 1995).

A sample of 1288 participants showed that VVIQ scores are negatively skewed (see Figure 6; Zeman et al., 2020). An analysis of 38 samples that took the VVIQ showed significant skew such that the average response is above the midpoint of the scale—a 95% confidence interval places the mean between 55.6 and 62.5 on the most commonly used scale that ranges from 16 (no mental imagery vividness) to 80 (maximum mental imagery vividness; confidence interval was converted from McKelvie, 1995, pp. 13–14). Other mental imagery measures have also produced negative skew such that many people report vivid mental imagery and few people report mental imagery that is low in vividness (see A. Richardson, 1994; Betts, 1909; Faw, 2009). Whether this skew is an artifact of measurement (e.g., it may be due to low metacognitive access to placing your own vividness relative to the vividness of others), or if this is truly how the visual mental imagery abilities of people are distributed is a mystery that I will attempt to illuminate in Study 2. Negative skew in the distribution is not necessarily a problem because it *may* reflect the true distribution of scores, but it will be interesting to see if the distribution changes once participants have more information about how *other* participants visualize, such that they may be able to more accurately place themselves in the vividness of imagery spectrum when completing the VVIQ.

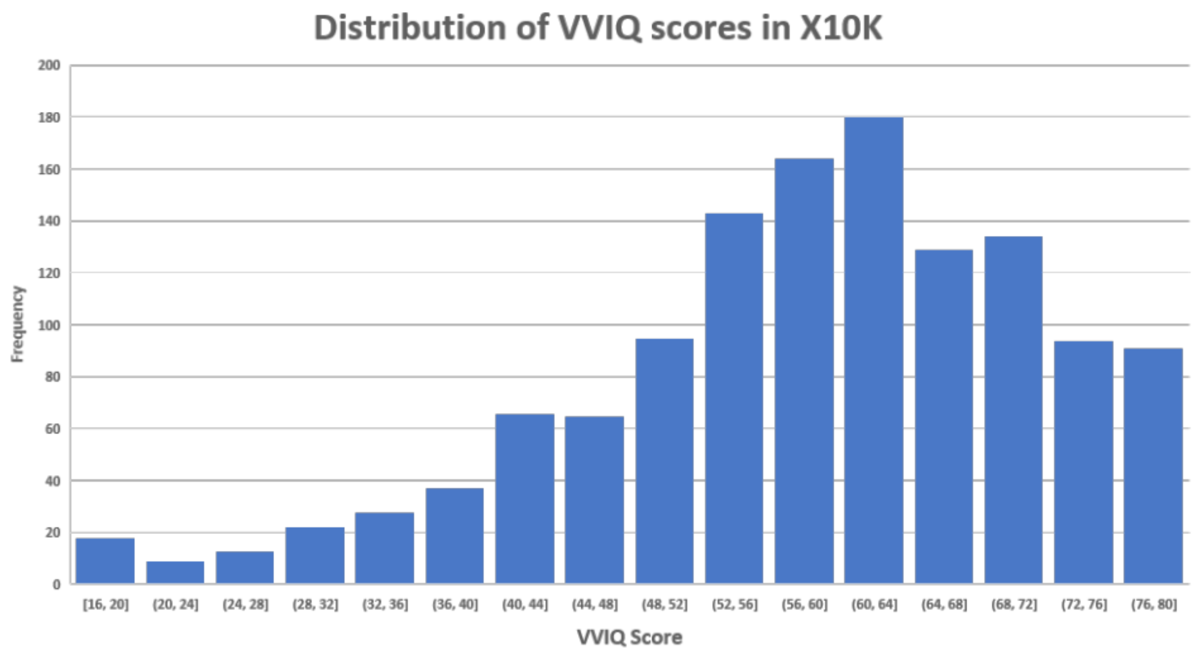


Figure 6. Adapted from Figure 8 of Zeman et al., 2020. This bar graph shows the skewed distribution of VVIQ scores. The skewness here is approximately -.74, indicating a moderately negative skew (Bulmer, 1979).

Aphantasia as VVIQ Score

Even before aphantasia was named, it was common for studies investigating individual differences in imagery vividness to use the VVIQ to identify extreme groups—but the exact method varied considerably. This variance is partially due to the fact that the VVIQ was designed as a *continuous* measure, and not as a way to sort people into *discrete groups* such as people with aphantasia versus hyperphantasia. The most straightforward way to create discrete groups was to include an equivalent number of lowest scorers and highest scorers following mass testing of potential participants, which is how the developer of the VVIQ approached it (Marks, 1973; Cairns & Coll, 1977). Occasionally, participants scoring in the bottom half of the VVIQ scale were binned together as "poor imagers" while high-scoring participants were binned as "vivid imagers," creating unequally-sized groups (Riske et al., 2000). In some cases, a median split produced equally-sized groups of "good" imagers and "poor" imagers (e.g., Berger & Gaunitz, 1977; Gur & Hilgard, 1975). More creative splits were also engineered, including a top-third/bottom-third extreme groups design that threw the middle third of scores out of the analysis entirely (Walczyk & Taylor, 2000). Sometimes seemingly arbitrary cut-off points on the scale were chosen *after* the scores of the participants were known, such as in Dobson & Markham (1993) where the "upper 25 per cent of the range [of the VVIQ]" and "lower 20 per cent" were used for grouping. In one case, the cut-off for non-VVIQ "picture imagery" scores assigned to each group was placed after the scores were known such that the design "included the maximum number of subjects while eliminating any overlap in the picture imagery scores of the two groups" (Johnson et al., 1979). In one case, the original design (extreme groups of the upper and lower 24% of scorers) was dropped in favor of scrambling to find enough participants "whose vividness scores fell into the appropriate ranges" (McKelvie & Rohrberg, 1978, p. 453).

The naming of aphantasia did not immediately improve how "poor imagers" are defined. Extreme group designs are still used in some cases (e.g., Fulford et al., 2018). The strict definition of aphantasia as having zero mental imagery—and a score of 16 on the VVIQ—has been used in some studies (e.g., Crowder, 2018). However, other studies that purport to use people with aphantasia as participants have chosen other operational definitions based on various cut-off points on the VVIQ (e.g., a score of 25 or below in Bainbridge et al., 2021 and in

Pounder et al., 2022). Further still, some studies allow people with aphantasia to self-identify and no cut-off point on the VVIQ is applied at all (e.g., Dawes et al., 2020; Keogh & Pearson, 2018). Therefore, while the VVIQ—and therefore *vividness*—is frequently used in conjunction with aphantasia studies, there is no single consistent way that this subjective measurement of imagery vividness is used in order to bestow the label of aphantasia to a particular person. However, a categorization scheme was recently suggested by Zeman et al. (2020)—which was a large-scale survey including around 2400 participants. According to that scheme, here is how VVIQ scores should be broken into distinct groups: a score of 16 is extreme aphantasia, 17–23 is moderate aphantasia, 24–74 is typical mental imagery vividness, 75–79 is moderate hyperphantasia, and 80 is extreme hyperphantasia. These definitions are arbitrary and potentially imperfect, but they will hopefully provide a consistent categorization scheme for future research, or at least serve as a starting point for further exploration of the category boundaries.

Objective Measurement of Imagery Ability

As previously mentioned, objective (task-based) measurement of mental imagery ability was attempted even in pre-behaviorism times (see Fernald, 1912, for an early discussion of "objective" measurement of mental imagery; see Kosslyn & Jolicoeur, 1980; D. G. Pearson et al., 2013; Woodworth, 1938, for additional reviews of objective measurements of imagery). More recent efforts are discussed below. I focus on objective tasks that have been used in conjunction with the VVIQ, because the VVIQ is the subjective measure used in the present studies.

One task that emerged both as a pre-behaviorism task and (seemingly separately) as a post-behaviorism task is the backwards spelling task, also called "reverse spelling task" or "spell-back latency" (Fernald 1912; Walczyk & Taylor, 2000). In this task, participants are asked to spell a word backwards after it is read out loud. Participants may speak their answers out loud or write the letters down. Fernald (1912) showed that there were many strategies that participants used while performing this task, and some of them were more dependent on visual images than others. Fernald wrote that they did not find “any convincing correlation between rate of spelling and the use of visual imagery” (1912). Notably, however, Fernald did not have a subjective mental imagery test to compare the results to in order to examine self-reported individual differences when compared to objective scores. Walczyk & Taylor (2000), on the other hand, did find a significant difference. They used a modified version of the VVIQ (i.e., they increased the Likert scale from 5-point to 6-point) and correlated it with both response time and accuracy for spelling words backwards out loud. The correlation for response time was -.25 such that people with higher scores on the VVIQ responded faster. The correlation for accuracy was .21 such that participants with higher scores on the VVIQ correctly spelled more words. Eighty-two of their eighty-nine participants reported that they had formed mental images of the words while spelling them. The words used are presented in Appendix B. A version of this backwards spelling task will be used as part of Study 2.

One task that has been theorized to correlate with visual mental imagery ability is the snowy pictures task, also called the “degraded pictures test” (created in Ekstrom et al., 1976; used in Blajenkova et al., 2006; Burton & Fogarty, 2003; Kozhevnikov et al., 2005). In this task, line drawings of objects that have been degraded (i.e., made harder to identify with the addition of a grainy “snow” pattern) are identified by the participant. It was originally developed as part of a battery of cognitive tests in 1976, although the intended purpose was not explicitly for mental imagery research (Ekstrom et al., 1976). Notably, Kozhevnikov et al. (2005) claimed that some participants in their experiment had significantly better accuracy and reaction times for this test, and those participants also tended to have high scores on the VVIQ, although they did not explicitly correlate these findings with VVIQ scores. Blajenkova et al. (2006) found that accuracy on the snowy pictures test was positively correlated at .26 with VVIQ scores. However, Burton & Fogarty (2003) found a correlation of -.04 between accuracy on the snowy pictures test and the VVIQ. Importantly, the version of the snowy pictures test used by Kozhevnikov et al. and Blajenkova et al. was modified to be more difficult than the version that appeared in both Ekstrom et al. and Burton & Fogarty. A version of this task was used as part of Study 2, and is presented in Appendix C.

The tail length test (TLT) is a frequently used objective measurement of mental imagery ability. This test requires participants to respond “short” or “long” in response to a list of animal names, regarding the relative length of their tail to their body (a standard version was created by Policardi et al., 1996; versions were used in Behrmann et al., 1994; Farah, Hammond et al., 1988; Milton et al., 2020). Animals with tails that are highly associated with their names are not used (e.g., peacocks, beavers). Performance on this task was shown to differ significantly when extreme groups were compared such that participants with extremely high VVIQ scores performed more accurately than participants with extremely low VVIQ scores, although no overall correlation was reported (Milton et al., 2020). People with acquired aphantasia have been shown to perform significantly worse in terms of accuracy on the TLT when compared to neurotypical controls (e.g., Farah, Hammond et al., 1988). The items on this test are presented in Appendix D. This task will be used as part of Study 2.

Some objective investigations of mental imagery were not originally intended to measure individual differences in mental imagery ability. Three such tasks were discussed previously in the “Great Imagery Debate” section: mental rotation (Cooper & Shepard 1973), scanning (Dror

& Kosslyn, 1994), and dynamic visual noise (Borst et al., 2012). Those tasks were used to show that mental images are depictive, but not used to discern individual differences in mental imagery ability. To my knowledge, an investigation of mental scanning using the modern definition of aphantasia has not yet been conducted. I will be using a scanning task in conjunction with aphantasia in Study 3.

Mental rotation tasks have not been shown to correlate well with mental imagery ability in the general population (see Reisberg, 2013), but at least one study with a strict definition of aphantasia (VVIQ score of 16) *did* find a significant difference such that people with aphantasia were slower compared to controls in responding to a mental rotation task when speed was measured as the ratio of the number of trials completed divided by the number of time "blocks" used to complete the items (see Crowder, 2018). At least one research group has reported that they did not find a difference in mental rotation task accuracy or response time when comparing people with VVIQ scores at or below 25 with people scoring 35 or higher on the VVIQ (Pounder et al., 2022). Notable, however, was that there were only 20 participants with aphantasia and 20 controls enrolled in the study, and the statistical test for a mental rotation response time difference between the aphantasia group and the control group barely missed the significance cut-off with an effect size that would conventionally be considered between medium and large ($F(1, 38) = 3.62$, $p = .07$, $\eta_p^2 = .087$).

Additionally, dynamic visual noise tasks require participants to visualize and maintain mental images of letters, which is not something that people with aphantasia can necessarily attempt, and could easily malinger, leading to inconclusive results. Therefore, the dynamic visual noise task and all tasks that require maintained revisualization (e.g., binocular rivalry paradigms) are avoided by the present studies.

There are also a few classic neuropsychological tasks, such as the Minnesota form board task (Likert & Quasha, 1970) and the paper folding task (French et al., 1963) that were used as more general cognitive tasks but were sometimes interpreted as evidence against the preservation of mental imagery when brain damaged patients were unable to complete them effectively. Such tasks may occasionally still be used in acquired aphantasia cases, but as will be discussed later, they do not seem to be highly correlated with general mental imagery abilities in the general population (see Burton & Fogarty, 2003). Therefore, those classic tasks will not be used in this set of studies.

Another task that has been assumed to use mental imagery is the curved segments task (created in Coltheart et al., 1975; used in Bridge et al., 2012; Milton et al., 2020; Policardi et al., 1996). This task requires a person to indicate whether a given capital letter contains at least one curve. Policardi et al. found that their participant with acquired aphantasia had lost the ability to do this accurately when compared to controls. However, Bridge et al. found the opposite, that this ability was preserved in a person with acquired aphantasia. Milton et al., administered this test to people with varying VVIQ scores but did not find significant differences, most likely because they had a ceiling effect in which most participants correctly answered on all trials. If Milton et al. had examined reaction time in addition to accuracy, they might have found an effect. As such, there is no current evidence that scores on this task correlate with scores on the VVIQ, and I will not be making use of this task in my series of studies.

Color, as a stimulus category, has been an occasional avenue of research into mental imagery ability. Researchers have correlated performance on the VVIQ with long-term memory accuracy of the colors of objects. In Reisberg et al. (1986), participants were asked to match their memory for colors of everyday objects (e.g., stop sign, green traffic light) to swatches out of a color swatch book. In that study, performance on the color-matching task was correlated with VVIQ at -.47, meaning that participants with *higher* vividness of imagery scored *worse* on the color matching task. Similarly, Heuer et al. (1986) found a correlation of .57 between error magnitude in chosen color during a match task to memory and score on the VVIQ—also indicating that participants with higher mental imagery vividness were worse at the color matching-to-memory task.

It should be noted that in both cases where a correlation indicated that superior mental imagery was related to worse color matching, that multiple-choice color options were presented for a *recognition* task (i.e., the possible answers were presented as colored squares to be chosen from). The nature of the task is important because this changes the task from a pure mental image generation *recall* task to a more general recognition memory task in which even participants with poor mental imagery could match the presented colored swatches to their memories rather than generating the correct colors themselves. Theoretically, if the problem the people with poor mental imagery are experiencing is partially due to poor backpropagation from higher visual areas to lower visual areas (e.g., from the color center in V4 back to a "visual buffer" in V1) then providing participants the opportunity to perceptually view the correct answer in a recognition task is tapping a different connectivity stream in the brain (i.e., the stream from V1 to V4). Therefore, match-to-memory recognition tasks in which the stimulus is presented may not tap mental imagery to the extent that such "objective" tasks may seem to at first glance. Study 4 will attempt to avoid a pure recognition task in favor of a task requiring image generation (a change identification recall task) that makes use of not only color, but also form, orientation, and position.

People with acquired aphantasia have been given versions of a color task that did not perceptually present color options to choose from, such as the version of the task used in Experiment 3 of De Vreese (1991). In such cases, people with acquired aphantasia sometimes show deficits for long-term color matching (e.g., Farah, Hammond, et al., 1988) and sometimes do not show a major deficit (e.g., Bridge et al., 2012). Another testing approach using color is a color congruency task that requires participants to react to matched or mismatched colors/words. When this was done such that participants had to react to a displayed color word (e.g., the word purple) on a matched or mismatched color background (e.g., an orange background or a purple background), participants with higher VVIQ scores showed a lower index of performance (congruent minus incongruent performance %) indicating that better mental imagery vividness assists in performance on congruent trials (Cui et al., 2007). However, that study had only 8 participants and had extreme range restriction on VVIQ scores. In a similar vein, Wantz et al. (2015) did not find a significant correlation between VVIQ score and a person's ability to react to congruent/incongruent presentations of color—notably, in that study they required their participants to imagine one color while presented with a different color, which may make it the more valid finding compared to Cui et al.

In sum, the objective measurement of mental imagery ability through the use of color as stimuli is mixed in its findings. Color is arguably a useful modality for examining visual mental imagery as opposed to spatial mental imagery due to its nature as a purely visual type of stimulus (for a discussion of this, see Farah, Hammond, et al., 1988). However, because the VVIQ contains questions about general *form* that do not necessitate the generation of color, it does not make sense for my studies to restrict themselves to color as the modality of objective measurement—although color will make an appearance in Study 4.

Reisberg and Leak (1987) reported a counterintuitive correlation between VVIQ score and memory for features of famous faces. Participants were asked to mentally image the faces of two celebrities and decide which had a particular feature (e.g., a broader nose, bushier eyebrows, etc.). The correlation of -.47 was negative, such that participants with higher VVIQ scores performed less accurately. I have not found a replication of this surprising finding. Additionally, this task made use of well-known celebrity faces from the 1980s, and using a similar task would require updated stimuli. Therefore, this task is not a good fit for my series of studies.

There are a number of semi-objective tasks that people with less vivid mental imagery claim to be worse at (e.g., autobiographical memory reporting in Palombo et al., 2018; binocular rivalry priming in Keogh & Pearson, 2018; people with aphantasia recalled fewer objects and colors when drawing details of studied scenes from memory than control participants in Bainbridge et al., 2021). However, these tasks are not good choices for "objective tasks" because they rely too heavily on self-reporting. Performance on such pseudo-objective tasks may be a simple matter of malingering such that people claiming to have poor mental imagery could artificially deflate their performance—intentionally or unintentionally. This possibility that participants are sometimes wrong in their self-reports is discussed below.

The Misalignment of Subjective and Objective Measurements

As discussed above, both subjective measurements and objective measurements have been used extensively in the historical examination of mental imagery. However, the findings from one rarely align with the findings of the other. Many researchers expect that people reporting higher vividness on mental imagery questionnaires like the VVIQ should show faster response times and higher accuracy on tasks that make use of mental imagery. As discussed previously in the section about the definition of vividness, accuracy and vividness are not always the same thing—but the misalignment issue goes deeper than that. As early as 1969, there have been discussions of potential methodological issues leading to the mismatch of "subjective reports of mental imagery" and "objective neurological, physiological or behavioral indices" (A. Richardson, 1969, p. xi).

Since then, the literature remains full of discussions of mixed and equivocal findings (see discussions in McAvinue & Robertson, 2007; McKelvie & Rohrberg, 1978; Reisberg, 2013; Reisberg & Leak, 1987). As an example, McAvinue and Robertson noted that, historically, "imagery ability was measured through subjective tests… and objective tests of spatial ability. These two kinds of test were found to be largely unrelated to one another" (2007, p. 205). Despite creating the VVIQ, Marks himself admitted that "attempts at correlating self-report measures of imagery with cognitive performance have frequently led to mixed or equivocal results" (Marks, 1977, p. 280).

For example, Reisberg pointed out that some people who claim to have no mental imagery can successfully complete tasks that are purported to use mental imagery, while people who claim exceptional mental imagery don't seem to have a clear advantage in response times for the classic mental imagery tasks of mental rotation or mental scanning (2013, p. 382). Even the most robust investigation into how the VVIQ is related to cognitive measures conducted by McKelvie (1995) finds only very spotty and unconvincing relationships between scores on the VVIQ and tasks involving visual memory, spatial transformation (e.g., mental rotation), and other tasks where a relationship tends to be expected by researchers.

The subjective surveys and objective tasks themselves seem to have decent reliability and validity (see D. G. Pearson et al., 2013 for an overview of subjective and objective tasks historically used in mental imagery research). The problem comes when the subjective and objective measures are combined. This summarizing quote comes from McAvinue & Robertson (2007):

> To measure imagery, then, there have been two kinds of measure available, subjective and objective. Both have been employed extensively and both have been accepted, by those who use them, as having at least adequate levels of reliability and validity. Given these facts, many researchers have been puzzled by the finding that these two types of imagery measure are, on the whole, unrelated to each other. (p. 203)

In fact, the status of the visual imagery measurement literature is altogether worrying, as the "disorganized smattering" of correlations that are rarely replicated suggests that there is "no substantial relationship between subjective measures of imagery and actual patterns of imagery use in cognitive tasks" (Schwitzgebel, 2011, p. 48).

As Slee (1988) pointed out, early theories and experiments did not account for whether the subjective and objective tasks were well-matched (e.g., did they truly both make use of vividness?). Indeed, "factor analytic studies have consistently shown that self-report and objective tests load on separate factors" (McAvinue & Robertson, 2007, p. 203). Many studies have reported this (e.g., Burton & Fogarty, 2003; Di Vesta et al., 1971; Lequerica et al., 2002; McLemore, 1976; Poltrock & Brown, 1984). One study with a sample of 213 participants measured 41 variables, including scores on the VVIQ and commonly used mental imagery tasks (Burton & Fogarty, 2003). The resulting correlation table showed a poor relationship between the VVIQ and some tasks that had been previously used to assess mental imagery. As examples, the Minnesota form board task (Likert & Quasha, 1970) and the paper folding task (French et al., 1963) each had a correlation of .13 with VVIQ scores (but they may be well-correlated with self-reported spatial imagery; Dean & Morris, 2003). More concerning was the finding that mental rotation reaction time (as discussed earlier) correlated with the VVIQ at -.03, which could indicate that mental rotation tasks do not correlate well with vividness of mental imagery, or something is wrong with the system of measurement, or both.

There appear to be several reasons why self-report questionnaires (subjective measurement), most notably the VVIQ, do not align with cognitive/behavioral findings (objective measurement). The reasons fall into two major categories of explanation, although other variations of these explanations have also been offered (see Schwitzgebel, 2002).

The first major misalignment explanation is that self-reporting the subjective experience of mental imagery is difficult and untrustworthy. There are many possible reasons for these poor self-reports, including low metacognitive access, differing interpretations of levels of imagery abilities due to the subjective nature of imagery, demand characteristics/social desirability, and malingering.

The second major misalignment explanation is that “objective” measurements are used incorrectly. Many tasks designed to use mental imagery may not *rely* on mental imagery, and instead allow for successful completion with rote memorization/prior knowledge, propositional knowledge, or other intact cognitive abilities. Additionally, as the implicit/explicit theory predicts, some tasks may be accomplished with implicit (i.e., not consciously experienced) imagery, unlike the vivid explicit imagery that is purportedly measured by the VVIQ.

These two reasons are summarized below as “the participants are wrong” and “the measurements are wrong.”

The Participants are Wrong: The Untrustworthiness of Self-Reporting Mental Imagery

At most, a person has access to information about vividness of mental images within their own brain. At least, they have access to *zero* information about mental imagery vividness, if metacognitive knowledge of mental imagery vividness via introspection is impossible for that person (or at a more philosophical level, impossible for all people). Either way, this level of knowledge about mental imagery vividness does not inspire confidence in the ability of participants in subjective research to accurately pinpoint their mental imagery experiences *relative to the experiences of others*, as is requested by the VVIQ and other imagery inventories, for they have no direct access to the experiences of others, and they may not even have direct access to their own experiences.

Indeed, as mentioned, people who discover that they possess impoverished mental imagery abilities most commonly report that they became aware of the deficit after the age of 20, most often following a conversation with someone who has different mental imagery abilities from their own (Zeman et al., 2020). Discovering such a difference in imagery ability so late in life reflects the inscrutable nature of mental imagery in the brains of others. J. T. E. Richardson summarized this well: "each person can only experience his own mental images… and since mental imagery is qualitatively different from any other sort of experience, the subject has no absolute or intersubjective criteria for evaluating the vividness of his experienced imagery" (1980, p. 124). J. T. E. Richardson went on to say that because individuals cannot adequately judge vividness, then "it follows that comparisons among experimental subjects in terms of their ratings of evoked mental imagery are neither valid nor meaningful, and it is quite unsurprising that they should fail to predict performance" (p. 125; see also Kaufmann, 1981; J. Pearson et al., 2011). Kosslyn and Jolicoeur summarized the general feeling towards self-rating mental imagery ability like this:

> The problem with self-rating techniques is clear: there is no way to be sure (1) that everyone knows the referent of the word *image*, or (2) that everyone sets his or her

> criterion to the same level in assessing images. Further, as Sheehan and Neisser have demonstrated, this technique seems especially susceptible to demand characteristics, response biases, and the like. (1980, p. 141, referencing Sheehan & Neisser, 1969)

J. T. E. Richardson (1980) agreed that demand characteristics and experimenter effects "might well operate" when measuring mental imagery using "introspective questionnaires" (p. 122). Farah also agreed that experimenter effects were a problem for the self-reported measurement of mental imagery (1988). Additionally, VVIQ scores are positively correlated with measurements of social desirability (with the Marlowe-Crowne Social Desirability Scale in McKelvie, 1995, p. 18; and within a factor analysis of imagery tests, vividness of mental imagery was heavily weighted on social desirability in Winograd et al., 1998; see also Di Vesta et al., 1971). Participants are biased towards responding in a way that indicates that they experience greater vividness because of "a general demand characteristic that an image will be experienced" (McKelvie, 1995, p. 9). It has been argued that: "introspective reports, as measures of imagery, do not possess construct validity… they must be considered to be confounded with response bias" (Di Vesta et al., 1971). There is also evidence that self-reported mental imagery scores can be pushed around (i.e., made to change) due to minor variations in instructions (e.g., Ahsen, 1993; McKelvie, 1995, p. 7; J. T. E. Richardson, 1980, p. 122).

Therefore, increasing participant ability to place themselves on a mental imagery ability spectrum relative to others, as well as reducing demand characteristics, is essential for any study involving the subjective measurement of mental imagery. To do this, researchers can emphasize that more vivid mental imagery is not necessarily better than less vivid imagery, give examples of the experiences that different people have with mental imagery vividness, and emphasize that some people do not experience much—if any—mental imagery.

There is some evidence that metacognitive ability to judge mental imagery vividness can be trained. In one study, it was found that a person's metacognitive ability to judge the vividness of their own mental imagery can improve with training, such that participants' self-ratings of their own imagery vividness better predicted perceptual bias in a binocular rivalry paradigm as the number of trials they completed increased (Rademaker & Pearson, 2012). In other words, as participant experience increased, they became more likely to report experiencing vivid mental imagery on trials in which the vivid mental imagery seemed to bias their responses to the binocular rivalry presentations. In other words, the "subjective" and "objective" measurement became more aligned. Thus, self-ratings of mental imagery vividness can be improved.

If the reason that subjective measurement and objective measurement of mental imagery ability are in misalignment is because of errors made by participants on the subjective measurements, then perhaps fixing the misalignment problem is as simple as providing subjective measurements that are less open to interpretation. Providing more information on which to make their subjective judgments may restore the expected alignment with objective measurements. This possibility will be explored in Study 2.

The Measurements are Wrong: Alternative Methods of Arriving at the Correct Answer and Misaligned Tasks

If there are people who claim that they cannot use mental imagery, then shouldn't their performance on tasks that make use of mental imagery be significantly worse than the average person's performance? The answer to this question intuitively seems like it should be "yes," but a problem exists. There does not appear to be any such thing as a "pure" mental imagery task. Frequently, there are alternative methods for arriving at the correct answer on so-called "visual mental imagery" tasks, whether via semantic knowledge or non-visual spatial representations.

Many tasks previously used as imagery tests are “by no means pure ‘measures of visual imagery’: other sources of knowledge, both general semantic knowledge and extra-visual but modality specific information, for example kinaesthetic imagery, can influence performance” (Milton et al., 2020, p. 36). Farah (1988) called this imagery-substitution approach in which the use of mental imagery is circumvented the “tacit knowledge” account for the mismatch between subjective mental imagery and objective mental imagery measurements. Some objective measurements of mental imagery do not actually require the direct use of vivid mental imagery in order to successfully complete them in a timely manner. Walczyk and Taylor summarize this argument well: “the self-reported ability to form vivid visual images often does not correlate with performance measures on imagery-dependent tasks. It may be that a number of tasks thought to involve imagery only involve imagery indirectly” (2000, p. 177).

Consider the curved segments task discussed earlier (Coltheart et al., 1975). Is it possible that some people simply *know* that the capital letter “B” contains curves while the capital letter “E” does not? Does it truly require bringing the form of the letter to mind, or is semantic information about the letters enough to accurately complete such a task? The most extreme version of “the measurements are wrong” hypothesis is that *no* tasks use pure mental imagery, and that it is entirely useless (e.g., Winch, 1908).

It does seem that people who have suddenly lost the ability to visualize (acquired aphantasia) retain a surprising number of faculties that one might assume require mental imagery. As Brain (1954) remarked:

> …perhaps the most surprising feature is how little the loss of voluntary visualization impaired functions in which visual imagery might have been expected to play some part…. It would seem, therefore, that a patient who has no power of voluntary

> visualization can, nevertheless, recognize objects and persons, accurately propositionize about them… and also reproduce objects graphically. (p. 290)

Indeed, as far back as Galton (1883), a person with aphantasia reported being able to draw while "assisted by trial and error on the paper or canvas" and relying on recognition from memory to determine if the drawn lines should be kept or erased (p. 91). Even functions that may seem to rely on mental imagery such as mental rotation can be completed using propositional (verbal) methods. This was discussed in Bethell-Fox & Shepard (1988) when some participants "who reported having difficulty" in a mental rotation task that used matrices of black-or-white boxes offered that they had been memorizing the locations of black boxes by naming their locations (e.g., "upper left, middle left") and then "working out the effects" of the rotation, avoiding mental imagery (p. 20).

Additionally, there are multiple conflations that can be made when choosing an objective measurement to match a subjective measurement. One example of this, as mentioned previously, is conflating visual imagery with spatial imagery.

The most commonly-used measure of mental imagery ability, the VVIQ, does not measure *spatial* imagery, while many of the classic tasks used to objectively measure a person's imagery abilities did make use of spatial imagery. As Reisberg wrote: "this, in short, is why studies find no relationship between VVIQ scores and mental rotation" (2013, p. 383). Such conflations between subjective measurement of mental imagery vividness and objective measurement of spatial imagery ability have been written about previously. As Reisberg & Leak wrote, "imagery self-report appears unrelated to performance of many of the standard imagery tasks… the standard imagery tasks require spatial judgements (e.g., about juxtaposition) or spatial manipulations (folding, rotation, etc.). If one argues that spatial representations need be in no way visual, then it is unsurprising that imagery self-report fails to predict performance on these tasks" (1987, p. 521). Conflations such as those between visual imagery vividness and spatial representation abilities may partially explain the discrepancies between subjective and objective measurements of mental imagery ability.

However, it is also important to note that both the ventral and dorsal pathways may be involved to some extent in both spatial and object imagery—"even when the tasks are designed specifically to draw on spatial processing, they require some object processing and vice versa" (Mazard et al., 2004, p. 688). Therefore, in theory, a measure such as the VVIQ still includes some measurement of spatial ability in addition to the more obvious visual imagery because they cannot be entirely separated—and therefore it would be unlikely that the above "wrong measurements" explanation can entirely explain the lack of correlation found between imagery tasks. Other possible forms of the "wrong measurements" explanation concern conflation between *vividness* and *accuracy*, conflation between *visual working memory* tasks and *visual mental imagery* tasks, and conflation between stimulus *recognition* from perception and stimulus *recall* from mental imagery generation.

The first of these, vividness versus accuracy, was described previously in the section about vividness. The important takeaway is that a mental image can be vivid (i.e., strongly resemble true perception) without being accurate. As discussed, several studies have indicated that having vivid mental imagery may make a person more susceptible to misinformation.

Visual working memory and visual mental imagery are similar concepts—both require the active representation of visual information—and some researchers have even written about whether they might be the same thing (e.g., Tong, 2013). The difference between visual working memory and visual mental imagery comes down to whether participants need to *maintain* a mental image or *generate* a mental image. In visual working memory tasks, an image is presented to the participant and they must maintain it and possibly transform it mentally. In a visual mental imagery task, a participant must generate an image from their memory. It is possible to combine these tasks such that a person must first generate a mental image, maintain it, and then transform it. The important difference is that when the image in question is perceptually provided in a working memory task, this bypasses the generation process of mental imagery and so cannot be called a pure mental imagery task. If both tasks are defined as requiring *maintenance* of an image, regardless of whether it was obtained perceptually (as in a working memory task) or from memory (as in a visual mental imagery task), then both tasks certainly overlap in the brain to some extent during this image maintenance, and neural representations examined during that maintenance stage have been shown to be similar using a multivariate analysis of activated voxels during fMRI (e.g., Albers et al., 2013; see also Dijkstra et al., 2019). One study investigated working memory tasks in a single person with developmental aphantasia, and concluded that there was evidence that she was impoverished relative to controls at the "most difficult" of the trials that required a high degree of precision (Jacobs et al., 2018). More research is needed to determine whether people with aphantasia respond differently to visual working memory tasks, and a task that could be called a working memory task will be used in Study 3 in order to investigate this.

Finally, the conflation between recognition tasks and recall tasks was discussed in the section on "Objective Measurement of Visual Imagery" in relation to Reisberg et al. (1986) and Heuer et al. (1986) in their pursuit of mental color studies (see also Study 4's discussion of the recall task in Gur & Hilgard, 1975, versus the recognition task in Berger and Gaunitz, 1977). The important takeaway is that recognition tasks may make partial use of neural pathways from V1 to extrastriate cortex, while recall tasks reliant on mental imagery in which no answers are presented perceptually makes use of the neural pathways in the opposite direction leading to V1—and such neural connectivity is not necessarily equivalent when reversing direction. If people with aphantasia encode visual information differently than controls—for example, using a propositional/categorical approach instead of a coordinate/depictive approach for complex visual information—then task performance on a recall task that requires image generation should be affected relative to recall task that does not necessarily require image generation (e.g., can be completed with propositional knowledge) in a way that can be detected in Study 4.

In the following experiments, I strive to eliminate (or at least identify) places where possible alternative methods of arriving at the correct answer or possible conflation may exist between the cognitive processes used and the processes relied upon by the objective task in question (e.g., spatial versus visual imagery, visual working memory versus visual mental imagery, etc.).

Purpose of the Current Studies

A summary of the theories that are to be tested by this series of studies is provided as Table 1. The series of studies makes use of several different “objective” tasks that encourage the use of mental imagery. They are designed such that evidence for the existence of aphantasia may be gathered, and so that evidence for and against three theories of aphantasia can be collected.

Table 1. *Theory Space for Aphantasia Explanations*

Representation Type	Neurotypical	People with Aphantasia
Explicit Depictive	Requires acceptance of subjective imagery accounts	Mistaken/Malingering theory places aphantasia here
Implicit Depictive	Established as possible by the great imagery debate	Implicit/Explicit theory places aphantasia here
Propositional	Established as possible by the great imagery debate	Perceptual/Conceptual theory places aphantasia here

The overall goals of this series of studies are as follows:

Expand understanding of the experience of aphantasia through survey methods, such as prevalence of dreaming, religiosity, and enjoyment of fictional stories (Study 1).

Ask all participants to report the extent they made use of visual mental imagery during the "objective" tasks in this series of studies (Studies 2, 3, and 4). This may inform the question of whether the "objective" tasks actually require mental imagery, or if they can be circumvented with other methods (e.g., propositional encoding/recall strategies).

Attempt to improve calibration of VVIQ scores with "objective" mental imagery tasks using cognitive measures (e.g., response time, accuracy) by providing video-based information about individual differences in mental imagery vividness that allows participants to more accurately evaluate their mental imagery abilities using the VVIQ (Study 2). I predict that providing imagery information before the VVIQ will result in stronger correlations between VVIQ scores and "objective" cognitive tasks (i.e., the Backwards Spelling Task, the Snowy Pictures Task, and the Tail Length Task) when compared to the correlation for participants who did not receive the video instructions.

Discover how well some previously used "objective" tasks of mental imagery can be accomplished by people with developmental aphantasia (Study 2).

Search for evidence of a difference in the way people with aphantasia solve the classic imagery task of mental scanning (Study 3). If scanning distance does not affect the reaction times of people with aphantasia in the same way that it affects control participants, perceptual/conceptual theory is supported. Conversely, if people with aphantasia are just as affected by mental scanning distances and task complexity as controls, implicit/explicit theory is supported.

Search for evidence for increased propositional (categorical) encoding of stimuli in people with aphantasia, or reduced depictive (coordinate) encoding, or both, supporting perceptual/conceptual theory. Or, show that people with aphantasia are just as capable of encoding and retrieving in a depictive format as control participants, supporting implicit/explicit theory (Study 4).

Compare the results of people with aphantasia who report visual images during nighttime dreaming with the results of people with aphantasia who do not report visual images during nighttime dreaming—people with aphantasia may constitute two separate groups, one that supports the perceptual/conceptual theory (do not dream due to missing backpropagation), and one that supports the implicit/explicit theory (can dream due to reduced processing resources while awake; Studies 3 and 4).

The perceptual/conceptual theory states: people with aphantasia do not use mental imagery (implicitly or explicitly) during cognitive tasks. Therefore, their performance in regard to reaction times and accuracy *should be* demonstrably different from control participants (i.e., people who *do* use mental imagery) due to the use of different strategies (e.g., propositional encoding/recall) on cognitive tasks. However, the reaction times and accuracy of people with aphantasia *should not be* affected by certain within-person manipulated task traits that rely on mental image generation such as scanning distance.

The implicit/explicit theory states: people with aphantasia are using implicit mental imagery, which is not perceivable to them, but does influence performance on tasks that make use of mental imagery. Therefore, their performance in regard to reaction times and accuracy *should not be* demonstrably different from control participants (i.e., people who *also* use mental imagery) due to the use of the same implicit strategies (e.g., depictive encoding/recall) on cognitive tasks. Reaction times and accuracy of people with aphantasia *should be* affected by certain within-participant manipulated task traits that rely on mental imagery generation such as scanning distance.

CHAPTER 2. STUDY ONE

Study 1 consisted of a survey, a video manipulation for the control group, and the Vividness of Visual Imagery Questionnaire (VVIQ). People with aphantasia have not been exhaustively surveyed about their life experiences, which leaves many demographic questions unanswered. Some qualitative research of individuals (Kendle, 2017) and quantitative survey-based research has been conducted (e.g., Dawes et al., 2020; Faw, 2009; Zeman et al., 2020). This survey tapped into additional questions that could lead to useful research avenues in the future. It also was intended to quantify some of the previous qualitative research (Kendle, 2017).

The video manipulation for the control group was designed to explore the possibility that people who are naïve about individual differences in mental imagery (i.e., do not understand that other people may experience mental imagery or a lack of mental imagery that is different from their own experiences) may be poorly calibrated in terms of their VVIQ scores and their performance on tasks that benefit from the use of mental imagery. To rectify this, half of the control group was shown an educational video about individual differences in mental imagery before taking the VVIQ, while the other half watched a “control” video about color perception which had nothing to do with mental imagery.

It is no great secret that comparisons of subjective scales can be problematic (e.g., Heine et al., 2002). In this case, comparing VVIQ scores across participants may be problematic if calibration is poor between VVIQ scores and performance on tasks that benefit from the use of mental imagery. If people vary in “their expectations of what images should be like” (N. H. Kerr & Neisser, 1983), then an informational video should help participants better understand where they fall on the spectrum of mental imagery ability. Because VVIQ scores are positively correlated with social desirability (e.g., McKelvie, 1995; Winograd et al., 1998), it is also important to emphasize that more vivid mental imagery isn't necessarily superior to less vivid imagery, and also include that some people don't experience much if any mental imagery.

I hypothesize that viewing the imagery video will change the way that participants subjectively report their imagery experiences (e.g., respond to the VVIQ) such that control participants who watched the imagery video have significantly different VVIQ scores than control participants who watched the color perception video. Because of the negative skew of VVIQ scores in the general population (Zeman et al., 2020), I predict that watching the imagery video should *lower* the average person's VVIQ score, such that there is less skew in the resulting distribution of scores when participants have more information about individual differences in mental imagery abilities. In other words, I believe that the average person who is naïve to individual differences in mental imagery vividness overestimates their imagery vividness, such as someone who has aphantasia but has not yet learned about aphantasia.

Method

Participants

The participants in the aphantasia group were recruited from online communities such as the aphantasia subreddit on Reddit and through Facebook groups for aphantasia (see Bainbridge et al., 2021, for a similar recruitment strategy). In this way, I recruited a larger sample of people with aphantasia than I would have been able to with mass testing, due to the low prevalence of aphantasia in the general population—perhaps 2–3% (Faw, 2009). Funding to recruit people with aphantasia was provided by Dissertation Completion Funding from the Psychology Department at Iowa State.

In order for people with aphantasia to participate, they needed to email me and express interest in participating. I screened these potential participants by asking them if they were at least eighteen years old, and if they identified as having poor mental imagery such as aphantasia. If the potential participant replied to the screening email with the answer of “yes” to both questions, I emailed them a link to the Qualtrics survey and they were officially enrolled.

The control group consisted of undergraduate students at Iowa State University that were enrolled in an introductory psychology or communication studies course. This control group recruitment was completed via the institution’s research (SONA) pool. As part of the study, the control group participants were asked whether they identified as someone with aphantasia. If they answered yes, they were moved from the control group to the aphantasia group (four participants were moved in this way).

Participants were compensated for their participation in this study. Participants enrolled at Iowa State University received partial course credit. Participants who were not enrolled were compensated with a gift card for $15, but were allowed to decline the gift card if they chose to do so. Permission to carry out all reported studies with these participants was obtained from the Institutional Review Board (see Appendix E).

There were 213 participants in the control group and 107 participants in the aphantasia group before participants were excluded for a variety of reasons. Thirty-nine participants in the control group and eight participants in the aphantasia group began the Qualtrics survey but did not complete the entirety of the study; some of these people dropped out because of technology limitations as evidenced by email follow-ups (e.g., their computer did not allow them to run the entirety of the study due to software limitations or lack of administrative access to install the necessary InquisitPlayer software), while others simply chose to quit because they felt that they did not have the time or mental energy to complete the study at the time that they chose to begin the asynchronous study. Those participants were excluded from all analyses.

One person in the control group failed an embedded attention check in the form of a question asking for a specific response (i.e., please respond “4” to this question), and this person was excluded from all analyses. Two people in the control group failed an attention check following the presentation of an educational video about mental imagery (i.e., “Did you watch the entire video?”) and were excluded from all analyses. Six participants in the control group completed the Qualtrics survey multiple times in a way that exposed them to both randomized video conditions (see Study 2) and were excluded from all analyses. Four people in the control group were excluded from all analyses because they produced unusable data, possibly because they tried to use an unapproved platform to complete the study (mobile phones and tablets were explicitly not allowed because the study protocol used a keyboard, but these directions may not have been followed) or because they did not correctly enter their identifier numbers that would allow the data from the two programs to be matched up. Next, four people in the control group were moved to the aphantasia group because their survey answers indicated that they believed themselves to have aphantasia. Following those exclusions, there were 99 participants remaining in the aphantasia group and 157 participants remaining in the control group, but these were not the final numbers.

Next, it was necessary to screen the data for quality due to the asynchronous and online nature of the study. This was accomplished by looking at a randomly chosen subset of data comprising less than 10% of the entire dataset. That subset was examined for response time patterns at the level of individual trials. From this subset, exclusion criteria were developed that were then applied to the entire dataset. Participants were excluded from all analyses if they showed great evidence of speeding through the study as fast as possible, or evidence of great distraction such that they had extremely long response times. More specifically, participants were removed from the dataset if they met any of the following four criteria.

1) Two or more response times of less than 100 ms. [Mean count of such responses before exclusion (*N* = 256) was 1.195 (*SD* = 3.882). Mean after all exclusions (*N* = 215) was .033 (*SD* = .178)]
2) Ten or more response times of less than 300 ms. [Mean count of such responses before exclusion (*N* = 256) was 4.457 (*SD* = 11.620). Mean after all exclusions (*N* = 215) was .355 (*SD* = .957)]
3) Three or more response times at or exceeding two minutes. [Mean count of such responses before exclusion (*N* = 256) was .082 (*SD* = .361). Mean after all exclusions (*N* = 215) was .042 (*SD* = .223)]
4) One or more response time at or exceeding ten minutes. [Mean count of such responses before exclusion (*N* = 256) was .008 (*SD* = .088). Mean after all exclusions (*N* = 215) was 0 (*SD* = 0)]

Thirty-seven participants in the control group and four participants in the aphantasia group were eliminated due to these data quality criteria. Many participants greatly exceeded the threshold(s) for elimination, producing several unrealistic or even impossible response times (e.g., of zero milliseconds), suggesting that they were rushing through the study and not completing the trials with care. In contrast, very long response times may indicate multitasking during the studies, but because no researcher was in the room with these participants while they were completing the asynchronous studies, I cannot be certain of the cause.

Thus, the final number of participants in this study was 120 participants in the control group and 95 participants in the aphantasia group. This final number exceeded my target of 66 participants in the aphantasia group. That target number was based on a power analyses run using GPower 3.1 (Faul et al., 2009), with the assumption of a small-to-medium effect size (i.e., partial eta squared = .05) resulting from a within-between interaction tested using ANOVA. It is worth noting that the reason I was able to exceed expectations for recruitment of people with aphantasia was that a large number of participants in that group agreed to forgo compensation in the interest of my recruiting additional people for the study, and for that I owe a debt of gratitude.

Materials and Procedure

As noted, Study 1 consisted of a survey, a video manipulation for the control group, the Vividness of Visual Imagery Questionnaire (VVIQ). A novel measure of imagery use during cognitive tasks, the Total Self-Reported Imagery Use (TSIU), will also be discussed as part of Study 1, although it was administered in pieces throughout the studies.

Survey

Qualtrics was used to administer the survey. The procedure was completed remotely and online because aphantasia is a rare enough condition that the participants were geographically dispersed beyond means of bringing them physically to the laboratory, and restrictions on travel and laboratory protocols due to the ongoing COVID-19 pandemic were prohibitive for in-person research.

The survey questions were drawn from a variety of sources. Some were taken directly from recent survey research (Dawes et al., 2020; Zeman et al., 2020). Others were generated from previous qualitative research in order to determine how universal the described experiences are among people with aphantasia (e.g., Kendle, 2017). Additional questions were generated based on speculations garnered from other sources in the literature, or curiosities from people discussing the nature of aphantasia online. The survey questions and rationale were as follows:

1) What is your age in years? (Fill in the blank)

Age is a standard demographic question. Galton thought that younger people had stronger imagery and older people had weaker imagery (Galton, 1883). There is some evidence that younger people use mental imagery more than older people (see Kosslyn, 1983, p. 20), but to my knowledge, this aging hypothesis has not been fully explored.

2) Please specify your gender identity: (Optional, fill in the blank)

3) Please specify your sex assigned at birth: (Male; Female; Intersex; Prefer not to disclose)

Gender and sex are standard demographic questions. Galton thought that women had stronger imagery (Galton, 1883). However, modern surveys have not found a significant difference for aphantasia incidence regarding sex/gender—although both recent large-scale surveys conflated sex and gender (i.e., equating gender with male/female rather than man/woman; Dawes et al., 2020; Zeman et al., 2020). Treating gender and sex as independent demographic questions is warranted, because they are not always equivalent.

4) Which racial and/or ethnic categories describe you? Please check all that apply, and provide details in the blank space: (White; Hispanic, Latino, or Spanish origin; Black; Asian; American Indian or Alaskan Native; Middle Eastern or North African; Pacific Islander; Some other race or origin)

This is a standard race/ethnicity demographic question that to my knowledge has not yet been explored in its interaction with the incidence of aphantasia. In order to improve classification, written elaboration of race/ethnicity was allowed (e.g., participants were allowed to specify further identification details such as Chinese, Japanese, etc.).

5) In general, to what extent do you enjoy reading fictional stories (i.e., about events that did not truly happen such as fantasy or science-fiction books): (Likert scale from 1-do not enjoy at all to 6-enjoy immensely)

6) In general, to what extent do you enjoy reading non-fictional stories (i.e., about events that truly happened such as biographies, true crime, or science books): (Likert scale from 1-do not enjoy at all to 6-enjoy immensely)

7) Given the choice, would you prefer to watch a fictional film (such as a fantasy or science fiction film) or a non-fictional film (such as a documentary): (Fiction; Non-fiction).

In theory, people with aphantasia enjoy fictional stories less than the average person because they are not able to picture faces, events, locations, etc. This question was inspired by M. Kerr & Pear (1931), Kendle (2017, pp. 66–67 & 71–77), and Stokes (2019).

8) To what extent do you consider yourself to be a religious person? Religiousness is defined here as "having belief in an organized divine entity with rituals and practices related to a higher power" (Likert scale from 1-I am not at all religious to 6-I am very religious)

9) To what extent do you consider yourself to be a spiritual person? Spirituality is defined for this question as "having deep feelings about belonging to a cosmic whole greater than myself, such as finding meaning in nature or the universe" (Likert scale from 1-I am not at all spiritual to 6-I am very spiritual)

It is possible that people with aphantasia are less religious/spiritual than the average person. Vivid mental images might encourage a person to identify as more spiritual because of their ability to call vivid meaningful images such as religious iconography into their mind, in a similar fashion as to how hallucinations experienced while using psychedelics are associated with increased spirituality (e.g., Móró et al., 2011). Similarly, visual hallucinations are often reported to be an important part of religious experiences in epilepsy patients (e.g., Devinksy & Lai, 2008). I do not believe that possible connections between aphantasia and religion or aphantasia and spirituality have been previously examined.

10) Upon waking up from sleep, how frequently do you recall a dream that you had or were having while asleep? (Likert scale from 1-never to 6-always)

11) During your dreams whilst asleep, to what extent do you experience mental imagery such as being able to see pictures, faces, and colors? (Likert scale from 1-I never experience dreams with mental imagery to 6-I always experience mental imagery in dreams)

Questions about dreams have been asked in a previous survey (Dawes et al., 2020). This question is important because it could potentially help differentiate between backpropagation aphantasia and attentional resource aphantasia. As Supplemental Analyses for Studies 3 and 4, people with aphantasia were split into two groups based upon whether they *never* experience visual dreams, or if they *sometimes/regularly* experience dreams for an additional set of analyses designed to see if there are two clear sub-types of aphantasia reflected in the cognitive results.

12) Can you distinguish between all of the colors currently displayed on your screen along with this question? (Yes, they all look like different colors to me; No, I cannot distinguish between two or more of the displayed colors).

12a) If you cannot distinguish between two or more of the colors, please specify which colors are giving you trouble. This will not affect your ability to participate in the remainder of the study. (Fill in the blank)

This question is important because color is used in Study 4. Question 12a was displayed only if participants indicated “No…” on question 12.

13) Do you identify as having aphantasia (i.e., someone who lacks the ability to visualize pictures in their imagination): (Yes; No)

Questions 14 and 15 were displayed only for participants who reported “yes” to this question about having aphantasia. Question 13 was also used to move control participants into the aphantasia group if they indicated “yes” on this question.

14) How old were you, in years, when you first discovered that you have aphantasia? (Fill in the blank)

Qualitative data show that, on average, people do not discover their aphantasia until they are adults (e.g., Kendle, 2017; Zeman et al., 2020).

15) Have you always had aphantasia for your entire life? Did you lose the ability to visualize mental images, or did you never have it at all? (I used to be able to form mental images; I have never been able to form mental images)

This question is important because it will allow me to separate out any people with acquired aphantasia who may have accidentally participated in the research. Only people with developmental aphantasia are included in the analyses.

Video Assignment

Following their responses to the Qualtrics survey, control participants were randomly assigned either to watch an informative video about individual differences in mental imagery, or to watch an informative video about individual differences in color perception (e.g., color blindness). The video was embedded in the end of the survey, which is why it is mentioned here. Participants in the aphantasia group did not watch a video, to avoid splitting the sample. This video manipulation is relevant for the discussion of all studies (see Study 2 Materials).

Vividness of Visual Imagery Questionnaire (VVIQ) Scores

After this, the Qualtrics page redirected all participants to the next part of the study which ran on a piece of software called InquisitPlayer. Participants were required to install the InquisitPlayer on their computer in order to continue the study. Participants were made aware of this necessary download before they began the Qualtrics survey.

Next, as part of the InquisitPlayer part of the study, all participants self-administered the Vividness of Visual Imagery Questionnaire (VVIQ; see Appendix A) by reading and then answering the questions about their mental imagery abilities. Some previous studies administered the VVIQ following the experimental phase in an attempt to disguise the nature of the experiment as having to do with mental imagery differences. However, McKelvie cautioned against administering the VVIQ following experimental tasks because “if subjects think that they have done well or if they are self-confident, they respond more leniently and report more vivid imagery” (1995, p. 27). Additionally, because some participants will be recruited due to their aphantasia, the purpose of the study cannot be disguised with a cover story in a way that it would have nothing to do with mental imagery, which would defeat the purpose of administering the

VVIQ following the experimental task. Therefore, the VVIQ was administered before the cognitive tasks phase of the study.

Self-Reported Imagery Use for Cognitive Tasks

Researchers have used the majority of the cognitive tasks seen in the present studies in order to assess mental imagery ability abilities—though not always related to aphantasia specifically—but there remains doubt that these cognitive measures actually make use of visual mental imagery. To investigate this, I asked participants to self-report the extent to which they felt that they had made use of visual mental imagery during each task following the completion of that task (e.g., “While completing the change identification task, to what extent did you use visual mental imagery, such as imagining pictures, in order to complete the task?”). They replied by providing a score from one to four, corresponding to responses of “I did not use mental images,” “I barely used them,” “I used them quite a bit,” and “I used them a whole lot.” This was done for all five tasks: the backwards spelling task (Study 2), the snowy pictures task (Study 2), the tail length task (Study 2), the square donut scanning task (Study 3), and the change identification task (Study 4). When all five “did you use visual mental imagery” questions are summed as “total self-reported imagery use” scores (TSIU), they are of theoretical interest when viewed alongside the initial discussion of VVIQ score differences, and so are discussed as part of Study 1, while the scores on the individual five questions (i.e., by study) about “self-reported imagery use” (SIU) are themselves discussed with the appropriate studies.

Study 1 Results and Discussion

All datasets used for this study have been anonymized and can be accessed through the OSF project page, and are usable with attribution (i.e., cite this project: Toftness, 2022; https://osf.io/u3bxj/).

Demographics

Participants in the aphantasia group were older on average (*M* = 33.46, *SD* = 11.94) than participants in the control group (*M* = 19.16, *SD* = 1.86). This was a significant difference after correcting degrees of freedom for a significant Levene's test ($F = 140.99$, $p < .001$), and thus, age is controlled for in the analyses presented later in this series of studies ($t(97.62) = 11.57$, $p < .001$, $d = 2.34$). This age difference makes intuitive sense because the undergraduate control population is more generally constrained in its age than is the population of Reddit and Facebook users from which the aphantasia group was mostly recruited (excepting the four aphantasia group members sourced via the control group recruitment method). In order to control for age in key analyses, a propensity score approach was attempted, but was not advisable given the poor availability of matched ages between group and the small control group sample size following the split due to the video manipulation (See Supplemental Analyses for details). Thus, age was entered as a covariate in the analyses of interest (ANCOVA). While this covariate approach is imperfect, and generally inadvisable for quasi-experiments, in this case, the aphantasia participants reported *always* having experienced aphantasia (see Aphantasia Characteristics, below), and therefore there is no reason to believe that aphantasia as a condition is dependent on age, and thus age and aphantasia status are not expected to share variance on a theoretical basis (see Miller & Chapman, 2001). That is, among the population from which the aphantasia group was sampled (i.e., all people with aphantasia), there is no reason to believe that age has a relationship with developmental aphantasia due to the very nature of developmental aphantasia being defined as a life-long condition. Thus, entering age as a covariate should, in theory, remove only noise from the analysis, and no substantive variance for the groups (aphantasia versus control). If the aphantasia participants are mistaken about the severity being lifelong, and age and aphantasia status do share variance (e.g., if aphantasia becomes worse over time from young adulthood to older adulthood)—an effect that has not yet been established in the literature but someday *might* be—then it is conceivable that controlling for age would inflate the Type II error rate for some reported analyses due to the act of controlling for age taking a bite out of the variance due to aphantasia status. Fortunately, effects were detected even after controlling for age (and sex, as discussed below), making the possible inflated Type II error rate a moot point for the final conclusions.

Sex assigned at birth was significantly different between groups, with a nearly equal number of females (n = 47) and males (n = 48) in the aphantasia group, but a disproportionately higher number of females (n = 82) than males (n = 39) in the control group ($\chi^2(1, N = 215) = 7.86, p = .005$). This result makes intuitive sense because the undergraduate population within the psychology major tends to skew female. Thus, sex was also entered as a covariate in the analyses of interest (ANCOVA), with the same logic as age.

Previous aphantasia surveys have not distinguished between sex assigned at birth and gender (Dawes et al., 2020; Zeman et al., 2020). In this study, the majority of the 214 participants who provided both pieces of information (one participant in the control group was excluded because they wrote "straight" for the free response question about gender) were cisgender (94.86% overall) such that their sex assigned at birth matched the Western gender identity that they most identified with (i.e., biological females identifying as women and biological males identifying as men). Participants in the aphantasia group were more likely to identify with a gender identity other than cisgender (e.g., non-binary, genderqueer, no gender, gender fluid) with 9.47% compared to 1.68% in the control group ($\chi^2(1, N = 214) = 6.58, p =$

.010). However, it is important to note that this chi-square analysis does not control for age or sex assigned at birth.

Overall, there were no significant differences detected between the groups for race and ethnicity. Participants could identify with one or more races or ethnicities. One person in the aphantasia group did not check any races or ethnicities and instead typed in “mixed” as a response, and was coded as “some other race or origin.” In the aphantasia group and control group respectively, a similar number of people identified as white (85% vs. 85%), Hispanic, Latino or Spanish in origin (7% vs. 6%), black (2% vs. 3%), Asian (7% vs. 8%), American Indian or Alaskan Native (0% vs. 2%), Middle Eastern or North African (1% vs. 2%), Pacific Islander (0% in both groups), and some other race or origin (2% in both groups). The percentages add up to more than 100% because categories include people who checked more than one category. No significant difference was detected between groups in terms of how many people identified as multiracial (i.e., picked more than one racial or ethnic category), with six people in the aphantasia group (6.32%) and seven people in the control group (5.83%) identifying as such ($\chi^2(1, N = 215) = .022, p = .883$).

One person was red/green colorblind, and their data was not used for the analyses in Study 4 because those analyses depend upon color perception.

VVIQ Scores and Self-Reported Imagery Use for Cognitive Tasks

VVIQ scores covered the full range of possible scores from 16 to 80. For the 95 participants in the aphantasia group, the median VVIQ score was 16—the lowest possible score, as expected from people identifying as having aphantasia—while the mean was 19.00 (*SD* = 5.88). The VVIQ scores within the aphantasia group ranged from 16 to 49, although the vast majority of scores (96.8%) fell at or below a score of 32. For the control group, the median score was 60 while the average was 58.83 (*SD* = 11.79). The scores for the control group ranged from 16 to 80, although the vast majority of scores (95%) fell at or above a score of 42. Thus, the VVIQ score distributions looked very different between the two groups, as expected.

Participants in the control group who watched the video that explained individual differences in mental imagery ability (the "imagery" video) reported significantly lower VVIQ scores ($n = 64$, $M = 56.45$, $SD = 12.99$, median = 57.5, range = 16–80) than control participants who watched an irrelevant video of identical length about individual differences in color perception ability ($n = 56$, $M = 61.55$, $SD = 9.66$, median = 62, range = 39–80), suggesting that people who are more informed about mental imagery are likely to deflate their self-reported estimate of their mental imagery vividness ($t(118) = 2.41$, $p = .017$, $d = .446$). Range restriction is noted here, with the color perception video producing a distribution restricted to scores from 39–80.

Total self-reported imagery use (TSIU) scores were calculated by adding together the five self-reported imagery use responses from each cognitive task for each participant (i.e., from the BST, SPT, TLT, square donut scanning task, and change identification task), and the resulting TSIU scores covered the full range of possible scores from 5 to 20. For the 95 participants in the aphantasia group, the median score was 5—the lowest possible score, as expected from people identifying as having aphantasia—while the mean was 5.73 (*SD* = 1.51). TSIU scores within the aphantasia group ranged from 5 to 14, although the vast majority of scores (96.8%) fell at or below a score of 9. For the control group, the median score was 14 while the mean was 14.04 (*SD* = 2.90). The scores for the control group ranged from 6 to 20, although the vast majority of scores (96.7%) fell at or above a score of 9. Thus, TSIU score distributions looked very different between the two groups, as expected.

Participants in the control group who watched the imagery video reported significantly lower TSIU scores (n = 64, M = 13.45, SD = 2.89, median = 14, range = 6–20) than control participants who watched the video about individual differences in color perception abilities (n = 56, M = 14.71, SD = 2.78, median = 15, range = 7–20), suggesting that people who are more informed about visual mental imagery are more likely to deflate their self-reported estimate of their use of visual mental imagery during cognitive tasks ($t(118) = 2.43$, $p = .017$, $d = .447$).

VVIQ scores and TSIU scores were consistently positively correlated such that participants reporting stronger mental imagery abilities tended to also report making more use of visual mental imagery during the cognitive tasks. This was true when looking at participants overall ($r(213) = .862$, $p < .001$; partial correlation while controlling for age and sex: partial $r(211) = .790$, $p < .001$), when looking only at the aphantasia group ($r(93) = .780$, $p < .001$), when looking at the entire control group ($r(118) = .294$, $p = .001$), when looking at only participants in the control group who watched the imagery video ($r(62) = .247$, $p = .050$), and when looking at only participants in the control group who watched the color perception video ($r(54) = .285$, $p = .033$).

The difference between the correlation of VVIQ and TSIU within the aphantasia group and the same correlation within the control group is significant (Fisher's $z = 5.33$, $p < .001$; for calculation method, see Diedenhofen & Musch, 2015). The fact that the correlation with TSIU is stronger for lower VVIQ scores (participants with aphantasia) than it is for higher VVIQ scores (control participants) makes intuitive sense: people with lower VVIQ scores have less ability to use imagery during tasks, and therefore are likely to consistently report not using imagery, and that consistency lends itself to a strong correlation. However, people with higher VVIQ scores have the *option* to either use imagery or to not use imagery, and so their reports of using or not using imagery are not consistent and therefore the correlation is weaker.

Entertainment Preferences

Participants were asked to report the extent to which they enjoyed reading fictional stories on a Likert scale from 1 to 6. When the aphantasia group (M = 4.61, SD = 1.68) and the control group (M = 4.73, SD = 1.16) were compared using an independent samples t-test, after correcting the degrees of freedom for a significant Levene's test (F = 20.78, p < .001), no significant difference was detected (t(160.39) = .567, p = .572, d = .090). The Levene's test appeared to be highly significant here because very few participants chose low scores (i.e., expressing that they do not enjoy fiction), with participants in both groups clustering near the top of the available range of scores (i.e., expressing that they do enjoy fiction), thus creating curves for each group that were visibly not normally distributed. Considering the shape of these data, I decided to bin them and run a categorical analysis. I treated scores of 1–3 as "low" and scores of 4–6 as "high" and ran a chi-square analysis to compare between the control group and the aphantasia group. This analysis was significant, with a smaller proportion of people in the aphantasia group (74.74%) than the control group (87.50%) reporting high enjoyment of fiction (χ^2(1, N = 215) = 5.82, p = .016).

Participants were also asked to report the extent to which they enjoyed reading non-fictional stories on a Likert scale from 1 to 6. When the aphantasia group (M = 3.92, SD = 1.40) and the control group (M = 3.83, SD = 1.18) were compared using an independent samples t-test, after correcting the degrees of freedom for a significant Levene's test (F = 5.583, p = .019), no significant difference was detected (t(183.12) = .505, p = .614, d = .075). Following the line of thought from the "fiction" data, and noting the shape of these data as well, I binned them to run a categorical analysis. As before, I treated scores of 1–3 as "low" and scores of 4–6 as "high" and ran a chi-square analysis to compare between the control group and the aphantasia group. This analysis was not significant, with a smaller but non-significant proportion of people in the aphantasia group (56.8%) than the control group (60.8%) having a high reported enjoyment of non-fiction ($\chi^2(1, N = 215) = .349, p = .554$).

There was not a significant correlation between VVIQ scores and preferences for fiction ($r(213) = .057, p = .406$), and there was not a significant correlation between VVIQ scores and preferences for non-fiction ($r(213) = .012, p = .861$). When this analysis was repeated excluding all control participants who did not watch the imagery video, a similar pair of weak and non-significant correlations emerged ($r(157) = .011, p = .899$; $r(157) = -.024, p = .767$). There was not a significant correlation between preferences for fiction and preferences for non-fiction ($r(213) = .007, p = .923$).

When asked whether they would rather watch a fictional film or a non-fictional film, the group's proportional responses were not significantly different. In the aphantasia group, 15.8% chose the non-fictional option, while in the control group, 20.8% chose that option ($\chi^2(1, N = 215) = .891, p = .345$).

Beliefs

Participants in the aphantasia group reported being far less religious than participants in the control group, but the groups did not statistically differ in their self-reported spirituality. Religiosity in the aphantasia group ($M = 1.85, SD = 1.51$) was significantly lower than religiosity in the control group ($M = 3.27, SD = 1.73$) after correcting the degrees of freedom for a significant Levene's test ($F = 6.73, p = .010$) when using an independent samples t-test ($t(210.94) = 6.40, p < .001, d = .881$). However, spirituality in the aphantasia group ($M = 3.02$, *SD* = 1.71) was not significantly different from spirituality in the control group (*M* = 3.24, *SD* = 1.48) after correcting the degrees of freedom for a significant Levene's test (*F* = 6.81, *p* = .010) when using an independent samples t-test (*t*(186.75) = .994, *p* = .321, *d* = .145).

There was a significant positive relationship between VVIQ scores and religiosity (*r*(213) = .344, $p < .001$). There was no significant relationship between VVIQ scores and spirituality (*r*(213) = .096, *p* = .159). This same pattern of correlations held when control participants who did not watch the imagery video were excluded from these two correlations (*r*(157) = .329, $p < .001$; *r*(157) = .066, *p* = .410). There was a significant positive relationship between religiosity and spirituality (*r*(213) = .452, $p < .001$).

After using an ANCOVA to control for both age and sex assigned at birth, the difference between groups in terms of religiosity remained significantly different, with the estimated marginal mean for the aphantasia group (*M* = 1.81, *SE* = .203) lower than for the control group (*M* = 3.30, *SE* = .174; *F*(1, 211) = 65.11, $p < .001$, $\eta_p^2 = .103$, observed power = .998). Meanwhile, there was no significant effects detected for age or sex (respectively, *F* (1, 211) = .393, *p* = .531, $\eta_p^2 = .002$, observed power = .096; *F* (1, 211) = 1.13, *p* = .289, $\eta_p^2 = .005$, observed power = .185). The religiousness difference appears to be partly driven by the majority of the people in the aphantasia group (68.4%, or 65 people) choosing the lowest response option (e.g., 1 = "I am not at all religious") compared to far fewer people in the control group (21.7%, or 26 people) choosing that option (see Figure 7).

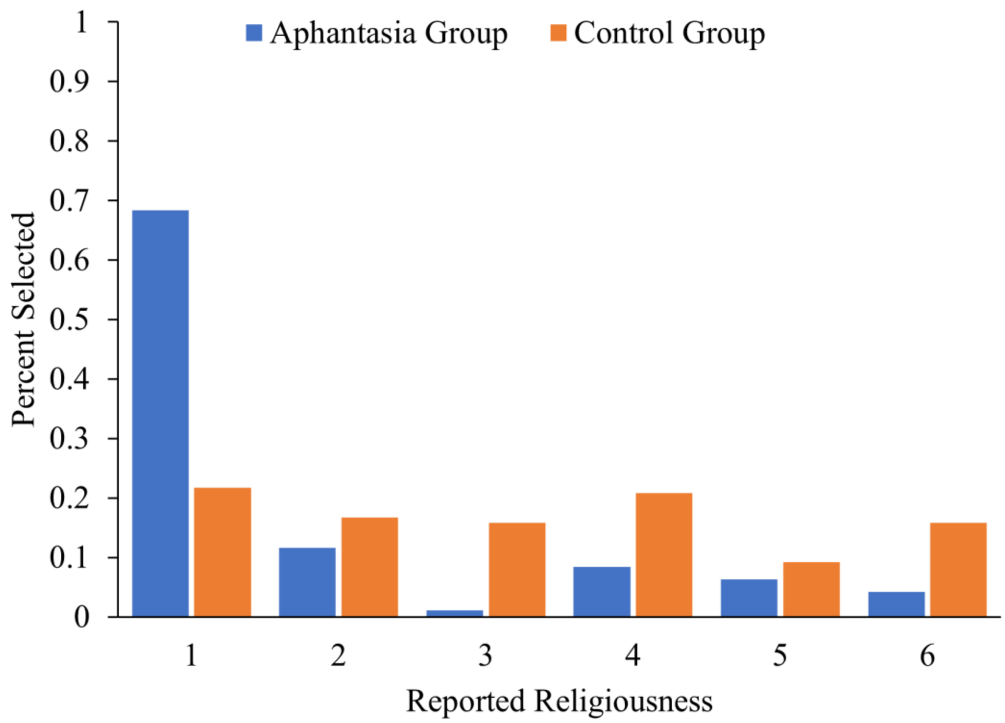


Figure 7. Proportion of participants in each group that selected each of the six possible responses to the question about religiousness: "To what extent do you consider yourself to be a religious person? Religiousness is defined here as "having belief in an organized divine entity with rituals and practices related to a higher power." Likert scale ranged from 1- *I am not at all religious* to 6- *I am very religious*.

Dream Content

Participants were asked to report the frequency with which they remember dreams upon waking from a period of sleep on a Likert scale from 1 to 6. When the aphantasia group (M = 2.89, SD = 1.09) and the control group (M = 3.40, SD = 1.03) were compared using an independent samples t-test, a significant difference was detected for reported dream frequency ($t(196.92) = 3.46$, $p = .001$, $d = .493$).

Participants were also asked to report the frequency for which their dreams contained visual imagery content using a similar Likert scale from 1 to 6. When the aphantasia group (M = 3.63, SD = 1.81) and the control group (M = 4.63, SD = 1.14) were compared using an independent samples t-test, after correcting the degrees of freedom for a significant Levene's test

(F = 48.70, p < .001), the controls reported significantly more frequent imagery within dreams (t(150.30) = 4.71, p < .001, d = .768) as shown in Figure 8.

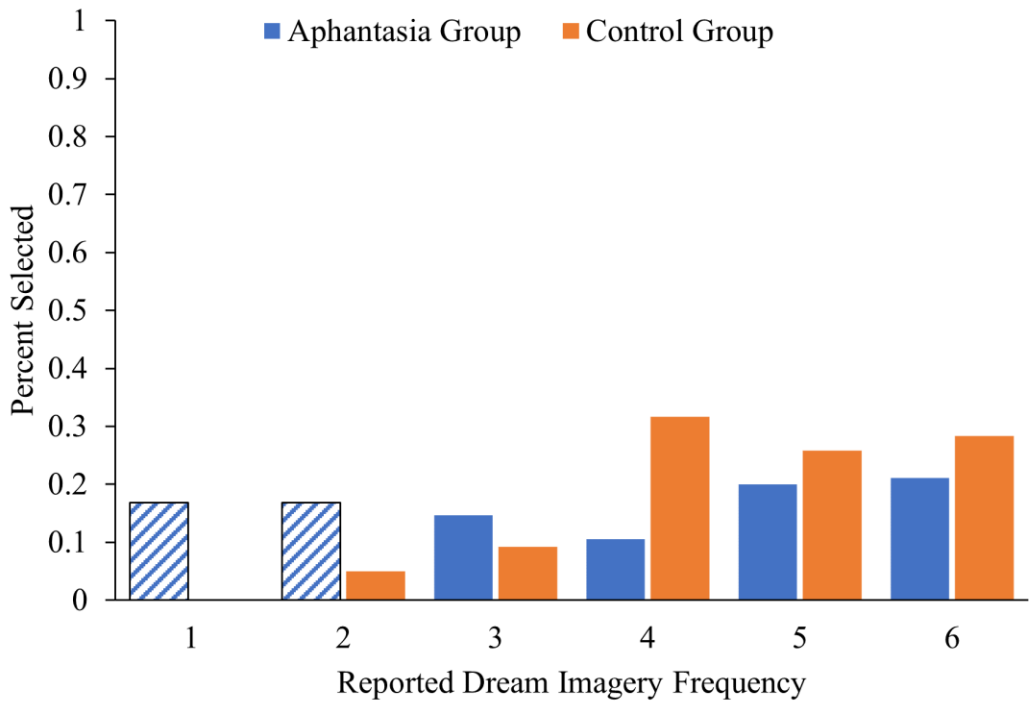


Figure 8. Proportion of participants in each group that selected each of the six possible responses to the question about dream imagery: "During your dreams whilst asleep, to what extent do you experience mental imagery such as being able to see pictures, faces, and colors?" Likert scale from 1- *I never experience dreams with mental imagery* to 6- *I always experience mental imagery in dreams*. The participants with aphantasia who responded "1" or "2," marked with diagonal shading, were treated as a subgroup to be compared with the other participants with aphantasia in further analyses (see Supplemental Analyses).

There was a significant positive relationship between VVIQ scores and reported frequency of dreams (r(213) = .280, p < .001). There was a significant positive relationship between VVIQ scores and reported dream imagery content (r(213) = .350, p < .001). This same pattern of correlations held when control participants who did not watch the imagery video were excluded from these two correlations (r(157) = .220, p = .005; r(157) = .283, p < .001). There was also a significant positive relationship between the reported frequency of dreams and reported dream imagery content (r(213) = .397 , p < .001).

After controlling for age and sex using an ANCOVA, reported dream frequency between groups dropped out of significance, with the estimated marginal means in the aphantasia group (*M* = 2.99, *SE* = .130) not differing significantly from the control group (*M* = 3.33, *SE* = .112; $F(1, 211) = 3.42$, $p = .080$, $\eta_p^2 = .014$, observed power = .416). Neither age nor sex were significant in this ANCOVA (respectively, F (1, 211) = .846, $p = .359$, $\eta_p^2 = .004$, observed power = .150; F (1, 211) = 2.59, $p = .109$, $\eta_p^2 = .012$, observed power = .360).

However, after controlling for age and sex, reported imagery within dreams remained highly significant, such that people with aphantasia reported fewer dreams with imagery content (*M* = 3.55, *SE* = .181) than did the control group (*M* = 4.70, *SE* = .156; $F(1, 211) = 18.17$, $p < .001$, $\eta_p^2 = .079$, observed power = .989). Neither age nor sex were significant in this ANCOVA (respectively, F (1, 211) = 1.57, $p = .211$, $\eta_p^2 = .007$, observed power = .239; F (1, 211) = 3.31, $p = .070$, $\eta_p^2 = .015$, observed power = .440).

Reported dream imagery warrants a further look, because according to the theories of how aphantasia happens in the brain, discussed in the "Theories of Individual Differences in Mental Imagery" section of the Introduction, people with aphantasia who report experiencing *less* imagery during dreams may be experiencing a more "complete" aphantasia than those who at least sometimes experience dreams that feature imagery. Therefore, I coded participants in the aphantasia group with scores of 1 or 2 on the imagery frequency in dreams question into a "few imagery dreams" group (i.e., responses of "never" or "very infrequently"), and coded participants with scores of 3 or above (i.e., responses of "somewhat infrequently," "somewhat frequently," "very frequently," and "always") as "some imagery dreams." This decision to split the group as such (i.e., split between responses of 2 and 3) was determined prior to data collection based on the theory that people who do not experience mental imagery in dreams may be *incapable* of generating mental images while people with aphantasia who do experience mental images during dreams may be capable of generating mental images but unaware of their presence during the waking day, and thus these two subgroups may be heterogeneous in terms of how they will be affected by the difficulty manipulations in Studies 3 and 4 (i.e., the "few imagery dreams" group may have an especially difficult time with those cognitive tasks).

Using these criteria, 32 participants from the aphantasia group were coded as "few dreams" and the remaining 63 were coded as "some dreams." These subgroups did not significantly differ in their age, enjoyment of fiction, religiousness, or VVIQ scores (all $ps > .112$). There was a significant difference for sex assigned at birth, with more men than women in the "few imagery dreams" condition (respectively, $n = 22$; $n = 10$) and fewer men than women in the "some imagery dreams" condition (respectively, $n = 26$; $n = 37$) with a significant chi-square analysis ($\chi^2(1, N = 95) = 6.41$, $p = .011$). These subgroups divided by imagery dream frequency were further investigated in Studies 3 and 4.

Aphantasia Characteristics

Within the aphantasia group, participants were asked about the age at which they discovered that they were experiencing aphantasia. A wide range of ages were reported, ranging from 5 to 65, with a mean of 30.31 and a standard deviation of 12.89. The median score was 28. Notably, very few participants reported discovering their aphantasia before they were adults, with only 12.8% reporting that they discovered the condition before the age of 18.

94.7% of participants in the aphantasia group ($n = 90$) reported that they had experienced aphantasia for as long as they could remember. The remaining five participants in the aphantasia group were asked to describe whether they lost mental imagery ability suddenly (e.g., due to a stroke or injury). None of the participants who had marked themselves as having "acquired" aphantasia had a definitive event that demarcated a clear line between being able to visualize and then becoming unable to visualize. Instead, they reported that they were not sure how they developed aphantasia. Therefore, these participants better fit the definition of developmental aphantasia and were included in all reported analyses rather than being excluded.

Study 1 Summary

Overall, the results of Study 1 provided some interesting insights. First of all, the fact that the means of VVIQ scores in the control group could be significantly altered by video condition suggests that people who are naïve to the topic of mental imagery before taking the VVIQ may not be well calibrated on their understanding of what their mental imagery abilities are like compared to other people. Similarly, participants' estimates of their visual mental imagery use during the cognitive tasks in these studies (TSIU scores) were also significantly affected by random assignment to video condition such that watching a video about individual differences in visual mental imagery vividness reduced estimates of imagery use. These findings suggest that the imagery video helped the control participants to calibrate their estimate of their mental imagery abilities and use relative to others, and overall that control participants were naïve about what imagery is like for other people. However, it is possible that watching the imagery video caused the VVIQ scores and self-reported imagery use scores to become *less* well-calibrated, such that the reduced scores reflect their true mental imagery experiences more poorly.

In order to make the argument that the VVIQ/SIU scores in the imagery video condition were *better* calibrated to true mental imagery experiences than the VVIQ/SIU scores in the color perception video condition, I must show that the VVIQ/SIU scores for the imagery video group have a stronger relationship with performance on cognitive tasks that benefit from the use of imagery than do the color video group. Studies 2, 3 and 4 will attempt to make just such a case.

Additionally, possible group differences were detected. People with aphantasia appear to be less likely to identify as cisgender, less likely to experience enjoyment when reading fictional stories, less likely to be religious, seem to have a decreased frequency of dreaming, and a reduced amount of imagery content within their dreams.

As far as I am aware, this is the first time that identifying as cisgender, enjoyment of reading fictional stories, and religiousness have been shown to have a relationship with aphantasia. Future studies into the demographics of aphantasia should certainly pursue these findings as potential additional individual differences, and perhaps administer more formal measures (e.g., of reading enjoyment, of religiousness, etc.).

The gender identity difference was not predicted and should be replicated before interpretation due to the fact that the two groups were recruited using different strategies that produced such significant intergroup differences (e.g., pre-existing age and sex differences). It might be interesting to intentionally look at this difference in a future set of studies.

People who identify as having aphantasia appear to be less likely to report having a high enjoyment for reading fictional stories. This makes intuitive sense because part of the enjoyment from reading a fictional story may come from being able to visualize scenes and action, and people with aphantasia may be less likely to be able to do so. However, as discussed, this result will need to be replicated before it is interpreted too closely, due to the exploratory nature of the analyses that produced the finding.

Regarding frequencies of dreams and imagery content of dreams, similar results in which people with aphantasia report fewer dreams and less imagery content within dreams were previously reported in a survey (Dawes et al., 2020). Individual differences in imagery content of dreams within the aphantasia group are used in the Supplemental Analyses.

CHAPTER 3. STUDY TWO

Previous studies claim that some cognitive tasks correlate with mental imagery ability as measured by the VVIQ. However, such cognitive tasks are administered inconsistently, and the VVIQ is also administered inconsistently. Many questions remain about whether these cognitive tasks are truly correlated with VVIQ scores, and if they make use of some underlying “visual mental imagery ability” construct. Therefore, one goal of Study 2 is to see whether these cognitive tasks correlate with VVIQ scores, and whether a latent variable can be extracted that also correlates with VVIQ scores. Specifically, I am investigating three cognitive tasks that the literature claims are related to visual mental imagery abilities, and for which at least some evidence exists (see “Objective Measurement of Imagery Ability”): the Backwards Spelling Task (BST; Fernald, 1912; Walczyk & Taylor, 2000), the Snowy Pictures Task (SPT; Ekstrom et al., 1976; Kozhevnikov et al., 2005), and the Tail Length Task (TLT; Policardi et al., 1996).

I hypothesized that presenting video information about imagery abilities to controls will strengthen the correlations between VVIQ scores and performance measures on cognitive tasks, relative to participants who watched the “control” video (see Study 1). Most notably, watching the imagery video should give participants a relative inter-individual spectrum on which to base their own introspective report of imagery vividness, rather than expecting them to rank their imagery vividness based solely on their own internal experience. This familiarity should produce more accurate self-reports of imagery vividness, and therefore a stronger correlation with cognitive tasks that (supposedly) tap imagery. If the imagery video improves the correlations between the “subjective” VVIQ and “objective” cognitive tasks, that may partially explain why the literature is full of poor correlations between the VVIQ and some cognitive tasks: because the participants in those studies had trouble accurately rating their imagery using the VVIQ.

Method

Participants

The participants with aphantasia and the control group were carried over from Study 1. One person in the aphantasia group answered zero questions correctly on the BST, and informed me that the sound did not play for them during the study. Thus, they are excluded from all Study 2 analyses. As a result, there were 94 people in the aphantasia group and 120 participants in the control group for Study 2.

Materials

The instructional videos were prepared by the experimenter. The length of the videos were each 3 min and 48 sec, well within the ideal length for engaging instructional videos which is believed to be less than six minutes (Guo et al., 2014). These videos are unlisted on YouTube, but are freely accessible by link for any interested researcher (Toftness 2021a; Toftness 2021b).

The cognitive tasks were adapted from existing research and have previously shown significant relationships with VVIQ scores. All of these tasks are examined in more detail in the Introduction section called “Objective Measurement of Imagery Ability.” The first task is the backwards spelling task (BST; Fernald, 1912; Walczyk & Taylor, 2000). Words used in my version of the BST were partially adopted from both Fernald (1912) and Walczyk & Taylor (2000) and are presented in Appendix B. The second task is the snowy pictures task (SPT; Ekstrom et al., 1976). A version of the task was created rather than using the original version (Ekstrom et al., 1976) or the modified “difficult” version (Kozhevnikov et al., 2005), because neither existing stimulus set was available to me. Items from my version of the SPT are presented in Appendix C. The third objective task is the tail-length task (TLT; Policardi et al., 1996). The items on the TLT are from Policardi et al. (1996), and are presented in Appendix D.

Procedure

Following the procedure described in Study 1, participants next completed five cognitive tasks of mental imagery. This series of tasks consisted of the BST, TLT, SPT, square donut scanning task (Study 3), and change identification task (Study 4). The order of the tasks was counterbalanced using a Latin square, and the order of the trials within all tasks was randomized. Each task included at least one practice trial to acquaint participants with expected responses, and participants were required to provide the correct answer(s) to the practice trial(s) in order to continue to the criterion part of each task. Both accuracy and response time were measured, and participants were always prompted to respond “as accurately and quickly” as they were able as part of the written instructions proceeding each task.

Following each of the five cognitive tasks, participants self-reported the extent to which they believed they had utilized visual mental imagery during the task that had just been completed. They reported this on a four-point Likert scale, with the available enumerated responses reading *1- I did not use mental images*, *2- I barely used them*, *3- I used them quite a bit*, and *4- I used them a whole lot*. These are referred to as the self-reported imagery use (SIU) measures, which also make up the TSIU scores when summed as discussed with Study 1.

Further details about the BST, SPT, and TLT are below, whereas the square donut scanning task (Study 3) and change identification task (Study 4) are detailed later on in their respective sections.

Backwards Spelling Task Procedure and Scoring

For the BST, words were presented aurally. Prior to the study, participants were told that they needed to use sound when completing the study. After listening to word, the participant was asked to type out the word backwards using a keyboard before hitting the enter key to submit the word. Each word was presented aurally twice, read by the experimenter via prerecorded audio files. Participants were allowed to begin typing as soon as each sound file began to play. One practice trial, using the word “sugar,” was completed before the criterion task, which included twenty words. Scoring for the BST was completed by comparing the string of characters entered to the correct string. Only 100% correct responses were scored as correct, and no deviations from the correct string were allowed.

Snowy Pictures Task Procedure and Scoring

For the SPT, a degraded and partially masked line drawing of an object was displayed on the screen, and the participant was prompted to type in the name of the object and then to press enter. One practice trial, using an image of an apple, was given prior to the criterion task, which included sixteen trials. In order to score the SPT, responses were coded according to their proximity to the correct object response, disregarding spelling. For example, spellings of “sissors” and “sisscors” were accepted as the correct answer for the “scissors” trial. If the spelling did not point towards a discernible object (e.g., a response of “siss”), then that trial was coded as incorrect.

Tail Length Task Procedure and Scoring

For the TLT, participants were told: “in the following task, you will be asked to determine whether a series of animals have long tails or short tails relative to the length of their bodies.” The name of an animal was presented on the screen, and the participant was asked to press one of two keys on the keyboard in order to indicate whether that animal has a long tail (by pressing the “L” key) or a short tail (by pressing the “S” key) relative to its body size. These response key options were displayed on the screen throughout all trials. Participants completed four practice trials (“panda,” “polarbear,” “peacock,” and “eel”) before attempting the criterion task, which consisted of twenty trials. Trials were spaced by a 500-second inter-stimulus interval. Responses were coded as correct or incorrect based on key pressed. Thus, chance performance was 50% for this task.

Results

All datasets used for this study have been anonymized and can be accessed through the OSF project page, and are usable with attribution (i.e., cite this project: Toftness, 2022; https://osf.io/u3bxj/).

Overall Performance on Study 2 Tasks

Considering raw scores, participants in the aphantasia group were significantly more accurate on the BST, SPT, and TLT than were control participants. The aphantasia group also had significantly longer response times on the TLT. However, when ANCOVA was used to control for age and sex, and the two dependent measures (accuracy and response times) were transformed to one dependent measure as rate correct scores (RCS; Vandierendonck, 2017), there were no significant differences in performance on any of the three tasks.

It is widely accepted in cognitive research that both accuracy and response times contain important information regarding the performance of a participant. However, because participants can intentionally trade response time for accuracy in cognitive tasks such as those used in this series of studies (e.g., respond slower in order to raise accuracy), it is important to consider these dependent measures both separately and together. For example, significant effects may point in opposite directions such that Group A has higher accuracy than Group B, but Group A has longer response times than Group B. In such a case, which group had "better" performance? To answer this question, we can transform the two dependent measures into one dependent measure. A variety of measures that integrate accuracy and response time have been proposed and used in cognitive research. The integrative measure of RCS, which is best conceptualized as "correct responses per second," was shown in a series of simulated data studies to account for additional variance than either of the component measures alone (Vandierendonck, 2017). RCS for any given participant was calculated by dividing their number of correct responses by the sum total of all response times for that particular participant. This formula was taken from Woltz & Was (2006), and is expanded upon in the Supplemental Analyses. Further details about the performance on the three tasks used in Study 2 are reported below.

Performance Measures on the BST

Mean accuracy for the aphantasia group on the backwards spelling task (M = 83.0, SD = 12.4) was significantly higher than the mean performance of the control group (M = 76.5, SD = 15.3; $t(212)$ = 3.35, p = .001, d = .467). After entering age and sex as covariates with ANCOVA, the difference in accuracy on the BST between groups remained significant, with the estimated marginal mean of the aphantasia group (M = 82.7%, SE = 1.7%) greater than the estimated marginal mean of the control group (M = 76.7%, SE = 1.5%; (F (1, 210) = 5.41, p = .021, η^2 = .025, observed power = .639).

Mean response time in ms for the aphantasia group on the backwards spelling task (M = 14560, SD = 5604) was not significantly different from the mean response time of the control group (M = 14086, SD = 5459; $t(212)$ = .622, p = .534, d = .085). After entering age and sex as covariates with ANCOVA, the difference in response time on the BST between groups remained non-significant, with the estimated marginal mean of the aphantasia group (M = 14527, SE = 687) not significantly different from the estimated marginal mean of the control group (M = 14112, SE = 587; F (1, 210) = .164, p = .686, η^2 = .001, observed power = .069).

Rate correct score (RCS) between groups was investigated using an ANCOVA that entered age and sex as covariates (see Figure 9). There was no significant difference between groups in terms of rate correct score on the BST, with the aphantasia group (*M* = .065, *SE* = .003) having similar marginal means of performance, measured in correct responses per second, relative to the control group (*M* = .060, *SE* = .002; F (1, 210) = 1.73, p = .189, η_p^2 = .008, observed power = .259).

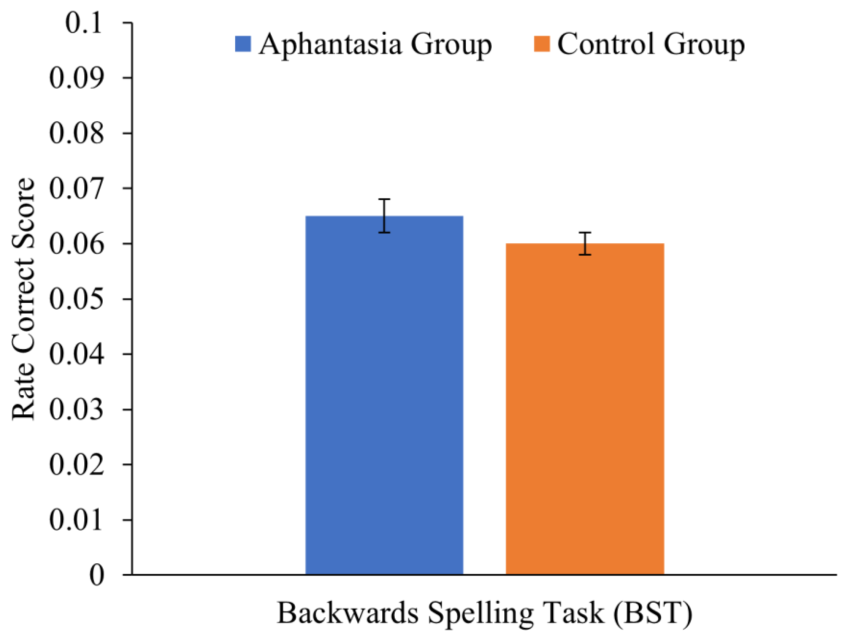


Figure 9. Performance on the BST as measured by rate correct score (correct responses per second). Error bars are standard error of the mean.

Performance Measures on the SPT

Mean accuracy for the aphantasia group on the snowy pictures task (*M* = 77.2, *SD* = 19.2) was significantly higher than it was for the control group (*M* = 65.0, *SD* = 22.5) after correcting for a significant Levene's test (F = 6.85, p = .009; t(210.52) = 4.27, p < .001, d = .589). After entering age and sex as covariates with ANCOVA, the difference in average accuracy on the SPT between groups remained *barely* significant, with the estimated marginal mean of the aphantasia group (*M* = 74.6%, *SE* = 2.5%) greater than that for the control group (*M* = 67.1%, *SE* = 2.2%; *F* (1, 210) = 3.91, *p* = .049, η^2 = .018, observed power = .503).

Mean response time in ms for the aphantasia group on the snowy pictures task (*M* = 7389, *SD* = 3745) was not significantly different from the mean response time of the control group (*M* = 7589, *SD* = 3637; *t*(212) = .310, *p* = .757, *d* = .043). After entering age and sex as covariates with ANCOVA, the difference in response time on the SPT between groups remained non-significant, with the estimated marginal mean of the aphantasia group (*M* = 7624, *SE* = 580) not differing significantly from the estimated marginal mean of the control group (*M* = 7405, *SE* = 496; *F* (1, 210) = .064, *p* = .800, η^2 = .000, observed power = .057).

Rate correct score (RCS) between groups was investigated using an ANCOVA that entered age and sex as covariates (see Figure 10). There was no significant difference between groups in terms of RCS on the SPT, with the aphantasia group (*M* = .142, *SE* = .009) having similar marginal means of performance, measured in correct responses per second, relative to the control group (*M* = .118, *SE* = .010; *F* (1, 210) = 2.45, *p* = .119, η_p^2 = .012, observed power = .344).

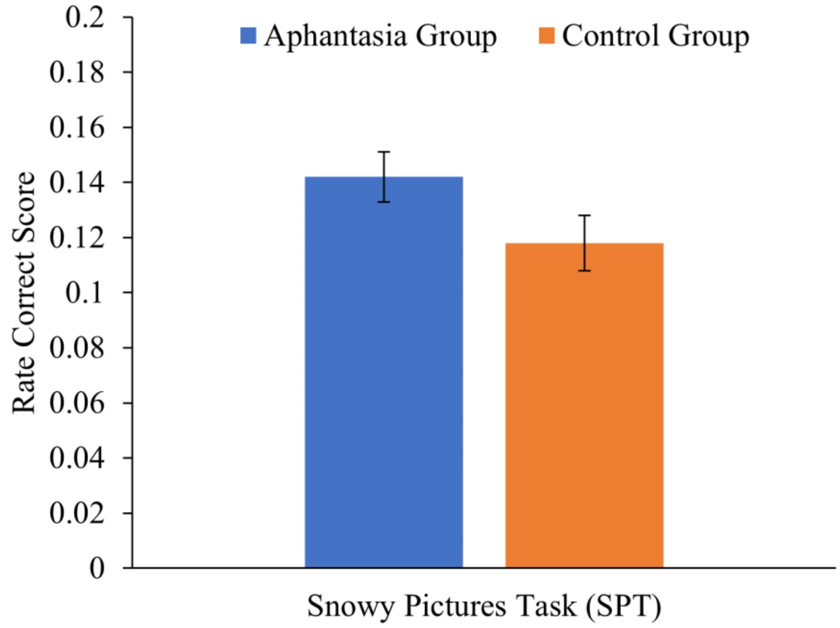


Figure 10. Performance on the SPT as measured by rate correct score (correct responses per second). Error bars are standard error of the mean.

Performance Measures on the TLT

Mean accuracy for the aphantasia group on the tail length task (*M* = 86.3, *SD* = 7.73) was significantly higher than the mean performance of the control group (*M* = 83.2, *SD* = 8.74; *t*(212) = 2.73, *p* = .007, *d* = .375). After entering age and sex as covariates with ANCOVA, the difference in average accuracy on the TLT between groups dropped out of significance, with the estimated marginal mean of the aphantasia group (*M* = 85.7%, *SE* = 1.0%) not significantly different from the estimated marginal mean of the control group (*M* = 83.7%, *SE* = 0.9%; *F* (1, 210) = 1.61, *p* = .207, η^2 = .008, observed power = .243).

Mean response time in ms for the aphantasia group on the tail length task (*M* = 1445, *SD* = 643) was significantly higher than it was for the control group (*M* = 1149, *SD* = 535) after correcting for a significant Levene's test (*F* = 6.98, *p* = .009; *t*(179.76) = 3.59, *p* < .001, *d* = .536). After entering age and sex as covariates with ANCOVA, the difference in response time on the TLT between groups dropped out of significance, with the estimated marginal mean of the aphantasia group (*M* = 1389, *SE* = 72.6) similar to the estimated marginal mean of the control group (*M* = 1193, *SE* = 62.0; *F* (1, 210) = , *p* = .071, η^2 = .015, observed power = .440).

Rate correct score (RCS) between groups was investigated using an ANCOVA with age and sex as covariates (see Figure 11). There was no significant difference between groups for rate correct score on the TLT, with the aphantasia group (*M* = .814, *SE* = .032) having similar marginal means of performance, measured in correct responses per second, relative to the control group (*M* = .737, *SE* = .037; *F* (1, 210) = 1.93, *p* = .166, η_p^2 = .009, observed power = .283).

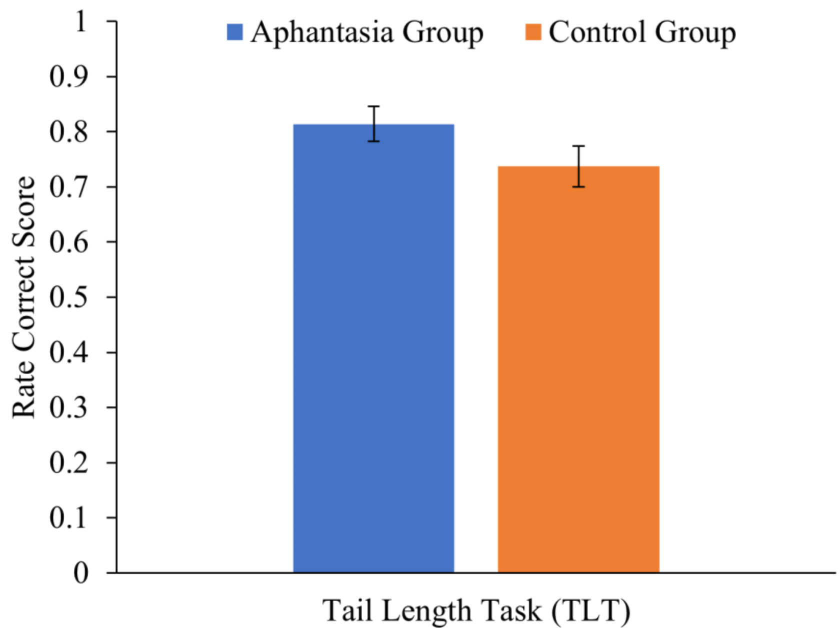


Figure 11. Performance on the TLT as measured by rate correct score (correct responses per second). Error bars are standard error of the mean.

Overall VVIQ Score Correlations

When considering all participants, VVIQ was negatively correlated with accuracy on all three tasks as well as with response time on the TLT (see below the diagonal in Table 2). After controlling for age and sex, the partial correlations between VVIQ and the performance measures dropped out of significance, with the exception of accuracy on the BST, which remained negatively correlated with VVIQ scores (see above the diagonal in Table 2).

Table 2. *Partial and Zero-Order Correlations of Study 2 Measures and VVIQ Including All Participants*

Variable	1	2	3	4	5	6	7
1. VVIQ	—	-.168*	.043	-.105	-.010	-.061	-.053
2. BST-ACC	-.221†	—	-.174*	.184†	-.040	.206†	-.071
3. BST-RT	.012	-.180†	—	-.032	.567†	.010	.472†
4. SPT-ACC	-.231†	.253†	-.042	—	-.263†	.056	.046
5. SPT-RT	.017	-.072	.569†	-.289†	—	.068	.510†
6. TLT-ACC	-.151*	.244†	.006	.123	.042	—	-.106
7. TLT-RT	-.181†	-.042	.470†	.089	.487†	-.066	—

Note. All participants from Study 2 are represented here ($N = 214$). Partial correlations controlling for age and sex appear above the diagonal while zero-order correlations appear below the diagonal. BST-ACC = average accuracy on the backwards spelling task. BST-RT = average response time on the backwards spelling task. SPT-ACC = average accuracy on the snowy pictures task. SPT-RT = average response time on the snowy pictures task. TLT-ACC = average accuracy on the tail length task. TLT-RT = average response time on the tail length task.
*$p < .05$. †$p < .01$

Intercorrelations with VVIQ Scores

After controlling for age and sex, VVIQ scores seemed to correlate with accuracy on the backwards spelling task, albeit in the opposite direction than would likely be predicted by researchers—people with lower mental imagery vividness as measured by the VVIQ tended to perform more accurately on the BST in this study. However, because the aphantasia group and the control group were recruited in different ways, there may have been pre-existing differences between the groups (e.g., motivation to complete the various cognitive tasks). A better test for whether VVIQ scores correlate with these three cognitive tasks is to look at the correlations within the control group only. Tellingly, when this is done, almost all evidence of correlation with VVIQ vanishes (see Table 3).

Next, I was interested in examining if there was a latent "imagery ability" variable hidden within the performance measures used in Study 2. Using principal axis factor analysis, I entered in the six performance measures from Study 2 (accuracies and response times for each of the three tasks), and I extracted latent variables with initial Eigenvalues above 1, resulting in two extracted factors (Kaiser-Meyer-Olkin Measure = .567). Initial Eigenvalues were 2.11 and 1.45, respectively, which together explained 59.36% of the variance. Following extraction, the first factor retained an Eigenvalue of 1.66 and explained 27.58% of the variance, while the second factor retained an Eigenvalue of .779 and explained 12.99% of the variance. Varimax rotation with Kaiser normalization was then applied, and rotated loadings greatly resembled the unrotated loadings, with the first factor retaining an Eigenvalue of 1.64 and explained 27.39% of the variance, and the second factor retaining an Eigenvalue of .790 and explained 13.17% of the variance. The loadings were clearly interpretable, with the first factor related to response times and the second factor related to accuracy (see Table 4).

The nature of this extraction process assumes that the sample used is equivalent to the population, which is incorrect. For example, not all possible VVIQ scores are represented in this limited sample. Therefore, the analysis of these latent variables is highly exploratory, and if researchers want to generalize these results, they will need cross-validation using additional, larger, samples (Field, 2017, p. 577).

Table 3. *Intercorrelations of Study 2 Measures and VVIQ for Controls Disaggregated by Video Condition*

Variable	1	2	3	4	5	6	7	8	9
1. VVIQ	—	-.152	.261*	.065	.037	.084	.082	.173	-.064
2. BST-ACC	.076	—	-.014	.313*	.150	.257*	.095	.091	.929†
3. BST-RT	.066	-.189	—	.122	.490†	.202	.376†	.817†	.132
4. SPT-ACC	-.011	.155	.116	—	-.065	.126	.155	.127	.517†
5. SPT-RT	.024	-.170	.650†	-.221	—	.196	.553†	.824†	.157
6. TLT-ACC	.130	.347†	-.333*	.133	-.242	—	.005	.165	.515†
7. TLT-RT	.047	-.301*	.604†	.121	.532†	-.262	—	.767†	.130
8. RT Factor	.051	-.252	.886†	.014	.864†	-.331	.811†	—	.181
9. ACC Factor	.101	.927†	-.148	.424†	-.261	.556†	-.242	-.249	—

Note. The results for the control participants who watched the imagery video (n = 64) are shown above the diagonal. The results for the control participants who watched the color perception video (n = 56) are shown below the diagonal. BST-ACC = average accuracy on the backwards spelling task. BST-RT = average response time on the backwards spelling task. SPT-ACC = average accuracy on the snowy pictures task. SPT-RT = average response time on the snowy pictures task. TLT-ACC = average accuracy on the tail length task. TLT-RT = average response time on the tail length task. RT Factor & ACC Factor = extracted latent variables from the six performance measures (BST-ACC, BST-RT, SPT-ACC, SPT-RT, TLT-ACC, and TLT-RT).
*p < .05. †p < .01

Table 4. *Results From a Factor Analysis of the Dependent Measures of Study 2.*

Item	Factor Loading	
	1	2
BST-ACC	-.066	**.702**
BST-RT	**.755**	.005
SPT-ACC	.061	**.365**
SPT-RT	**.753**	-.038
TLT-ACC	-.064	**.401**
TLT-RT	**.704**	-.033

Note. All controls are included (n = 120). Factor loadings above .300 are bolded and underlined.

Within the aphantasia group, there were no significant correlations between VVIQ scores and the performance measures (accuracy and response time) from Study 2, regardless of whether the correlation included all participants with aphantasia, just participants with aphantasia who reported few dreams (self-reported dream imagery frequency scores of 1–2), or just participants with aphantasia who reported at least some dreams (self-reported dream imagery frequency scores of 3–6). This lack of significant correlations makes intuitive sense because the vast majority of participants with aphantasia (67%) had a VVIQ score of 16, severely restricting the range. See the Supplemental Analyses for additional correlation tables showing these results.

Moderation Models using PROCESS

Next, I used PROCESS to test whether video condition acted as a moderator on the relationship between VVIQ scores and the six performance measures as well as the two extracted factors (Model 1; Hayes, 2018). In short, video condition was never a significant moderator on any of these relationships. See the Supplemental Analyses for a table depicting the universally non-significant results of these moderation analyses.

Self-Reported Imagery Use

Self-Reported Imagery Use on the BST

As measured by the self-reported imagery use (SIU) question, the aphantasia group reported significantly less use of visual mental imagery during the BST (n = 94, M = 1.03, SD = .177) than did the control group (n = 120, M = 2.38, SD = 1.02) after adjusting the degrees of freedom for a significant Levene's test (F = 250.99, p < .001; t(128.03) = 14.14, p < .001, d = 2.50). Within the control group, participants who were randomly assigned to watch the imagery video reported significantly less use of visual mental imagery during the BST (n = 64, M = 2.20, SD = 1.06) than did the control participants who watched the color perception video (n = 56, M = 2.57, SD = .951; t(118) = 2.00, p = .048, d = .368).

SIU on the BST had a strong positive correlation with VVIQ scores (r(212) = .657, p < .001). This was true even after controlling for age and sex (partial r(210) = .562, p < .001). When considering only participants with aphantasia, the correlation between SIU on the BST and VVIQ scores was strong and positive (r(92) = .329, p = .001). When considering only participants in the control condition, the correlation is smaller (r(118) = .208, p = .023). The difference between these two correlations is not significant (Fisher's z = .935, p = .350). When considering only participants in the control condition who watched the imagery video, that correlation is not significant (r(62) = .233, p = .064). Similarly, when only participants in the control condition who watched the color video are considered, the relationship is not significant (r(54) = .080, p = .559).

Within the control group, SIU on the BST did not appear to be correlated with performance on the BST in terms of accuracy (r(118) = .048, p = .603), response times (r(118) = -.002, p = .984), or rate correct score (r(118) = .003, p = .974). This continued to be true when the control group was split by video condition. For control participants who watched the imagery video, there was no correlation between SIU on the BST and performance on the BST as measured by accuracy ($r(62) = -.011$, $p = .932$), response time ($r(62) = -.065$, $p = .610$), or rate correct score ($r(62) = .040$, $p = .753$). For control participants who watched the color perception video, there was no correlation between SIU on the BST and performance on the BST as measured by accuracy ($r(54) = .099$, $p = .470$), response time ($r(54) = .060$, $p = .662$), or rate correct score ($r(54) = -.071$, $p = .604$).

Self-Reported Imagery Use on the SPT

As measured by the self-reported imagery use (SIU) question, the aphantasia group reported significantly less use of visual mental imagery during the SPT ($n = 94$, $M = 1.09$, $SD = .349$) than did the control group ($n = 120$, $M = 2.58$, $SD = .984$) after adjusting the degrees of freedom for a significant Levene's test ($F = 139.97$, $p < .001$; $t(155.12) = 15.39$, $p < .001$, $d = 2.47$). Within the control group, participants who were randomly assigned to watch the imagery video did not significantly differ in terms of their SIU on the SPT ($n = 64$, $M = 2.50$, $SD = 1.01$) when compared to the control participants who watched the color perception video ($n = 56$, $M = 2.66$, $SD = .959$; $t(118) = .891$, $p = .375$, $d = .164$).

SIU on the SPT had a strong positive correlation with VVIQ scores ($r(212) = .689$, $p < .001$). This was true even after controlling for age and sex (partial $r(210) = .577$, $p < .001$). When considering only participants with aphantasia, the correlation between SIU on the SPT and VVIQ scores was strong and positive ($r(92) = .369$, $p < .001$). When considering only participants in the control condition, the correlation is smaller ($r(118) = .187$, $p = .041$). The difference between these two correlations is significant (Fisher's $z = 1.42$, $p = .157$). However, when considering only participants in the control condition who watched the imagery video, that relationship is not significant (*r*(62) = .106, *p* = .404). When only participants in the control condition who watched the color video are considered, the relationship is moderate and positive (*r*(54) = .288, *p* = .032).

Within the control group, SIU on the SPT correlated with performance on the SPT in terms of accuracy (*r*(118) = .183, *p* = .045) but not for response times (*r*(118) = .000, *p* = .998) or rate correct scores (*r*(118) = .019, *p* = .841). For control participants who watched the imagery video, there was a significant correlation between SIU on the SPT and performance on the SPT as measured by accuracy (*r*(62) = .268, *p* = .032), but not for response time (*r*(62) = -.014, *p* = .914) or rate correct score (*r*(62) = .181, *p* = .153). For control participants who watched the color perception video, there was no correlation between SIU on the SPT and performance on the SPT as measured by accuracy (*r*(54) = .057, *p* = .675), response time (*r*(54) = -.002, *p* = .987), or rate correct score (*r*(54) = -.143, *p* = .295).

Self-Reported Imagery Use on the TLT

As measured by the self-reported imagery use (SIU) question, the aphantasia group reported significantly less use of visual mental imagery during the TLT (*n* = 94, *M* = 1.29, *SD* = .598) than did the control group (*n* = 120, *M* = 3.53, *SD* = .634) after adjusting the degrees of freedom for a significant Levene's test (*F* = 5.85, *p* = .016; *t*(204.83) = 26.55, *p* < .001, *d* = 3.71). The effect here was enormous, with participants in the aphantasia group overwhelmingly (94.7%) choosing the minimum response of "1" or a response of "2" while controls overwhelmingly (94.2%) chose "3" or the max response of "4," indicating that people with aphantasia did not use mental imagery to complete the TLT while controls depended upon imagery to complete the TLT (See Figure 12). Within the control group, participants who were randomly assigned to watch the imagery video did not significantly differ in terms of SIU on the

TLT (n = 64, M = 3.52, SD = .642) when compared to the control participants who watched the color perception video (n = 56, M = 3.55, SD = .630; $t(118)$ = .326, p = .745, d = .060).

SIU on the TLT had a strong positive correlation with VVIQ scores ($r(212)$ = .863, p < .001), even after controlling for age and sex (partial $r(210)$ = .790, p < .001). When considering only participants with aphantasia, the correlation between SIU on the TLT and VVIQ scores was strong and positive ($r(92)$ = .762, p < .001). When considering only participants in the control condition, the correlation is smaller ($r(118)$ = .228, p = .012). The difference between these two correlations is significant (Fisher's z = 5.50, p < .001). When considering only participants in the control condition who watched the imagery video, that correlation is not significant ($r(62)$ = .173, p = .171). When only participants in the control condition who watched the color video are considered, the relationship is moderate and positive ($r(54)$ = .316, p = .018).

Within the control group, SIU on the TLT correlated with performance on the TLT in terms of accuracy ($r(118)$ = .242, p = .008) but not for response times ($r(118)$ = -.044, p = .637) or rate correct scores ($r(118)$ = .045, p = .626). For control participants who watched the imagery video, there were no significant correlations between SIU on the TLT and performance on the TLT as measured by accuracy ($r(62)$ = .235, p = .062), response time ($r(62)$ = .054, p = .671), or rate correct score ($r(62)$ = -.024, p = .849). For control participants who watched the color perception video, there was no correlation between SIU on the TLT and performance on the TLT as measured by accuracy ($r(54)$ = .255, p = .058), response time ($r(54)$ = -.133, p = .330), or rate correct score ($r(54)$ = .128, p = .346).

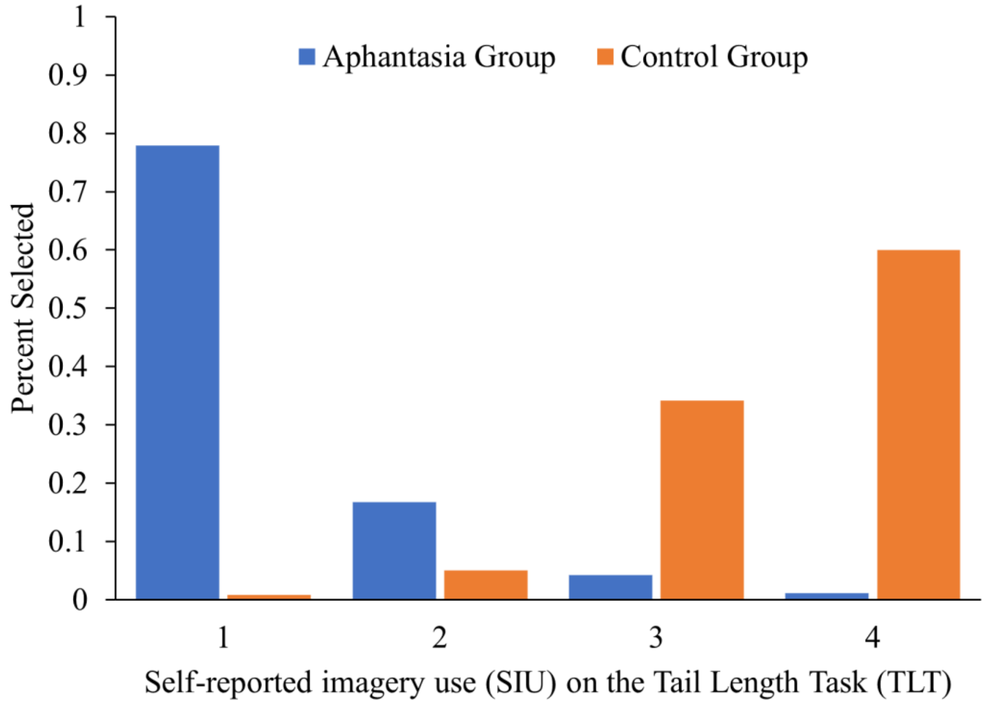


Figure 12. Self-reported usage of visual mental imagery (SIU) during the TLT. Scores from one to four correspond respectively to responses of “I did not use mental images,” “I barely used them,” “I used them quite a bit,” and “I used them a whole lot.”

Study 2 Discussion

The lack of evidence from Study 2 seems to indicate that the VVIQ is a poor measure to use when attempting to correlate subjective mental imagery “ability” to objective mental imagery “use.” However, I must be careful in how I approach this argument.

Because this is a quasi-experiment, it could be argued that it isn’t really fair to compare the “total” performance of the aphantasia group and the control group on these tasks. For example, the participants were recruited in different ways, and I did not ask participants how much effort they exerted during the trials, and so I cannot say for sure whether the aphantasia group was more interested in the tasks, or if there were other uncontrolled group differences. Such pre-existing differences between the groups could obscure differences in performance or various correlations when the entire sample is analyzed. Furthermore, unlike Studies 3 and 4, there were no “within-participant” experimental manipulations within these tasks. Therefore, there are no group x manipulation interactions to examine within the tasks that could show differences in how the groups approached the tasks (e.g., manipulable dependencies on the use of mental imagery per trial within each task). Therefore, some of my conclusions focus on patterns *within* the control group (which eliminates the pre-existing differences issue), and my conclusions that compare people with aphantasia to controls are tempered by controlling for age and sex (which differed significantly between the groups, see Study 1).

Within the Control Group

The main goal of Study 2 was to see if the correlations between VVIQ scores and performance measures on a variety of cognitive tasks could be strengthened if participants first watched a video informing them about individual differences in mental imagery vividness, when compared to participants who watched an irrelevant video (e.g., one about individual differences in color perception). Watching the imagery video was theorized to improve the “calibration” of the VVIQ scores to better match task performance. However, this result was mostly not found. Despite picking three cognitive tasks that had been previously established in the literature to have a relationship with VVIQ scores (i.e., BST, SPT, and TLT), and measuring two performance measures (accuracy and response times), only one relationship (response time on the backwards spelling task for participants who watched the imagery video) even attained *significance*. The VVIQ also did not correlate with “latent” variables extracted from the tasks. Furthermore, there was no evidence found in terms of the video condition acting as a moderator on the relationship between VVIQ scores and the performance measures on the three tasks.

Between Groups

Overall, the results suggest no group differences in performance on the three tasks. After controlling for age and sex, the aphantasia group performed better at two of the three tasks in terms of accuracy (BST and SPT), but when responses were measured as “correct responses per second” (rate correct score; RCS) which accounts for both accuracy and response time, there were no significant group differences.

When analyzing all participants, after controlling for age and sex, VVIQ scores and the performance measures were almost universally uncorrelated, with the exception of accuracy on the BST, despite previous investigations claiming the existence of relationships between these tasks and the VVIQ.

Why are the VVIQ and Performance Correlations Poor and Missing?

Why was there a lack of correlations between VVIQ scores and performance on the tasks? There are several possibilities.

Maybe the VVIQ Measures Vividness but not the Use of Mental Images

One possibility is that a person’s mental imagery *vividness* as measured by the VVIQ does not strongly predict a person’s *use* of mental imagery during a given task. That is, having the *ability to generate* vivid mental images may not necessarily lead a person to *use* such vivid images during a task for which such images might be used.

However, when all participants are considered, self-reported use of imagery during these three tasks was strongly correlated with VVIQ scores, even after controlling for age and sex, which seems to suggest that the VVIQ *does* capture at least some measurement of the *use* of mental imagery during these tasks. The picture becomes clearer as we look within the groups. VVIQ scores tended to be more weakly correlated with self-reported imagery usage within the control group than when looking within the aphantasia group (although, the difference between the larger correlations in the aphantasia group and the smaller correlations in the control group attained significance only for the TLT, due to the limited sample size).

Why might this be? As mentioned as part of Study 1, stronger correlations for the aphantasia group between VVIQ scores and self-reported use of imagery during tasks makes intuitive sense because people with lower VVIQ scores don't really have a choice in whether they use depictive or propositional strategies—they must use propositional strategies, and therefore their "imagery use" scores should be almost universally low. However, for participants with higher VVIQ scores (e.g., in the control group), they should have more choices as to how to approach a task. For example, when asked whether a rabbit's tail is long or short they could either depict a rabbit by generating a mental image, or, they could use semantic knowledge about the tail length of rabbits (i.e., a propositional strategy). Because they have this option of strategy, their "imagery use" scores are more variable, making the correlation with VVIQ scores weaker (see Figure 12). Thus, the VVIQ does seem to correlate with the self-reported use of mental imagery on a task in an interpretable fashion, and therefore, it would be expected for VVIQ score to correlate with task performance measures—but again, such correlations are not found.

However, there *are* correlations between self-reported mental imagery use (SIU) during tasks and the performance on those tasks. There were significant correlations within the control group (i.e., the participants who *had the option* of using either depictive or propositional methods of solving the tasks) between SIU and the accuracy on the respective tasks. Such correlations existed for the SPT and the TLT. That is, there *was* a relationship between self-reported imagery use and one measure of performance, suggesting that using mental images versus not using them did make a difference! Is it possible that the relationship between the VVIQ and self-reported mental imagery use is just too weak for VVIQ scores to have a detectable relationship with the performance measures? Such a mediation investigation is beyond the scope (and sample size, due to the low number of low VVIQ scores in the control group) of the present study. Regardless of whether the VVIQ is a valid measure of a person's *vividness* of mental imagery, for people who can make use of mental imagery, the VVIQ may not be a valid measure of their choice to *use* mental imagery for any given cognitive task.

Maybe Performance on the Tasks Doesn't Use Mental Imagery Vividness

An alternative explanation is that VVIQ scores are generally not correlated or poorly correlated with the performance measures on these three cognitive tasks because performance on these tasks does not make use of mental images, *even if a person believes otherwise*. That is, accuracy and response time may depend on some other cognitive calculation that is entirely separate from the *experience* of a mental image (i.e., the explicit depiction has no bearing on accuracy or response time). As discussed in the Introduction, the experience of a mental image may be purely epiphenomenal, and therefore, no differences in performance would be expected based on whether a person experienced a mental image. The implicit/explicit theory would suggest this interpretation.

Perhaps, then, solving these three cognitive tasks that have been purported to use mental imagery does not actually require tapping into mental imagery vividness in order to achieve accurate and quick completion. The backwards spelling task, for example, allows a person to use *recognition* to compare a string of letters that they are currently typing out, rather than *recall* in which a person may generate a vivid mental image—the second approach certainly taps more mental imagery than the first. The tail length task, as another example, may not require a more *vivid* depiction of a rabbit in order to more accurately and more quickly determine the length of its tail. That is, a barely-there non-vivid depictive representation of a rabbit may be just as sufficient for the task as a highly vivid depictive representation of a rabbit.

Maybe Using Images During these Tasks Doesn't Impact Task Performance

Another possibility is that people can use mental images to complete the tasks in a way that the task completion depends on the mental image (i.e., the depiction is made use of in the cognitive calculations), but using depiction does not produce a superior result—at least not detectably—when compared to completing the tasks using propositions. This is what the perceptual/conceptual theory would suggest. That is, being forced to use propositions to complete the three tasks due to having impoverished mental imagery may not interfere with the task in a detectable way. As James (1890a) would suggest, perhaps the destination is the same (e.g., determining the relative length of a tail), but the internal representations are wildly different (e.g., depictive vs. propositional).

Takeaway Points

The most important takeaway from the results of Studies 1 and 2 is that VVIQ scores are potentially a poor way of approaching mental imagery research. VVIQ scores were almost entirely unrelated to any measure of task performance in Study 2 (with the exception of a minor relationship with the BST, see Tables 2 and 3). Also, not only were VVIQ scores significantly affected by the video manipulation (i.e., watching a video about individual differences in imagery vividness made people report less vivid imagery), but the video manipulation also significantly affected self-reported imagery usage during the BST, such that control participants who watched the imagery video were less likely to report the usage of imagery than participants who watched the unrelated color perception video! If subjective reports of mental imagery vividness and use can be pushed around so easily, perhaps questionnaires such as the VVIQ need more strict rules of administration to ensure validity. This is further supported by existing literature demonstrating the impact of demand characteristics (e.g., experimenter effects, social desirability) on self-reported imagery vividness scores (e.g., Ahsen, 1993; Di Vesta et al., 1971; Farah, 1988; J. T. E. Richardson, 1980; McKelvie, 1995; Winograd et al., 1998).

Fortunately, efforts are already underway to construct a new self-report measure that will be specific to aphantasia research (Reeder, 2022). Therefore, this series of studies is ready to leave VVIQ scores behind and focus on self-identified group membership (e.g., whether a person identified themselves as having aphantasia). Importantly, even the control group was given the option to report as having aphantasia, and those that did were included in the aphantasia group. Are there any cognitive tasks that will reveal group differences between the participants reporting aphantasia and those who did not in terms of task performance? Studies 3 and 4 investigate this and will further test the theories of aphantasia with within-task manipulations.

Furthermore, within the control group, whether or not accuracy on the SPT correlated with self-reported imagery use (SIU) depended on which video condition was watched by the participant. That is, the correlation existed for the imagery video group ($r(62) = .268$, $p = .032$) but not for the color perception video group ($r(54) = .057$, $p = .675$). Was this a fluke, or were the control participants naïve in their understanding of mental imagery to the point that watching a video about individual differences of imagery allowed them to better calibrate their estimates about whether they were making use of imagery that was then better related to their actual performance on a task? In order to explore this, SIU was also investigated in Studies 3 and 4 to determine if the video watched mattered for the correlation with measures of task performance.

CHAPTER 4. STUDY THREE

If the classic cognitive task of mental scanning depends upon mental imagery, then shouldn't people with aphantasia perform differently from controls on such a task? This study examines that question. Specifically, this study compares the performance of people with aphantasia against controls on a mental scanning task that purportedly makes use of mental imagery: the square donut scanning task (see "Objective Measurement of Imagery Ability").

The dependent variable of most interest is reaction time because this was the dependent variable used in the previously conducted studies that used the square donut scanning task (Dror et al., 1993; Dror & Kosslyn, 1994). This task has not been previously used in conjunction with people with aphantasia.

Two within-participant independent variables are also crucial for this study: "distance" and "squares." First, in order to better understand the role of scanning distance in the performance of the groups, two different scanning distances are used: near and far. Secondly, two different "squares" conditions that varied in relative cognitive load were introduced; in one condition, the locations of three black squares needed to be mentally represented, while in the other condition, the locations of six black squares demanded representation.

The distance manipulation is based on a large amount of existing research that demonstrated increased response times when important stimuli were separated by more physical space (e.g., Dror & Kosslyn, 1994; see "The Great Imagery Debate" in the Introduction). At least, that is the relationship when mental imagery is used during a cognitive task that encourages "mental scanning" of previously presented stimuli. However, if the perceptual/conceptual model of aphantasia is correct, then the speed at which people with aphantasia respond to tasks that depend on mental imagery should not differ depending on the depictive visual properties of the task—i.e., "mental scanning” a near distance should take the same amount of time as scanning a far distance in people with aphantasia if the distance is being represented propositionally instead of being depicted. More specifically, if there is a significant interaction—slope differences between the near-distance and far-distance versions of the task—between the groups such that the aphantasia group is less affected by the distance manipulation than is the control group, this finding would support the perceptual/conceptual model. In other words, it would be evidence that people with aphantasia do not use depictive mental imagery because they were not affected by a manipulation of distance that has been established to be related to mental imagery use.

On the other hand, if the implicit/explicit model is correct, then the visual properties of the task should affect the response times of the people with aphantasia even if they do not report consciously experiencing mental imagery. If there is a significant effect in response times between the near-distance and far-distance versions of the task both for people with aphantasia and for controls, but no distance x group interaction, then this would be evidence for the implicit/explicit theory. In other words, it would be evidence that people with aphantasia do mentally represent distance, because their response time patterns are the same as people without aphantasia. Therefore, it could be argued that because scanning distances affected response times, people with aphantasia make use of depictive mental images, even if they are not aware of them. This prediction is based on the main effect of scanning distance detected by Dror & Kosslyn (1994) in which it takes the average person longer to respond to a “far” trial than to a “near” trial, ostensibly because mental representations depict distance which is then “scanned.”

When it comes to the manipulated number of squares, the intent here is to see if the relationship between group and performance on the square donut scanning task depends upon the demands of the task. If the perceptual/conceptual theory of aphantasia is correct, a person with aphantasia should be able to represent a 3-square trial using a propositional representation more easily than a 6-square trial, allowing them to respond faster (e.g., a propositional list of three items denoting locations such as "bottom-left, top-right, right-middle-bottom" can be represented and searched more quickly than a propositional list containing six items denoting locations). In contrast, if people in the control group are using depictive mental imagery, they should have a more similar level of difficulty for the 3-square trials and the 6-square trials, because they are supposedly using a picture of all of the possible squares, and not a propositional list, and such a picture of squares should take up the same number of mental "pixels" regardless of how many squares are black and how many are white. Therefore, participants with aphantasia should show a clear deficit on the 6-square trials relative to the 3-square trials, all relative to the same comparison in the control group (i.e., there should be a slope difference; an interaction effect). Under the implicit/explicit theory, such a slope difference should not be detected due to all participants using a similar depictive strategy, but there may or may not be a main effect of number of squares due to the increase in difficulty (number of squares was not tested in Dror & Kosslyn, 1994).

To summarize, if the perceptual/conceptual theory is correct, then the "distance" manipulation should impact the control group more than the aphantasia group because of the propositional strategy of the aphantasia group, while the "squares" manipulation should impact the aphantasia group more than the control group because of the depictive strategy of the control group. Both of these effects should be detectable as between-within interaction effects, where the group that is more impacted should show a steeper slope between the two categories for that within-participants manipulation.

Additionally, a three-way interaction was of potential interest (group x distance x squares), with participants in the aphantasia group possibly showing a slope difference for response times between scanning distances relative to controls, but dependent on the number of squares—or a similar complex relationship. In one study with a participant with developmental aphantasia, only the most difficult of a series of visual working memory tasks that required fine discrimination were impaired relative to controls (Jacobs et al., 2018). Therefore, different patterns of effects are possible at the different levels of each variable.

Method

Participants

The participants were the aphantasia group and the control group from Study 1. There were 95 participants in the aphantasia group and 120 participants in the control group.

Materials

Participants completed a version of the square donut scanning task (Dror et al., 1993; Dror & Kosslyn, 1994). Specifically, this task was designed to resemble the mental imagery task from Experiment 3 of Dror and Kosslyn (1994). See Appendix F for a list of changes made between the original version of the task and the version used in this study.

The task was administered remotely through an online experimental procedure in conjunction with the other studies (see Study 2). An example illustration of this task can be seen in Figure 3 in the Introduction chapter, and in Appendix F. The task was designed with three levels, one between and two within, creating a 2 (aphantasia x control) by 2 (near x far) by 2 (3-square x 6-square) matrix of conditions. The square donut and arrow stimuli were created with image-editing software and displayed as static images. For all stimuli, black squares never appeared in corners. For the 3-square stimuli, black squares never shared a side. For the 6-square stimuli, all sides featured at least one black square, and no side featured more than two black squares (see Figure 13).

Procedure

Participants completed the square donut scanning task on a personal computer. On each trial, a blank screen appeared for 500 ms. Participants then viewed a “donut” stimulus consisting of an arrangement of twenty squares, six to a side, such that they formed a larger hollow square (see Figure 13). Most of the squares were white but some of the squares (3 or 6) were black. Participants studied the stimulus for 750 ms. The donut stimulus then faded. A visual mask was presented for 750 ms. An arrow then appeared on the screen, pointing to the previous position of one square. The participant was asked to indicate if the square in question was colored white or black using one of two keys on the keyboard (“W” if white, and “B” if black). The response key options were presented on the screen during all trials. Feedback on accuracy was not provided, except during practice trials. The number of completed trials, as well as the total number of trials to be completed, were displayed during the task.

Half of the trials consisted of “near” trials in which the arrow appeared close in proximity to the square requiring analysis. The other half of the trials were “far” trials, such that the arrow was located further away from the square in question. Additionally, half of the trials were “3-square” trials, in which three of the twenty squares were colored black. The other half of the trials were “6-square” trials, in which six of the squares were colored black. It was intended that the 6-square trials would be challenging: more difficult to complete quickly and accurately.

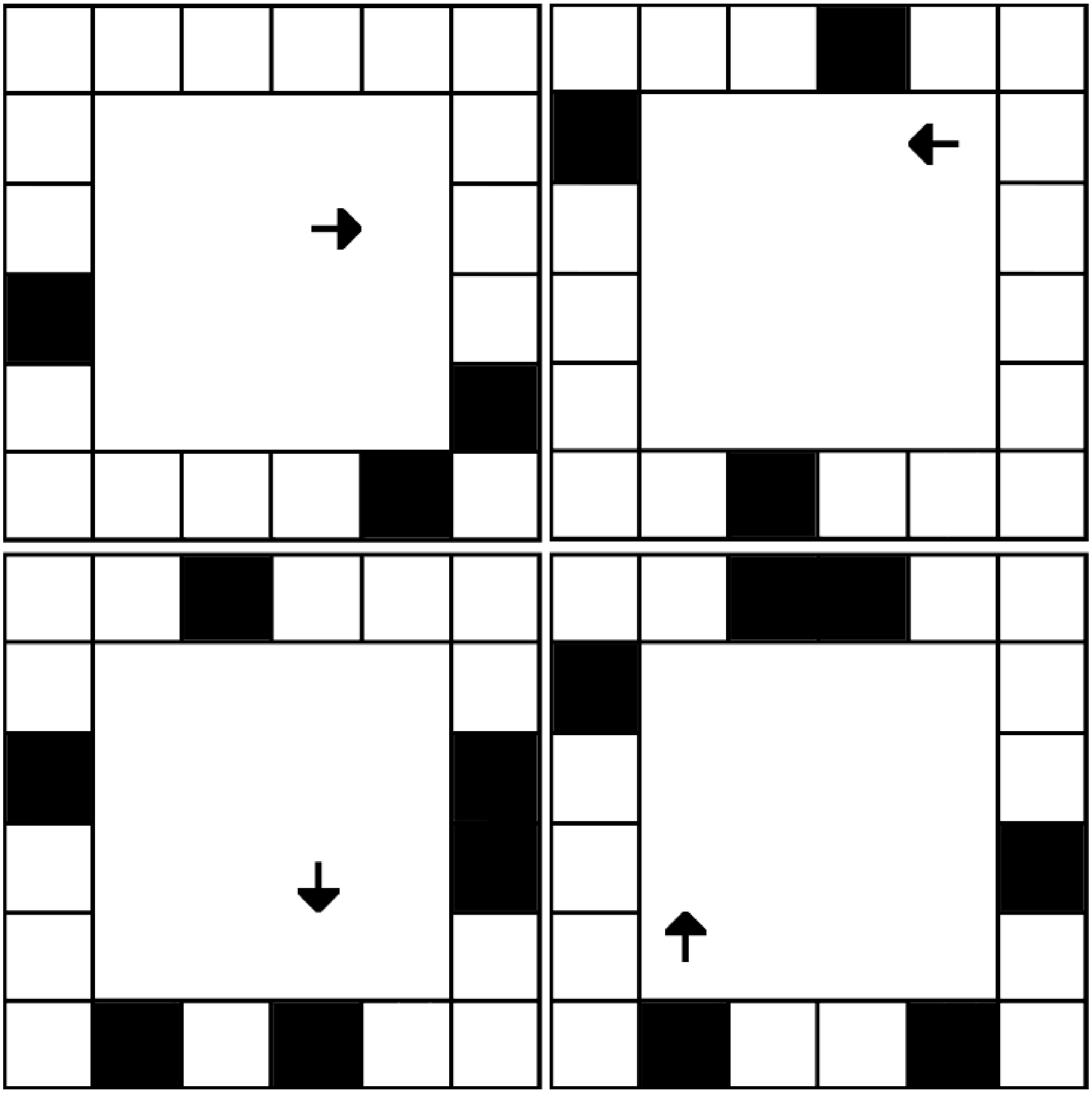


Figure 13. Four trials that were used during Study 3. The upper left trial is a 3-square near trial in which the correct answer was “white.” The upper right trial is a 3-square far trial in which the correct answer was “black.” The lower left trial was a 6-square near trial for which the correct answer was “black.” The lower right trial was a 6-square far trial for which the correct answer was “white.” Please note that this figure is for illustrative purposes, and at no point during the actual procedure were the arrows and square donut stimuli on the screen at the same time. A visual mask was always used to space them. See Appendix F for a depiction of this procedure.

Participants completed eight practice trials prior to the criterion trials in order to get used to the task. There were 64 criterion trials, with 32 trials at each distance and 32 trials of each black square number (3 or 6), resulting in 16 trials of each distance x square combination. Dror

& Kosslyn (1994) had 16 participants and detected a difference in response time as well as error rate between the far and near scanning trials with a mere 20 trials of each distance.

Study 3 Results

All datasets used for this study have been anonymized and can be accessed through the OSF project page, and are usable with attribution (i.e., cite this project: Toftness, 2022; https://osf.io/u3bxj/).

Overall Performance on Study 3

Raw performance measures were investigated before attempting to control for age or sex. Overall response time on the square donut task, measured in ms, significantly differed between groups, such that the aphantasia group (M = 1171, SD = 369) had longer average response times than the control group (M = 972, SD = 203) after correcting for a significant Levene's test (F = 20.5, $p < .001$; $t(138.25) = 4.73$, $p < .001$, d = .805). Overall accuracy on the square donut scanning task was not significantly different between groups, with the aphantasia group (M = 81.1, SD = 9.20) scoring about as well as the control group (M = 80.7, SD = 11.0; $t(213) = .244$, $p = .808$, d = .033). Group differences in performance measures after controlling for age and sex are discussed as part of the ANCOVA analyses below.

ANCOVA Investigation into Response Times

Between the measures of response time and accuracy, response time is of most interest because it was the examined dependent measure in the original studies using a similar square donut scanning task (Dror et al., 1993; Dror & Kosslyn, 1994). ANCOVA was used to examine the response times of participants. A 2 x 2 x 2 mixed ANCOVA compared group, scanning distance, and number of black squares ((aphantasia x control) x (near x far) x (three-square x six-square)) while entering age and sex as covariates.

Covariates

There was a main effect of age on response times (F (1, 211) = 54.38, p < .001, η_p^2 = .105, observed power = .999). Older participants had longer response times than younger participants such that the correlation was strikingly positive (r(213) = .443, p < .001). The ANCOVA model evaluated age at a value of 25.48 for the estimated marginal means reported below, which, for example, brought the mean response times of the aphantasia group (comprised of older people) and the control group (comprised of younger people) closer together. There was not a main effect of sex on response times (F (1, 211) = 2.83, p = .094, η_p^2 = .013, observed power = .387). The ANCOVA model evaluated sex at a computed value of 1.60 (where 1 = male and 2 = female).

There was also a significant interaction between age and number of squares (F (1, 211) = 5.79, p = .017, η_p^2 = .027, observed power = .668). Age was further investigated categorically by entering it into the ANCOVA as a between-subjects factor rather thana covariate (age > 25.48 were assigned as "older" and age < 25.48 were assigned as "younger"; see Figure 14). Simple effects were then calculated. There was no significant differences detected in reaction time between the younger participants and the older participants when looking at only the 3-square trials (F (1, 210) = .193, p = .661, η_p^2 = .001, observed power = .072) nor when looking at only the 6-square trials (F (1, 210) = .476, p = .491, η_p^2 = .002, observed power = .106). However, as would be expected, there was a significant difference in reaction time between the 3-square trials and the 6-square trials when looking at only the younger participants (F (1, 210) = 35.41, p < .001, η_p^2 = .144, observed power = 1) and when looking at only the older participants (F (1, 210) = 7.95, p = .005, η_p^2 = .036, observed power = .801). No other interactions with age were significant, and no interactions with sex were significant (all ps > .05).

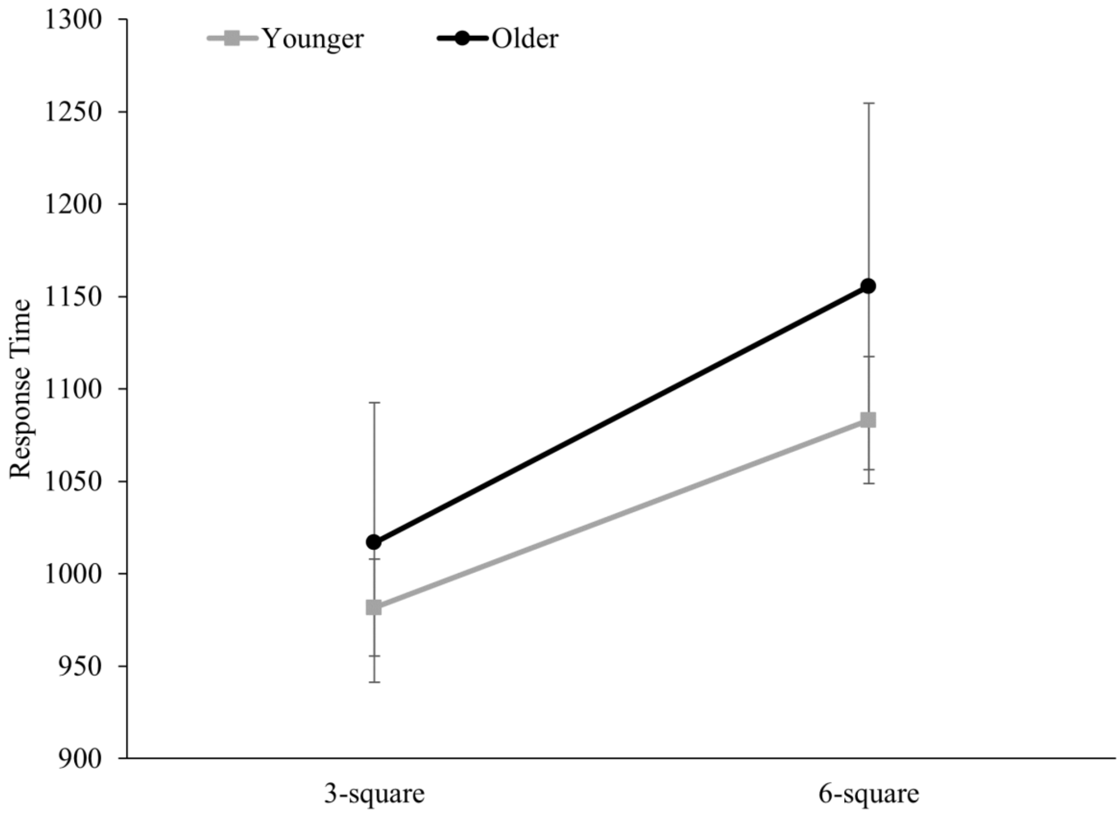


Figure 14. The significant age x squares interaction on response times for Study 3. For the purposes of graphing the interaction, age was split into younger (*n* = 147) and older (*n* = 68) groups at the age of 25.48 based on the ANCOVA-computed value. There was also a significant main effect of age when age was entered into the model as a continuous covariate. Response times appear in terms of ms on the y-axis. Mean response times have been adjusted for sex, and the estimated marginal means are depicted. Error bars are standard error.

Main Effects

A main effect was not detected for group: estimated marginal means were not significantly different, with participants in the aphantasia group having an estimated mean response time of 1085 ms (*SE* = 33.7) compared to the control group's average of 1039 ms (*SE* = 29.0; F (1, 211) = .830, p = .363, η_p^2 = .004, observed power = .148). There was, however, a main effect of scanning distance, with the near trials having an estimated marginal mean response time of 1002 ms (*SE* = 18.2) while for the far trials it was 1122 ms (*SE* = 20.7; F (1, 211) = 6.69, p = .010, η_p^2 = .031, observed power = .731). A main effect was not detected for the number of squares, with the estimated marginal means of response times for the 3-square trials (*M* = 1007, *SE* = 16.9) not differing significantly from the six-square trials (*M* = 1118, *SE* = 22.1, F (1, 211) = .424, p = .516, η_p^2 = .002, observed power = .099).

Two-Way Interactions

No significant interaction was detected between group and scanning distance (see Figure 15), with the difference in estimated marginal means between scanning distances relatively equal between groups while number of squares was held constant (F (1, 211) = .051, p = .821, η_p^2 = .000, observed power = .056). Because the implicit/explicit theory predicts that there should be simple effects here such that scanning distance was significant when considering *only* participants with aphantasia, the simple effects were investigated. Indeed, the simple effect of distance was significant when only the participants with aphantasia were considered (F (1, 211) = 41.05, $p < .001$, η_p^2 = .163, observed power = 1). It was also significant when only the participants in the control group were considered (F (1, 211) = 61.46, $p < .001$, η_p^2 = .226, observed power = 1). In both groups, the “far” scanning distance led to significantly longer response times than the ”near” scanning distance.

There was a highly significant interaction between group and squares (see Figure 16) such that, when scanning distance was held constant, the impact of the square manipulation on response time was larger on the aphantasia group than on the control group (F (1, 211) = 7.54, p = .007, η_p^2 = .035, observed power = .781). The simple effects of group x squares were then calculated using the estimated marginal means, revealing no simple effect between the groups at 3-squares (F (1, 211) = .009, p = .925, η_p^2 = .000, observed power = .051), nor a simple effect between the groups at 6-squares (F (1, 211) = 2.20, p = .140, η_p^2 = .010, observed power = .314). Simple effects were calculated between the levels of “squares” for each group using the estimated marginal means, revealing a significant simple effect of squares for the aphantasia group (F (1, 211) = 56.27, p < .001, η_p^2 = .211, observed power = 1) and a significant simple effect of squares for the control group (F (1, 211) = 15.57, p < .001, η_p^2 = .069, observed power = .975), such that response times for the 6-square trials were significantly longer in both groups.

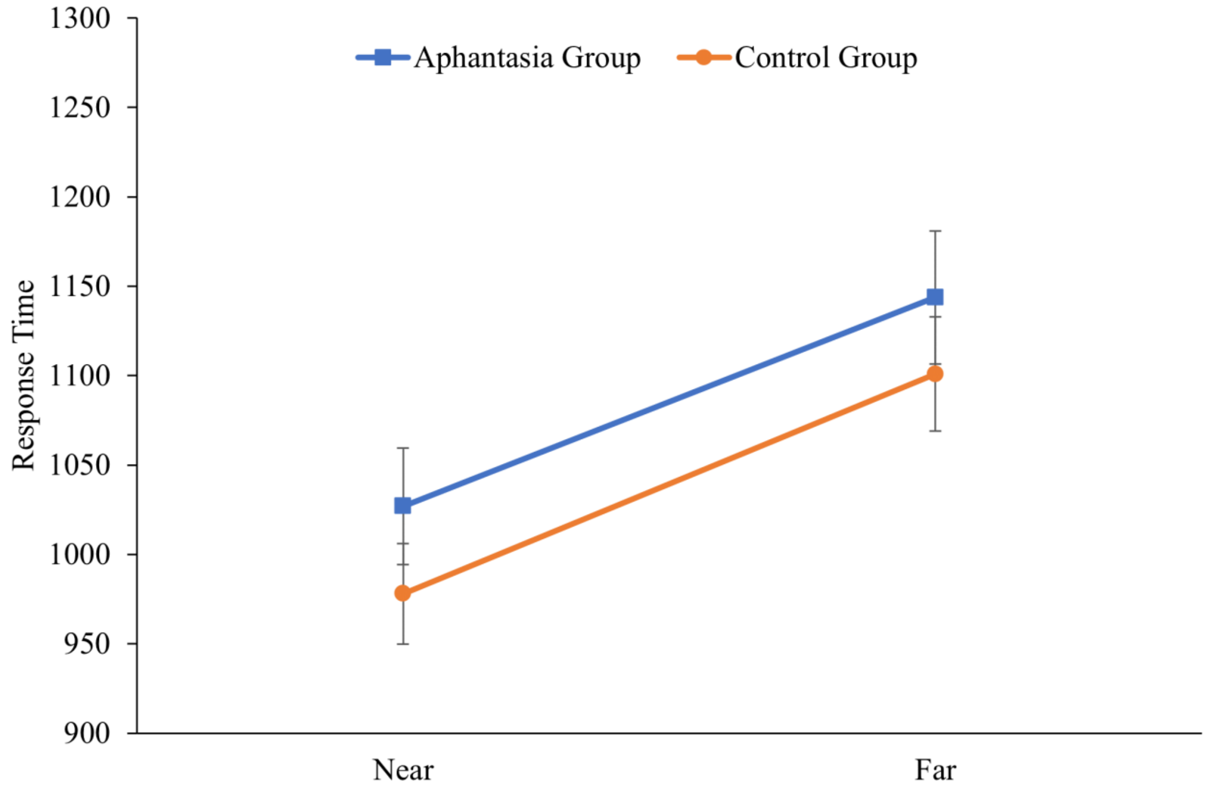


Figure 15. The non-significant group x scanning distance interaction on response times for Study 3. The main effect of group was not significant, but the main effect of scanning distance was significant. Response times appear in terms of ms on the y-axis. Mean response times have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error of the mean.

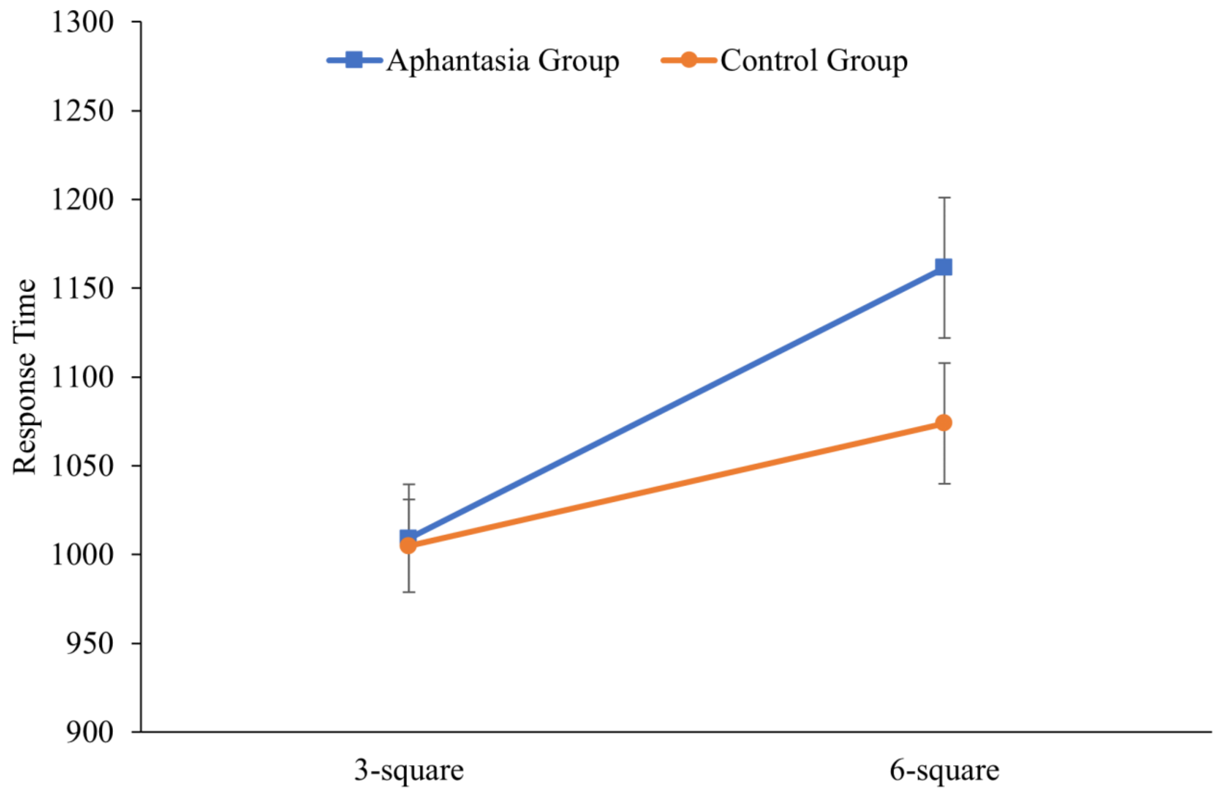


Figure 16. The significant group x squares interaction on response times from Study 3. Neither the main effect of group nor the main effect of squares was significant. Response times appear in terms of ms on the y-axis. Mean response times have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error of the mean.

There was not a significant interaction between scanning distance and squares, with the difference in estimated marginal means of response times relatively equal when group was held constant (F (1, 211) = 2.45, p = .119, η_p^2 = .011, observed power = .345).

Three-Way Interaction

There was a significant three-way interaction between group, scanning distance, and squares (F (1, 211) = 4.31, p = .039, η_p^2 = .020, observed power = .542). As can be seen in Figure 17, the relationship of the steepness of the slopes was complex.

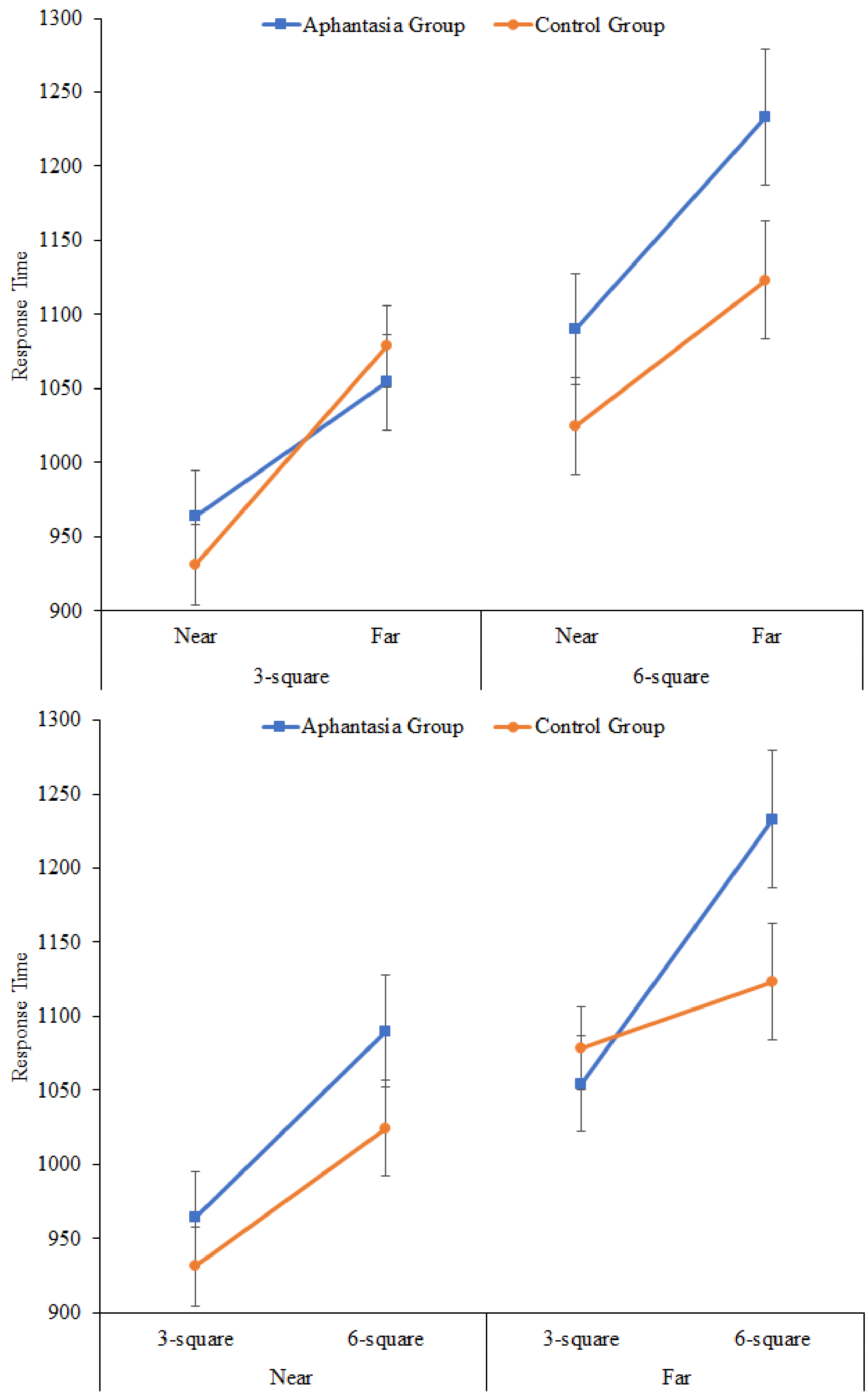


Figure 17. The significant three-way interaction between group, scanning distance, and number of squares on response times for Study 3, depicted two ways. Response times appear in terms of ms on the y-axis. Mean response times have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

The effects of the group x distance interaction were then examined at each level of the squares variable. Age and sex were again controlled for using a 2 x 2 ANCOVA approach (group x distance). When only three-square trials were considered, the interaction between group and distance was just barely outside of significance (*F* (1, 211) = 3.76, *p* = .054, η_p^2 = .018, observed power = .489), with the control group being more affected by the distance manipulation than was the aphantasia group, in terms of how steep the slopes between the distance conditions were for the respective groups. For six-square trials, the interaction between group and distance was not significant, and with a much smaller relative effect size (*F* (1, 211) = 1.09, *p* = .298, η_p^2 = .005, observed power = .180).

The effects of the group x squares interaction were then investigated at each level of the distance variable. Age and sex were again controlled for using a 2 x 2 ANCOVA approach (group x square). When only near trials were considered, the interaction between group and squares was not significant (*F* (1, 211) = .901, *p* = .343, η_p^2 = .004, observed power = .157). However, when only far trials were considered, the interaction between group and squares was highly significant (*F* (1, 211) = 9.77, *p* = .002, η_p^2 = .044, observed power = .875), with the aphantasia group being much more affected by the squares manipulation than was the control group, in terms of how steep the slopes between the squares conditions were for the respective groups.

ANCOVA Investigation into Accuracy

ANCOVA was also used to examine the accuracy of participants. A 2 x 2 x 2 mixed ANCOVA compared group, scanning distance, and number of black squares ((aphantasia x control) x (near x far) x (three-square x six-square)) while entering age and sex as covariates.

Covariates

There was a main effect of age on accuracy (F (1, 211) = 4.34, p = .038, η_p^2 = .020, observed power = .545). Older participants had lower accuracy than did younger participants, but the correlation between age and accuracy did not attain significance (r(213) = -.092, p = .178). The ANCOVA model evaluated age at a computed value of 25.48 for the estimated marginal means reported below. There was not a main effect of sex on accuracy (F (1, 211) = 2.70, p = .102, η_p^2 = .013, observed power = .372). The ANCOVA model evaluated sex at a computed value of 1.60 (where 1 = male and 2 = female).

There was a significant interaction between sex and number of squares (see Figure 18; F (1, 211) = 5.79, p < .001, η_p^2 = .061, observed power = .958). To further investigate the sex x squares interaction, sex was entered into the model as a between-subjects variable instead of as a covariate. The simple effects were then investigated. The difference in accuracy between males and females was not significant when looking at only the 3-square trials (F (1, 210) = .029, p = .866, η_p^2 = .000, observed power = .053) but it *was* significant when looking at only the 6-square trials such that males were more accurate (F (1, 210) = 9.16, p = .003, η_p^2 = .042, observed power = .854). Also, the difference in accuracy between the 3-square and 6-square trials was significant when looking only at males (F (1, 210) = 62.36, p < .001, η_p^2 = .229, observed power = 1) as well as when looking only at females (F (1, 210) = 237.79, p < .001, η_p^2 = .531, observed power = 1).

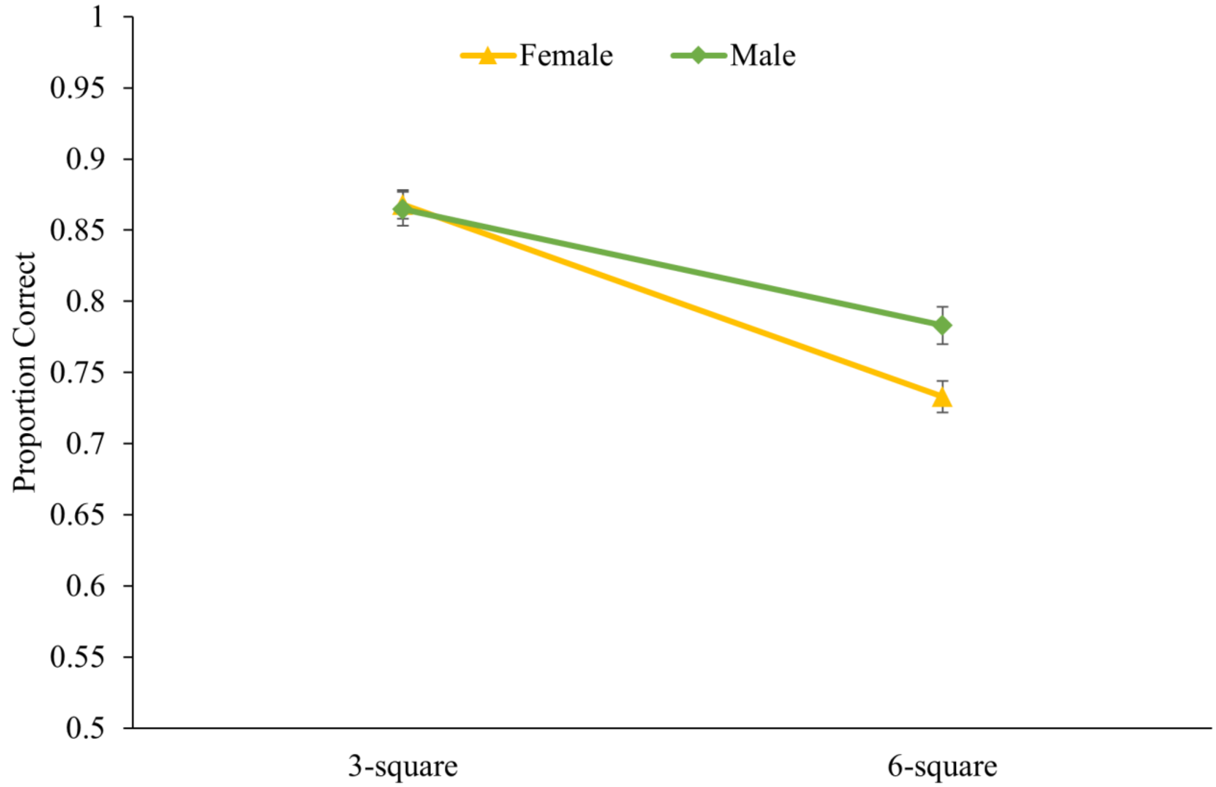


Figure 18. The significant interaction of sex x squares on accuracy for Study 3. The main effect of sex was not significant. Accuracy appears on the y-axis. Mean accuracies have been adjusted for age, and the estimated marginal means are depicted. Error bars are standard error.

There was a significant three-way interaction between age, scanning distance, and squares (See Figure 19), which was unusual, not-at-all hypothesized, and was not interpreted (F (1, 211) = 5.19, p = .024, η_p^2 = .024, observed power = .621). No other interactions with age were significant, and no other interactions with sex were significant (all *ps* > .05).

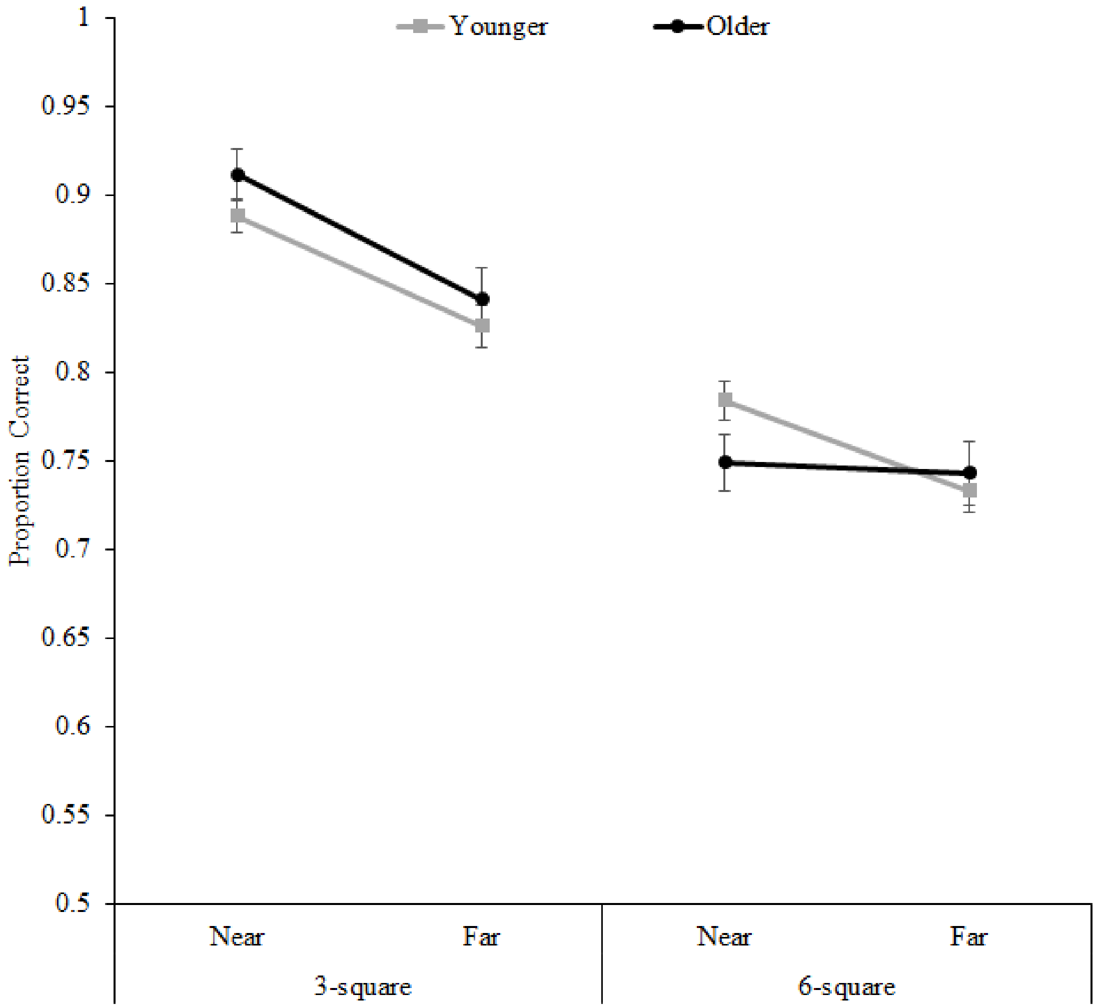


Figure 19. The significant interaction between age, scanning distance, and number of squares on accuracy in Study 3. For the purposes of graphing the interaction, age was split into younger (*n* = 147) and older (*n* = 68) groups at the age of 25.48 based on the ANCOVA-computed value for the covariate. There was a main effect of age when age was entered into the model as a continuous covariate. Accuracy appears on the y-axis. Mean accuracies have been adjusted for sex, and the estimated marginal means are depicted. Error bars are standard error.

Main Effects

There was no main effect of group; estimated marginal means were similar, with participants in the aphantasia group having an estimated marginal mean accuracy of 82.3% (*SE* = 1.3) while participants in the control group had an estimated marginal mean of 79.8% (*SE* = 1.1; F (1, 211) = 1.73, p = .189, η_p^2 = .008, observed power = .259). There was not a main effect of scanning distance, with the near trials having an estimated marginal mean accuracy of 83.5% (*SE*

= 0.7%) and the far trials having an estimated marginal mean accuracy of 78.5% (*SE* = 0.9%; *F* (1, 211) = 3.12, *p* = .079, η_p^2 = .015, observed power = .420). Similarly, a main effect was not detected for the number of squares, with the estimated marginal means of accuracy for the 3-square trials (*M* = 86.6%, *SE* = 0.7%) not differing significantly from the six-square trials (*M* = 75.5%, *SE* = 0.8%, *F* (1, 211) = .120, *p* = .730, η_p^2 = .001, observed power = .064).

Two-Way Interactions

No significant interaction was detected between group and scanning distance (see Figure 20), with the difference in estimated marginal means across scanning distances relatively equal between groups while number of squares was held constant (*F* (1, 211) = .434, *p* = .511, η_p^2 = .002, observed power = .101).

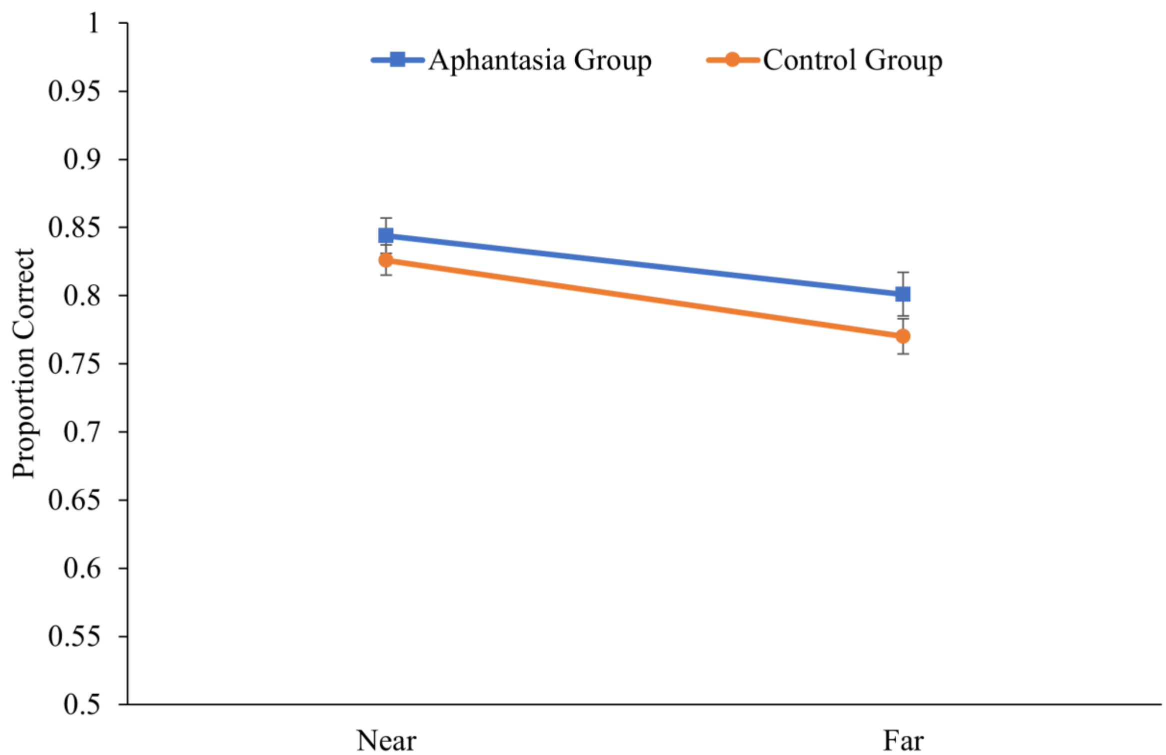


Figure 20. The non-significant group x scanning distance interaction on accuracy for Study 3. Neither the main effect of group nor the main effect of scanning distance were significant. Accuracy appears on the y-axis. Mean accuracies have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

There was a significant interaction between group and squares (see Figure 21) such that the impact of the square manipulation was larger on the aphantasia group's accuracy when scanning distance was held constant—that is, the amount of decrease in accuracy between the three and six square trials was significantly greater for the aphantasia group than for the control group (F (1, 211) = 4.20, p = .042, η_p^2 = .020, observed power = .532). Simple effects were then investigated. When looking only at the 3-square trials, the difference in accuracy between the aphantasia group and the control group was significant such that the aphantasia group was more accurate (F (1, 211) = 4.70, p = .031, η_p^2 = .022, observed power = .578), but there was no significant difference for the 6-square trials (F (1, 211) = .084, p = .772, η_p^2 = .000, observed power = .060). The difference in accuracy between the 3-square trials and the 6-square trials when looking only at the aphantasia group was significant (F (1, 211) = 114.71, p < .001, η_p^2 = .352, observed power = 1) and this was also significant for the control group (F (1, 211) = 79.01, p < .001, η_p^2 = .272, observed power = 1). There was not a significant interaction between scanning distance and squares, with similar estimated marginal means of accuracy when group was held constant (F (1, 211) = 2.27, p = .134, η_p^2 = .011, observed power = .322).

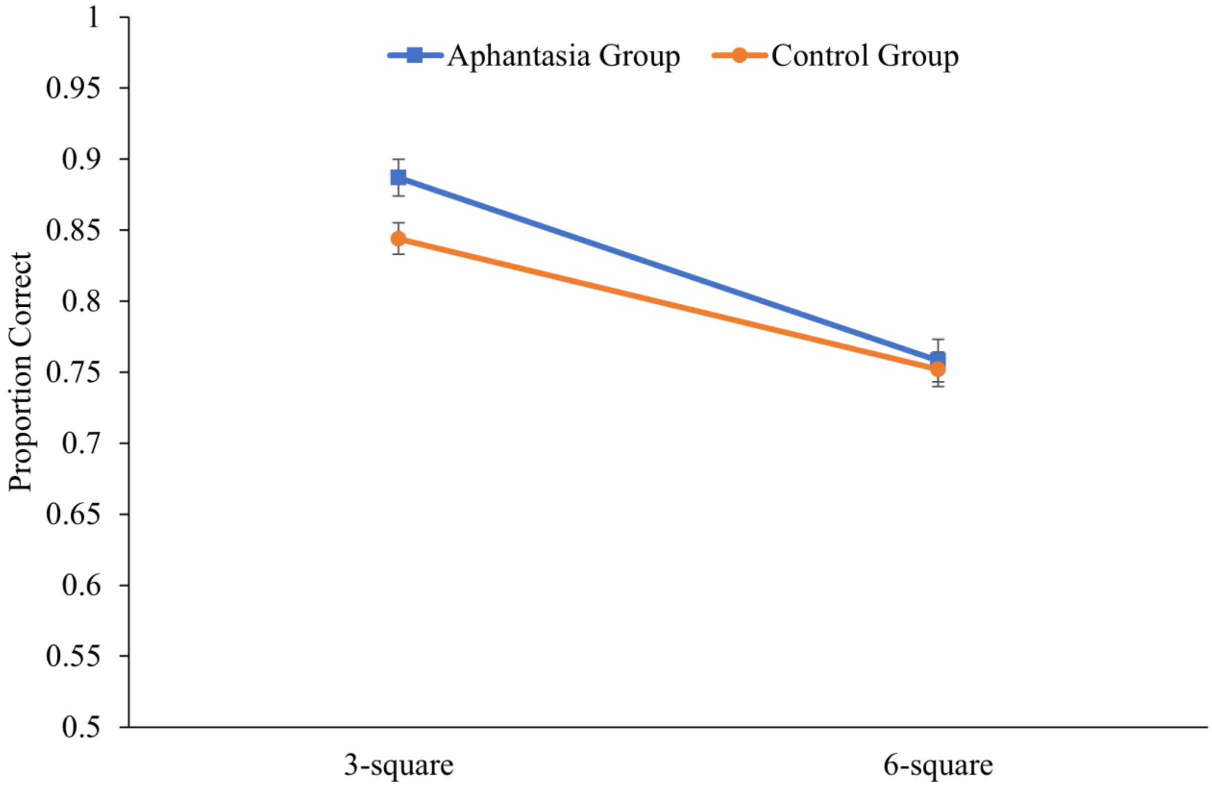


Figure 21. The significant group x squares interaction on accuracy for Study 3. Neither the main effect for group nor the main effect for squares were significant. Accuracy appears on the y-axis. Mean accuracies have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

Three-Way Interaction

There was not a significant three-way interaction between group, distance, and squares for accuracy (see Figure 22; F (1, 211) = .839, p = .361, η_p^2 = .004, observed power = .149).

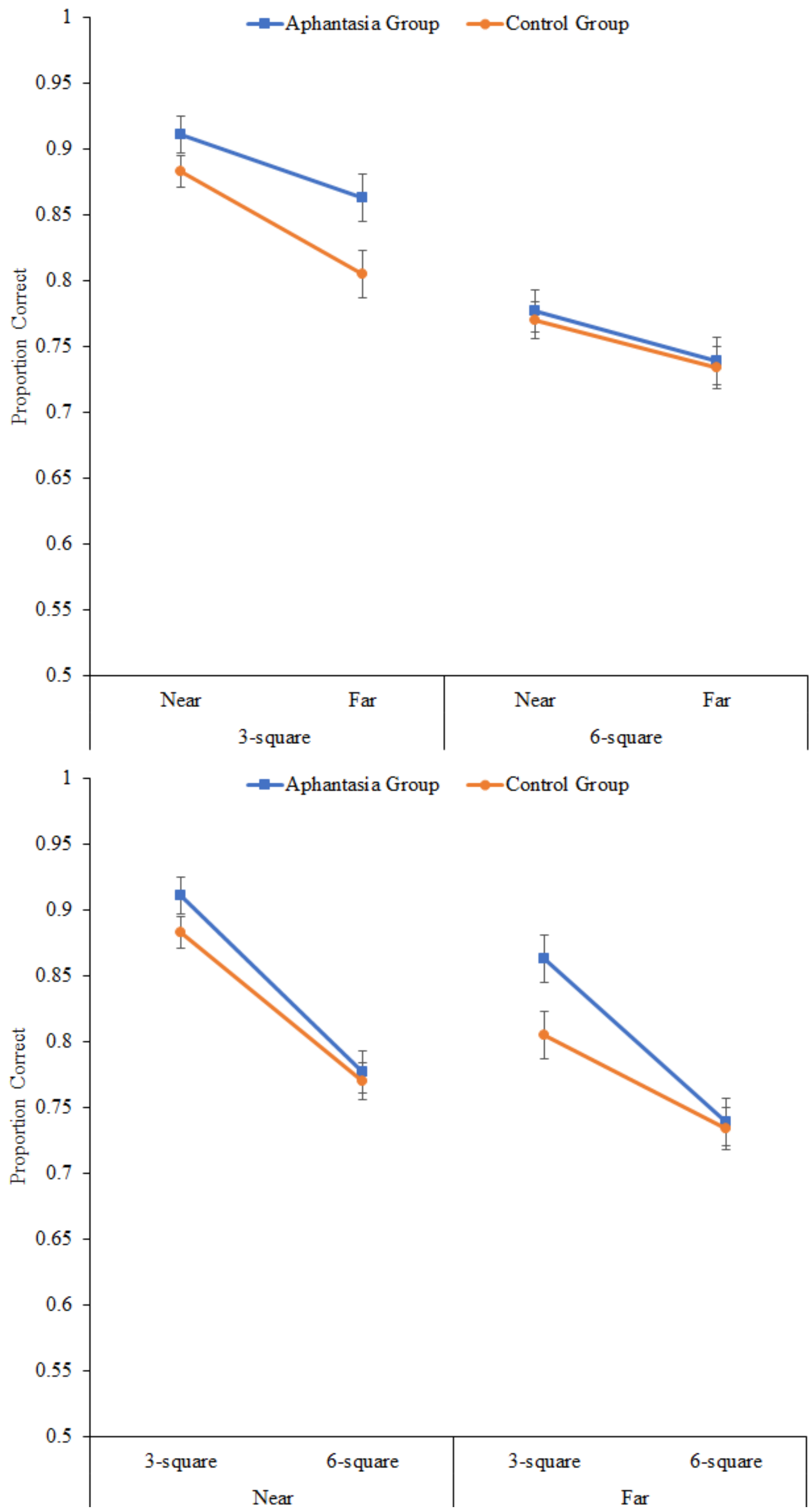


Figure 22. The non-significant three-way interaction between group, scanning distance, and number of squares on accuracy for Study 3, depicted two ways. Accuracy appears on the y-axis. Mean accuracies have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

ANCOVA Investigation into Rate Correct Score for Study 3

Because it could be argued that different participants could prioritize accuracy and response time differently, I also created a measure that captured the variance from both accuracy and response time by combining those two measures in a linear fashion: rate correct score (RCS), which is best conceptualized as "correct responses per second." This additional approach is justified and detailed in the Supplemental Analyses.

ANCOVA was used to examine the RCS of participants. A 2 x 2 x 2 mixed ANCOVA compared group, scanning distance, and number of black squares ((aphantasia x control) x (near x far) x (three-square x six-square) while entering age and sex as covariates.

Covariates

There was a main effect of age on RCS (F (1, 211) = 14.81, p < .001, η_p^2 = .066, observed power = .969). There was also a main effect of sex on RCS (F (1, 211) = 6.92, p = .009, η_p^2 = .032, observed power = .745). No interactions with age were significant, and no interactions with sex were significant (all *ps* > .05).

Main Effects

A main effect was not detected for group: estimated marginal means were not significantly different, with participants in the aphantasia group having an estimated mean rate correct score of .830 (*SE* = .028) compared to the control group's average of .844 (*SE* = .024; F (1, 211) = .106, p = .745, η_p^2 = .001, observed power = .062). There was, however, a main effect of scanning distance, with the near trials having an estimated marginal mean rate correct score of .911 (*SE* = .017) while for the far trials it was .763 (*SE* = .016; F (1, 211) = 28.24, p < .001, η_p^2 = .118, observed power = 1). There was also a main effect for the number of squares, with the estimated marginal means of rate correct scores for the 3-square trials (M = .927, *SE* = .017) significantly higher than the six-square trials (M = .747, SE = .015, F (1, 211) = 12.52, p < .001, η_p^2 = .056, observed power = .941).

Two-Way Interactions

No significant interaction was detected between group and scanning distance (see Figure 23), with the difference in estimated marginal means between scanning distances relatively equal between groups while number of squares was held constant (F (1, 211) = .766, p = .382, η_p^2 = .004, observed power = .140).

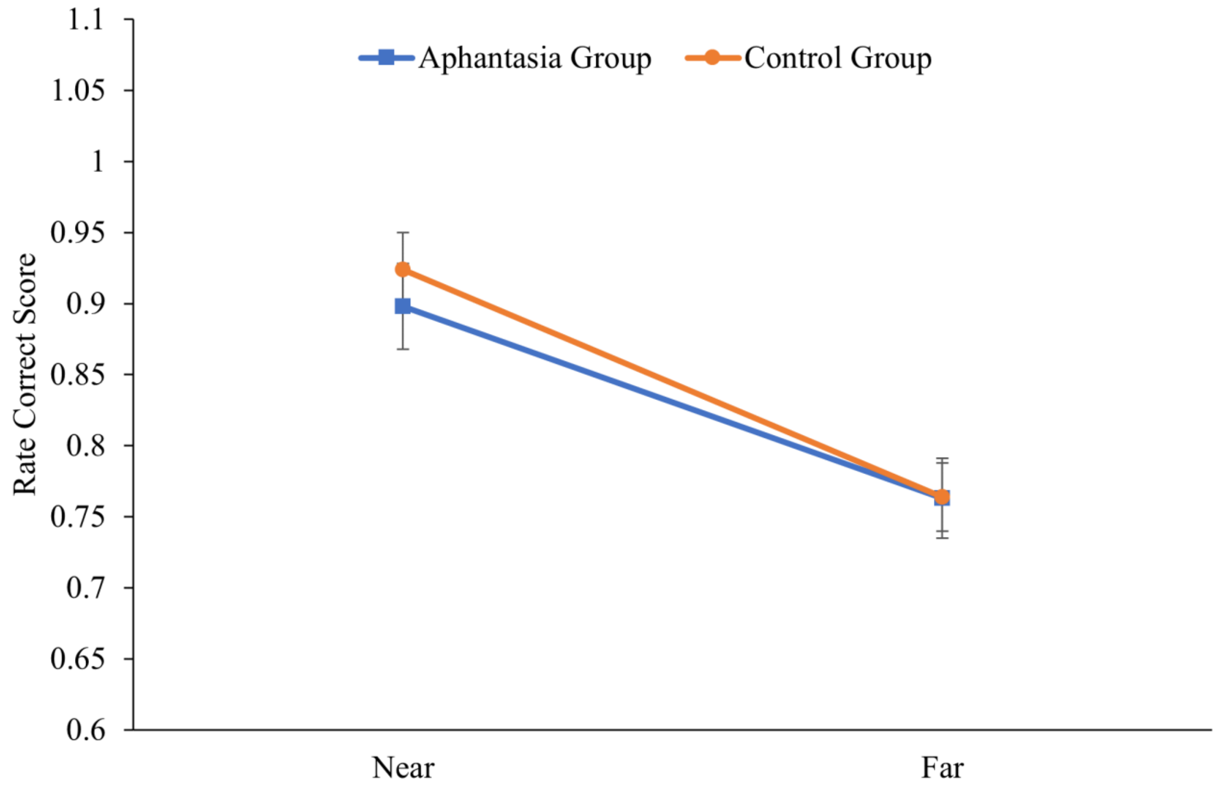


Figure 23. The non-significant group x scanning distance interaction on rate correct scores for Study 3. The main effect of group was not significant, but the main effect of scanning distance was significant. Rate correct score appears in terms of correct responses per second on the y-axis. Means have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error of the mean.

There was a highly significant interaction between group and squares (see Figure 24) such that, when scanning distance was held constant, the impact of the square manipulation on

RCS was larger on the aphantasia group than on the control group (F (1, 211) = 8.45, p = .004, η_p^2 = .038, observed power = .825). The simple effects of group x squares were then calculated using the estimated marginal means, revealing no simple effect between the groups at 3-squares (F (1, 211) = .262, p = .609, η_p^2 = .001, observed power = .080), nor a simple effect between the groups at 6-squares (F (1, 211) = 1.51, p = .220, η_p^2 = .007, observed power = .232). Simple effects were calculated between the levels of "squares" for each group using the estimated marginal means, revealing a significant simple effect of squares at the level of the aphantasia group (F (1, 211) = 163.46, p < .001, η_p^2 = .437, observed power = 1) and a significant simple effect of squares at the level of the control group (F (1, 211) = 96.25, p < .001, η_p^2 = .313, observed power = 1).

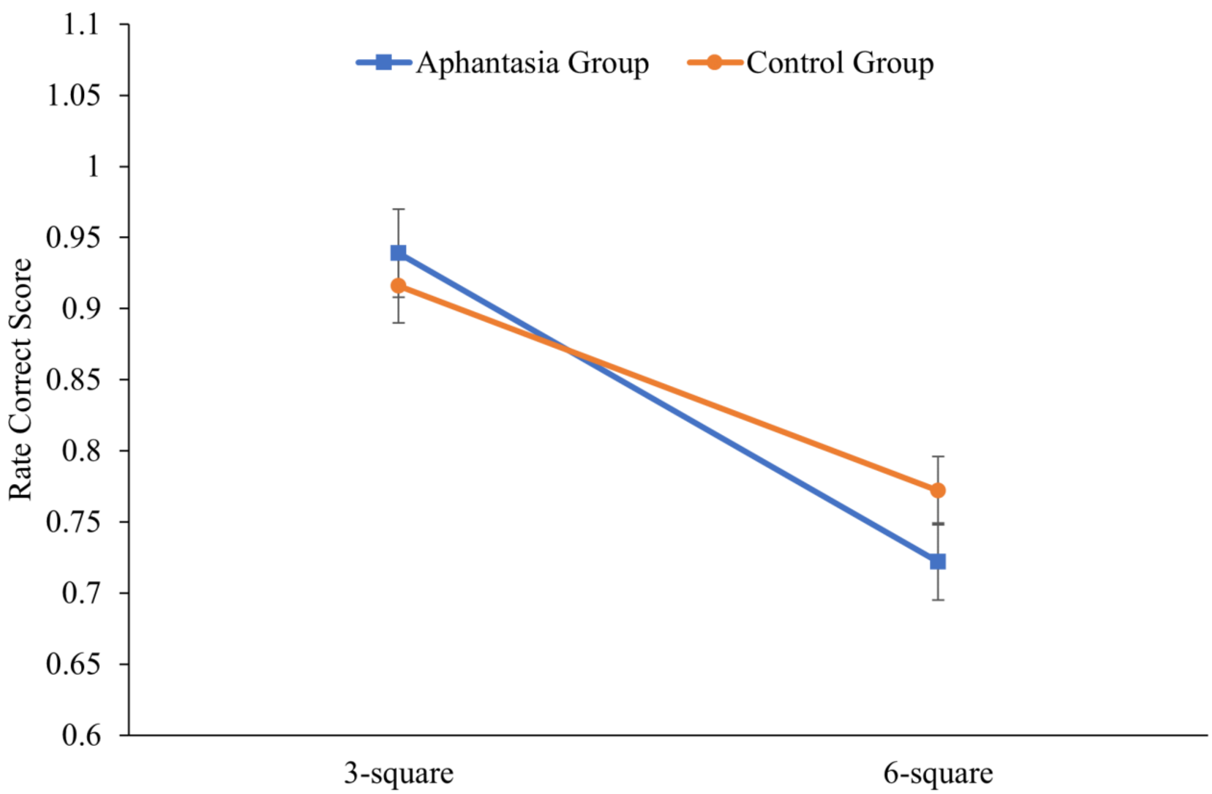


Figure 24. The significant group x squares interaction on rate correct scores from Study 3. The main effect of group was not significant. The main effect of squares was significant. Rate correct score appears as correct responses per second on the y-axis. Means have been adjusted for age and sex, and estimated marginal means are depicted. Error bars are standard error of the mean.

There was not a significant interaction between scanning distance and squares, with the difference in estimated marginal means of rate correct scores relatively equal when group was held constant (*F* (1, 211) = .381, *p* = .538, η_p^2 = .002, observed power = .094).

Three-Way Interaction

There was no three-way interaction detected between group, scanning distance, and squares (see Figure 25; *F* (1, 211) = 1.47, *p* = .226, η_p^2 = .007, observed power = .227).

Self-Reported Imagery Use and VVIQ During Study 3

The aphantasia group reported significantly less use of visual mental imagery during Study 3 (*n* = 95, *M* = 1.21, *SD* = .563) than did the control group (*n* = 120, *M* = 2.99, *SD* = .884) after adjusting the degrees of freedom for a significant Levene's test (*F* = 17.91, *p* < .001; *t*(204.28) = 17.95, *p* < .001, *d* = 2.51). Within the control group, participants who were randomly assigned to watch the imagery video did not significantly differ in terms of their SIU for Study 3 (*n* = 64, *M* = 2.86, *SD* = .906) when compared to the control participants who watched the color perception video (*n* = 56, *M* = 3.14, *SD* = .841; *t*(118) = 1.77, *p* = .080, *d* = .326).

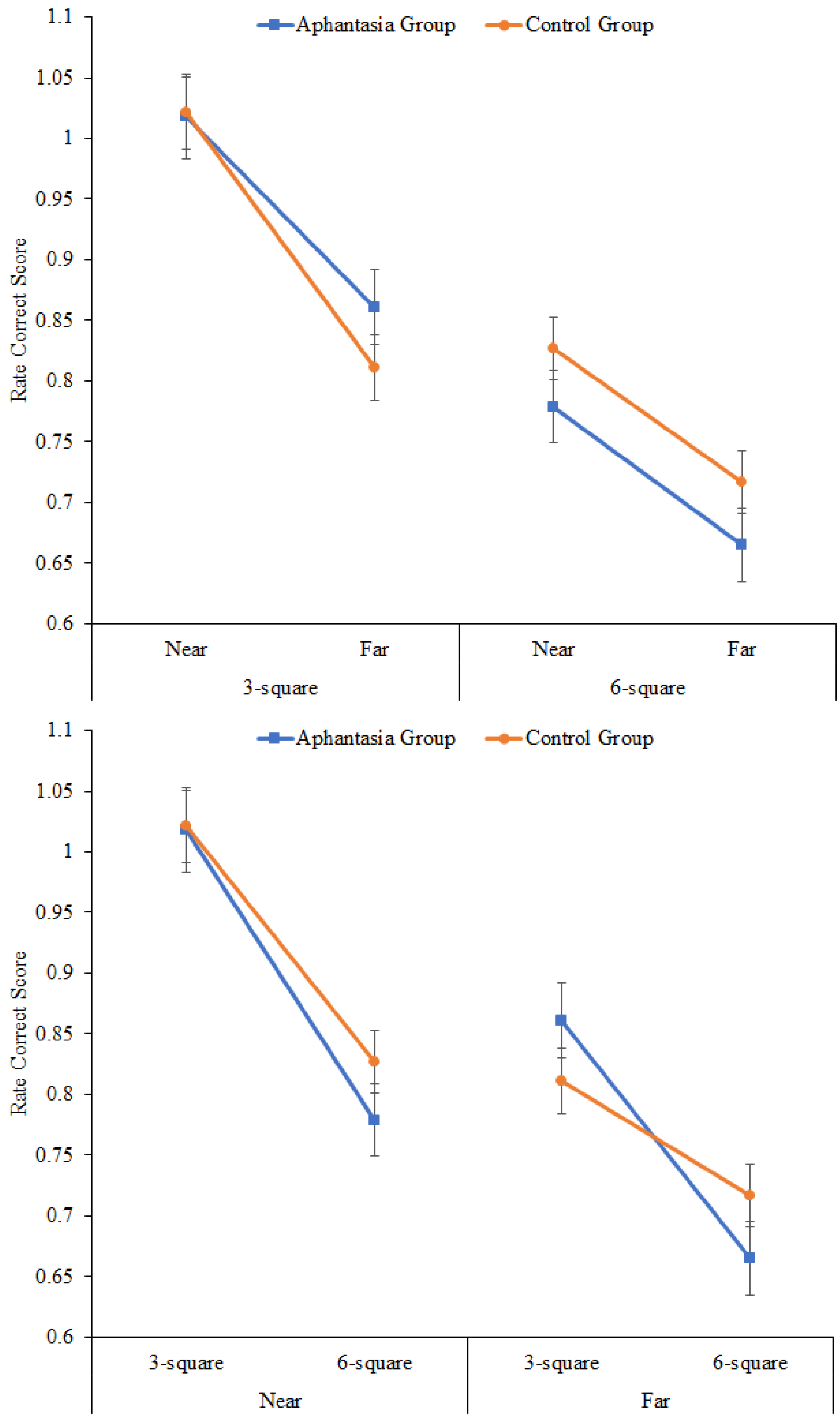


Figure 25. The non-significant three-way interaction between group, scanning distance, and number of squares on rate correct score for Study 3, depicted two ways. Rate correct score appears in terms of correct responses per second on the y-axis. Means have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

SIU on the square donut scanning task used in Study 3 had a strong positive correlation with VVIQ scores ($r(213) = .765$, $p < .001$). This was true even after controlling for age and sex (partial $r(211) = .667$, $p < .001$). When considering only participants with aphantasia, the correlation between SIU and VVIQ scores was strong and positive ($r(93) = .604$, $p < .001$). When considering only participants in the control condition, the correlation is smaller ($r(118) = .206$, $p = .024$). The difference between the correlation of VVIQ and SIU for Study 3 within the aphantasia group and the same correlation within the control group is significant (Fisher's $z = 3.52$, $p < .001$). However, when considering only participants in the control condition who watched the imagery video, that relationship is not significant ($r(62) = .124$, $p = .328$). When only participants in the control condition who watched the color video are considered, the relationship is moderate and positive ($r(54) = .270$, $p = .044$).

Within the control group, SIU for Study 3 was correlated with performance on the square donut scanning task in terms of accuracy ($r(118) = .345$, $p < .001$) but not response times ($r(118) = -.122$, $p = .183$). This picture is more interesting when the control group is split by video condition. When considering control participants who watched the imagery video, overall accuracy on the square donut scanning task was significantly correlated with SIU for Study 3 ($r(62) = .430$, $p < .001$), but this relationship was not significant for control participants who watched the color perception video ($r(54) = .222$, $p = .101$). As depicted in Table 5, the picture becomes even more nuanced when sub-conditions are considered, with the imagery video condition showing stronger correlational relationships than the color perception video condition.

Table 5. *Intercorrelations of Performance Measures with SIU Scores for the Control Group Disaggregated by Video Condition*

Variable	1	2	3	4	5	6	7	8	9	10
1. SIU	—	.124	-.227	-.106	-.240	-.290*	.379†	.159	.422†	.389†
2. VVIQ	.270*	—	.053	.182	.125	.107	-.142	-.149	-.044	.017
3. 3NearRT	-.036	-.221	—	.789†	.748†	.686†	-.195	.023	-.335†	-.294*
4. 3FarRT	-.116	-.223	.827†	—	.711†	.787†	-.163	.098	-.209	-.239
5. 6NearRT	-.086	-.128	.814†	.807†	—	.764†	.091	.189	-.049	-.067
6. 6FarRT	.114	-.141	.778†	.857†	.812†	—	-.004	.194	-.076	-.150
7. 3NearACC	.297*	.274*	-.472†	-.301*	-.291*	-.173	—	.460†	.561†	.450†
8. 3FarACC	.057	.220	-.470†	-.388†	-.237	-.186	.554†	—	.412†	.438†
9. 6NearACC	.274*	.136	-.412†	-.348†	-.303	-.131	.569†	.669†	—	.466†
10. 6FarACC	.092	.198	-.298*	-.294*	-.106	-.228	.336*	.580†	.404†	—

Note. The results for the control participants who watched the imagery video (n = 64) are shown above the diagonal. The results for the control participants who watched the color perception video (n = 56) are shown below the diagonal. SIU = self-reported imagery use for Study 3. VVIQ = Vividness of Visual Imagery Questionnaire. 3NearRT = Response times on the 3-square near trials. 3FarRT = Response times on the 3-square far trials. 6NearRT = Response times on the 6-square near trials. 6FarRT = Response times on the 6-square far trials. 3NearACC = Accuracy on the 3-square near trials. 3FarACC = Accuracy on the 3-square far trials. 6NearACC = Accuracy on the 6-square near trials. 6FarACC = Accuracy on the 6-square far trials.
*$p < .05$. †$p < .01$

When considering all participants, there was a significant negative correlation between VVIQ scores and response times on the square donut task such that participants with lower VVIQ scores had longer average response times ($r(213) = -.296$, $p < .001$). However, this correlation disappears when age and sex are controlled for (partial $r(211) = -.054$, $p = .430$). There was not a significant correlation between VVIQ score and accuracy on the square donut task: neither the zero-order correlation ($r(213) = -.004$, $p = .955$) nor the partial correlation controlled for age and sex (partial *r*(211) = -.066, *p* = .337) were notable. When considering RCS, there was a significant correlation with VVIQ scores before controlling for age and sex (*r*(213) = .202, *p* = .003). However, the correlation drops out of significance once age and sex are controlled for (partial *r*(211) = .023, *p* = .742) Thus, there is no correlational evidence that overall performance on this task and VVIQ score are related.

Dream Imagery

Results of these ANCOVAs for response time and accuracy were also analyzed after splitting the aphantasia group into two groups based on the frequency of their dreams with imagery content (see Study 1). It was possible that some effects are only detectable for participants with aphantasia who do not experience mental imagery in dreams because their aphantasia may have a different neural basis in the brain (i.e., the lack of mental imagery generation capabilities in the brain may be more severe). However, these additional analyses did not reveal any additional findings worth discussing in the main body of this study, and so have been relegated to the Supplemental Analyses.

Study 3 Discussion

Aphantasia Theories

The results of Study 3 are encouraging for the perceptual/conceptual theory of aphantasia. That is, it seems that the members of the aphantasia group were completing the square donut scanning task while favoring the use of propositional representations while the control group was more likely to use depictive representations—and not just because participants with aphantasia had significantly lower self-reported imagery use (SIU) scores than did control participants. Let’s walk through why the performance measures on the square donut scanning task seem to agree with this perceptual/conceptual interpretation.

Crucial to supporting this theory was the three-way interaction between group, scanning distance, and number of squares when average response time was investigated (see Figure 17). Additional evidence for this theory comes from the significant group x distance interaction on response times (see Figure 16), the significant group x squares interaction on accuracy (see Figure 21), and the general shape of the rate correct score graphs (see Figures 23, 24, and 25).

First, let's look at the upper panel and the left side of Figure 17, which shows three-square trials only. Notice how the average response times of the participants with aphantasia are less affected by the scanning distance manipulation, as indicated by the shallower slope of their line relative to the slope of the control group's line. This is the pattern of results that we would expect if the aphantasia group was more often using propositional representations while the control group participants were more often using depictive representations as the theory suggests—the scanning distance manipulation is less disruptive to a propositional representation strategy (i.e., a strategy that does not include metric space as part of the encoded image in the brain) than it is to a depictive representation strategy (i.e., a strategy that represents metric space as part of the encoded image in the brain). There is more disruption because the depictive representation (mental image) becomes more complex as more space between the arrow and black square is added. However, the propositional representation does not become more complex when more space is added between the arrow and black square (e.g., a person could attempt to encode a 3-square trial propositionally as "TOP-LEFTMOST, LEFT-BOTTOMMOST, RIGHT-UPPERMIDDLE," which does not represent distance). Thus, when there were three black squares, the control group was more affected by the distance manipulation than was the aphantasia group, possibly because of their use of mental imagery during the task relative to less mental imagery used by the aphantasia group. This pattern fits with perceptual/conceptual theory.

Next, let’s look at the right side of the upper panel of Figure 17, which shows six-square trials only. Notice how the average response times for the aphantasia group are now universally slower than the control group (see also Figure 16). However, the control group performed about as fast on these 6-square trials as they, the control group, did on the 3-square trials. That is, moving from 3 squares to 6 squares was massively disruptive to the aphantasia group, but not very disruptive to the control group. This is the pattern of results that we would expect if, for the aphantasia group, propositional strategies had now become too difficult and inconsistent to be used when there were six squares instead of three, and depictive strategies were less available to such participants. This also fits with perceptual/conceptual theory.

Next, let’s look at the lower panel of Figure 17. These are the same data as in the upper panel, but displayed differently in order to show the slopes resulting from the squares manipulation rather than from the distance manipulation. Notice that the steepest slopes in the bottom panel belong to the aphantasia group. That is, adding more black squares was very disruptive to the aphantasia group and caused their response times to skyrocket relative to the control group. In contrast, the control group’s slopes are much flatter, which is the pattern that we would expect if the control participants used a depictive strategy for both the 3-square and 6-square trials—adding more black squares is less disruptive to a depiction representation strategy than it is to a propositional representation strategy.

As would be predicted by perceptual/conceptual theory, when both accuracy and response time were accounted for in a single dependent measure, RCS, the general shape of interaction slopes were preserved (see Figures 23, 24, and 25). This can especially be seen on the left-hand side of Figure 25, where the line is steeper for the control group for the “distance” manipulation, while the line is steeper for the aphantasia group for the “squares” manipulation. The general shape of Figure 23 and the significant interaction of Figure 24 also support this interpretation. Not all of those RCS interactions attained significance, and so they are not as strong of evidence as the significant response time interaction graphs, but their general shape fits the perceptual/conceptual theory exactly as would be predicted.

Additionally, the aphantasia group was significantly more accurate than the control group on the 3-square trials, which is where a propositional representation strategy would be expected to be most effective, but there was no significant difference in accuracy between groups for the 6-square trials, where one would not expect a propositional strategy to be very effective due to a higher number of squares that may overwhelm, for example, a person’s verbal working memory that may be used during a propositional strategy (see Figure 21).

These patterns are evidence that the cognitive strategies of people with aphantasia are categorically different when it comes to their performance on a task that has been previously shown to make use of mental imagery, the square donut scanning task.

Also, accuracy on the square donut scanning task correlated with self-reported imagery use (SIU) within the control group, supporting the idea that the use of mental imagery was useful during the square donut scanning task when participants had a choice between depictive and propositional strategies.

In contrast, we must ask: is the implicit/explicit theory supported by these findings? One might try to argue that the significant “main effect” of scanning distance on response times, and the simple effect of scanning distance on response times when considering only participants in the aphantasia group, are evidence for depictive representations being used by all participants to complete the scanning task, as well as the *lack* of a significant group x scanning distance interaction for response times (see Figure 15). However, when the more complex model is evaluated, it turns out that scanning distance is part of a complicated three-way interaction as discussed above. Therefore, it isn't that participants with aphantasia showed the same pattern of response times in reaction to scanning distance relative to the control group, instead, their pattern of response times varied based on scanning distance and the number of black squares. Also, there could be alternate ways of explaining the impact of scanning distance that account for the increase in response times for the aphantasia group in Figure 15. As an example, the effect of scanning distance may have showed up for both the aphantasia and control groups because it was more difficult to determine which square was being pointed at whenever the arrow was far away during a "far" trial, regardless of whether a mental image was used during the trial. Therefore, the implicit/explicit theory is not convincingly supported by these results.

This study has demonstrated that people with aphantasia do indeed perform differently on a "classic" task of mental imagery—mental scanning. But questions remain. What exactly do people with aphantasia fail to mentally represent? For instance, the square donut scanning task is partially spatial and partially visual in nature. There is some evidence that spatial imagery and visual imagery dissociate in people with aphantasia (e.g., Keogh & Pearson, 2018; Kozhevnikov et al., 2005; see also Blazhenkova & Pechenkova, 2019). To investigate differences in the type of stimuli used for cognitive tasks (e.g., visual vs. spatial), Study 4 introduces several trial types that attempt to manipulate one specific characteristic of an image at a time from a set of four: color, form, orientation, or position.

Impact of Imagery Video Manipulation

Even though the video manipulation did not significantly affect SIU scores for Study 3, several measures of performance in Study 3 correlated significantly with SIU scores for control participants who watched the imagery video, more so than for control participants who watched the irrelevant color perception video (see Table 5). This improved calibration between subjective reports and performance measurements provides evidence that watching the mental imagery video helped the control participants better understand mental imagery and whether or not they were making use of it during the square donut scanning task. In parallel to Study 2, VVIQ scores were not useful in detecting such differences within the control group. Thus, SIU scores seem to be somewhat effective subjective measurements, and VVIQ scores are not as useful as prior literature has concluded because they seem to be so detached from the actual use and usefulness (e.g., promoting accuracy) of imagery during actual performance on tasks.

CHAPTER 5. STUDY FOUR

According to categorial/coordinate models of object recognition, one possible way that a person can remember visual relationships is by encoding categorical relationships between objects or parts of objects (e.g., Brooks & Cooper, 2006). Importantly, this categorical encoding can be accomplished with propositional representations (i.e., by remembering words describing the visual scene) and does not require depictive representations. In contrast is the possibility of encoding using depictive (coordinate) representations, by encoding points in space using metric information relative to, for example, a spatial grid. It has been argued that people can usually make use of both of these types of processing—similar accounts for multiple processing systems have also been proposed (e.g., categorical vs. coordinate spatial relations encoding in Kosslyn, 1994; dual coding theory in Paivio, 1971). These theories agree that there are multiple systems for processing incoming information. Perhaps, then, a difference in people with aphantasia could be that one of these systems—the coordinate system—is underutilized or unutilized when it comes to a) visual encoding and b) generating visual imagery for the purpose of recall.

People with aphantasia claim not to experience depictive mental imagery, so perhaps they do not *encode* visual scenes in a depictive way either. Such an encoding scheme would mean that they are not making use of the depictive/coordinate system when memorizing a visual scene. If people with aphantasia are not encoding visual information as depictive representations via coordinate processing, then perhaps they are only encoding visually-perceived images using a propositional (categorical) approach. For example, a person with aphantasia may favor remembering that their phone is on top of a cabinet by using a categorical/propositional approach of encoding the relationship ON(CABINET, PHONE), as opposed to encoding the specific coordinate/depictive relationship between the position of the phone and the cabinet such as by encoding a picture of *where* the phone is located. If this is true, it would support the perceptual/conceptual theory of aphantasia, and be evidence against the implicit/explicit theory. However, it is also possible that people with aphantasia are encoding visual images using a depictive (coordinate) approach, but that they are unable to generate an *explicit* mental image from that memory despite having an *implicit* depiction—which would support the implicit/explicit theory of aphantasia, and be evidence against the perceptual/conceptual theory.

In order to test these competing theories, participants were shown colored line drawings of shapes and angled lines. Then, they were shown modified (i.e., changed) versions of those stimuli and were asked to report *what changed* about the stimuli since the previous presentation. There were four types of changes: color, form, orientation, and position. Their task was to identify which of those four possible changes occurred for each stimulus. The task is not a change *detection* task because participants knew that a change occurred on each presentation—instead, this is a change *identification* task. Examples of these types of changes are provided in Figures 26–29.

Some of the changes were *categorical* in nature, such that a nominal boundary was crossed in a way that mismatched with propositional encoding (e.g., a rectangle becoming an oval). The rest of the changes were *coordinate* in nature, such that the change was metric without crossing a nominal boundary (e.g., a rectangle could become a longer rectangle or a shorter rectangle, but it remained a rectangle). The idea is that the coordinate changes were more likely to force a person to use depictive mental imagery in order to determine what the change was (e.g., they cannot simply remember the word "rectangle," they must retrieve an image of the length of the rectangle).

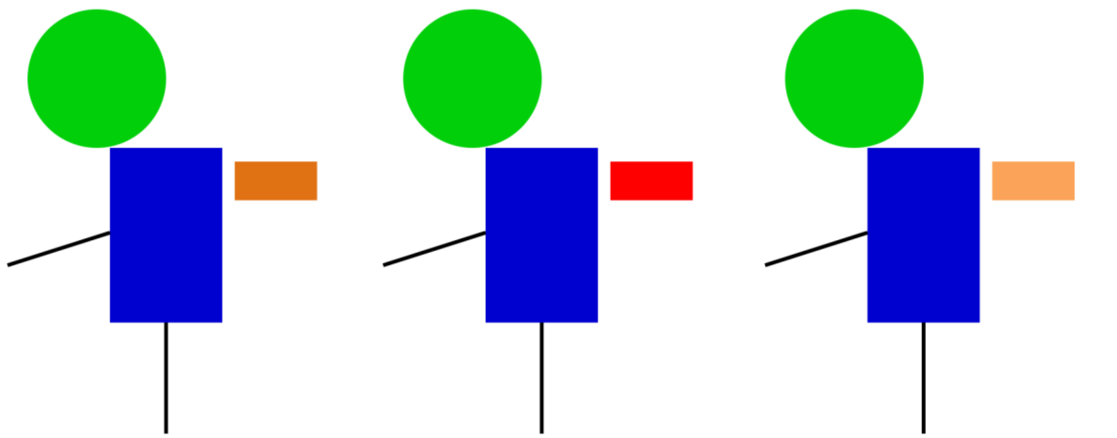


Figure 26. An example of a color trial. The figure on the left is the originally presented stimulus, where the smaller rectangle is a dark orange color. The figure in the middle is the “categorically” changed stimulus, where the smaller rectangle is a red color. The figure on the right is the “coordinately” changed stimulus, where the smaller rectangle is a light orange color. Color trial stimuli were displayed in color during the study. This trial did not appear as a criterion trial in the study.

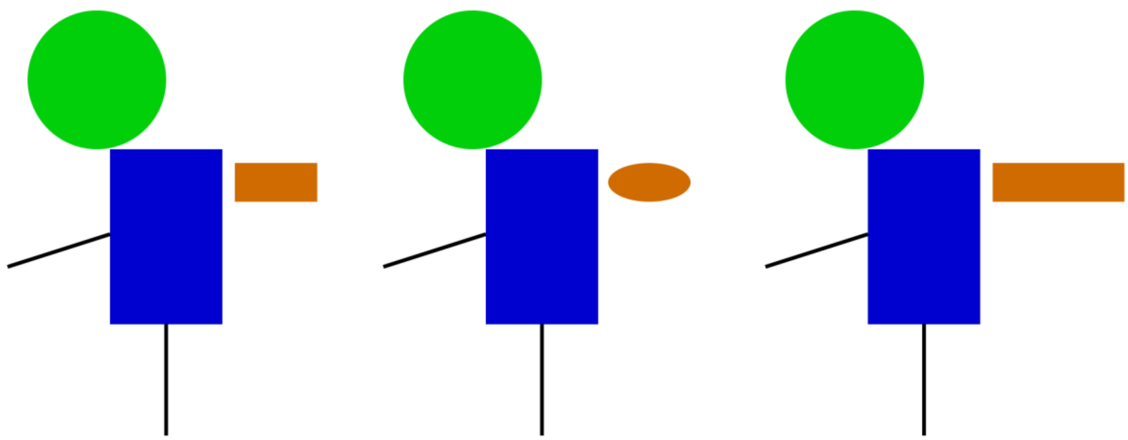


Figure 27. An example of a form trial. The figure on the left is the originally presented stimulus. The figure in the middle is the “categorically” changed stimulus. The figure on the right is the “coordinately” changed stimulus. This trial did not appear as a criterion trial in the study.

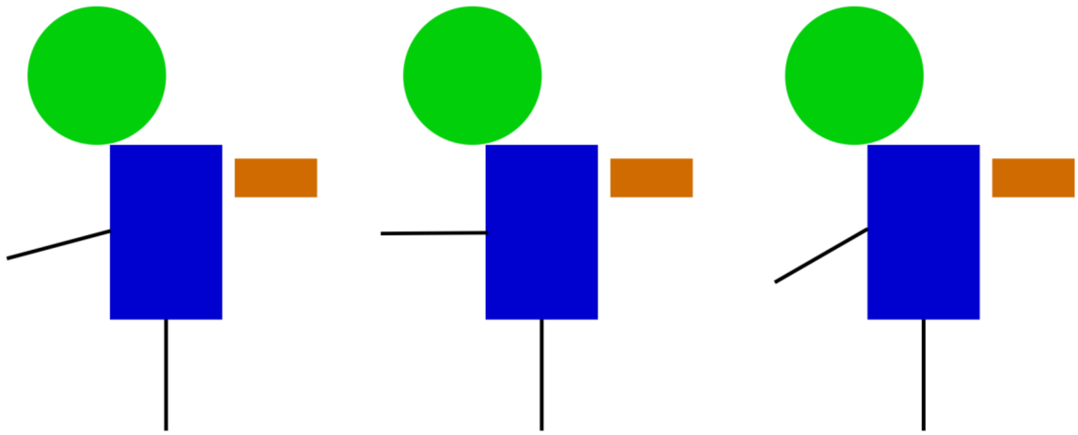


Figure 28. An example of an orientation trial. The figure on the left is the originally presented stimulus. The figure in the middle is the “categorically” changed stimulus. The figure on the right is the “coordinately” changed stimulus. This trial did not appear as a criterion trial in the study.

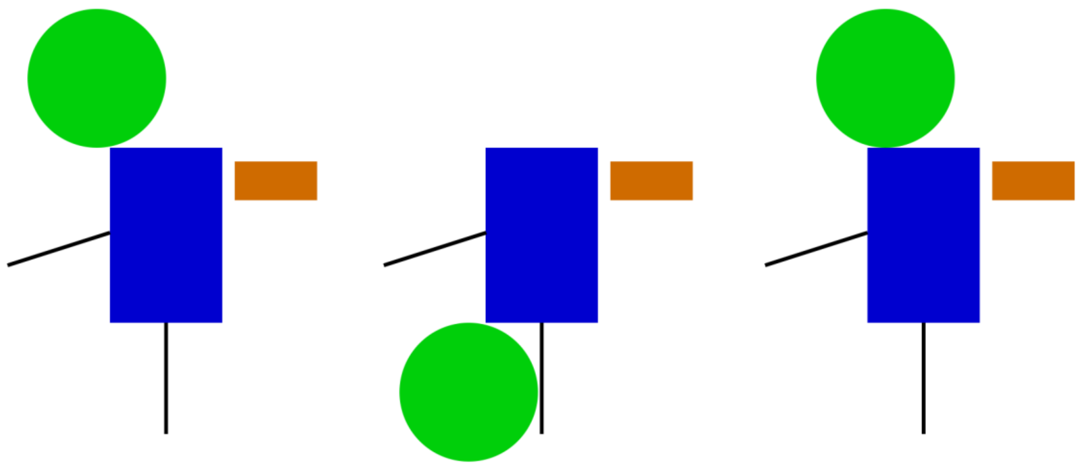


Figure 29. An example of a position trial. The figure on the left is the originally presented stimulus. The figure in the middle is the “categorically” changed stimulus. The figure on the right is the “coordinately” changed stimulus. This trial did not appear as a criterion trial in the study.

In order to better describe the logic of this experiment and justify the use of a recall-based change identification task over a recognition-based change detection task, I will describe what may be happening in a person's brain. First, when encoding a stimulus (e.g., the far-left stimulus from Figure 26), all people in this study will view the stimulus, sending neural information from their retina, to the lateral geniculate nucleus, to V1, and from there to the extrastriate regions, the posterior parietal lobe, and the inferior temporal lobe. The *pattern* of neural activity in those latter brain regions will form part of their memory of that stimulus. Then, in a recognition task such as a change *detection* task, what would happen is they would view a test trial of either the same stimulus as before or a changed version of the stimulus, and reply with a response of either "old" (seen previously) or "new" (not previously seen). Importantly, during the test trial of the stimulus in a recognition task, neural information can take the same path as during the encoding process, which will either trigger a feeling of familiarity or *not* trigger a feeling of familiarity. This feeling of familiarity due to the pattern of activation in those "latter" brain regions such as the inferior temporal lobe may be enough at times to answer the old/new question quickly and accurately. Therefore, in theory, during a recognition task a person with aphantasia can use the feeling of familiarity even without generating a mental image in order to accomplish the task. As such, a recognition task is not a strong way to force someone to generate a mental image, unlike a recall task in which no identical visual information is presented. My task in Study 4 does present visual information during the testing phase, and so resembles a recognition task—however, my task differs from a recognition task because they are not deciding whether the stimulus is old or new (i.e., the stimulus will always be 'new' because something will have been changed about it). Instead, they will always have the aforementioned feeling of familiarity for each stimulus because the majority of the image will match their encoding scheme (regardless of whether the encoding was propositional or depictive). Therefore, the feeling of familiarity will not be a useful diagnostic cue for completing the task, which should encourage the generation of mental images.

It was predicted that manipulations to the stimuli in terms of color, form, orientation, and position would differentially affect participants making use of propositional representations versus participants making use of depictive representations. That is, categorically changed stimuli should be relatively easy to identify propositionally (e.g., rectangle versus oval, red versus orange, right angle versus acute angle, on top of versus below) while coordinately changed stimuli should be more likely to require depictive representations in order to quickly and accurately identify the change (e.g., short rectangle versus longer rectangle, medium orange versus darker orange, narrow acute angle versus wider acute angle, above and greatly to the left versus above and slightly to the left). That is, because the coordinate changes do not cross a categorical boundary, if recall by proposition is attempted during a coordinate trial, quickly and accurately identifying the change to the stimulus will be more difficult than if that same strategy was applied to a categorical trial. Therefore, if the perceptual/conceptual theory is correct and people with aphantasia are using *only* propositional representations when recalling, then people with aphantasia should have a deficit in detecting changes to coordinately-changed stimuli relative to the control group. Additionally, according to the perceptual/conceptual theory, participants with aphantasia should show no deficit for categorically-changed stimuli, and may even have an advantage due to saving processing resources when they do *not* generate a mental image, unlike the control group who presumably will generate a mental image even for the categorically-changed stimuli—although, the control group also has access to propositional representations and so there may be no difference between the groups in terms of performance on categorically-changed stimuli. Therefore, the strongest evidence for the perceptual/conceptual model of aphantasia would be a significant interaction effect, where participants with aphantasia perform worse on coordinately-changed stimuli relative to categorically-changed stimuli, all relative to the control group (i.e., a slope difference; an interaction effect).

However, if the implicit/explicit theory of aphantasia is correct, then people with aphantasia should perform just as well as the control group at this task because even though they do not have conscious access to a depiction of the mental image, they can still use cognitive depictions to solve the question of what features changed in the displayed stimuli regardless of whether the changes are categorical or coordinate.

As an example, according to perceptual/conceptual theory, the average person with aphantasia that encoded the far-left stimulus from Figure 27, and then was tested on the center version of the categorically changed stimulus, should be as quick and as accurate as the average control participant would be in identifying that the correct answer is “form change.” Or, the average person with aphantasia *could* be faster and more accurate than the average control participant at identifying the form change in the middle stimulus due to saving the processing costs of generating a depictive mental image. However, also according to perceptual/conceptual theory, the average person with aphantasia that encoded the far-left stimulus from Figure 27 that was then tested on the far-right version of the coordinately changed stimulus should be slower and less accurate than the average control when identifying the correct answer of “form change.” In contrast, implicit/explicit theory makes no such predictions of differences in performance between the aphantasia group and the control group.

This study is not the first that has examined the effect of differences in reported imagery vividness on the memory for visual stimuli (see also McKelvie, 1995, p. 71). For example, Gur and Hilgard (1975) performed a median split on a group of 20 participants along VVIQ scores and found that when participants were asked to point at guaranteed differences between successively presented illustrated cards in a change identification task, the “good” visual imagers had significantly faster response times compared to the “poor” visual imagers. However, this effect failed to appear in Berger and Gaunitz (1977) when a larger sample of 48 was used, a double-blind procedure was added, and a recognition-based forced choice change detection task between “same” or “changed” was used instead of a task in which a guaranteed change on the card was identified. That last change was implemented in order to avoid measuring response time on a stopwatch, which could be biased by the experimenter—a problem which is avoided in the present study due to digital timing. It is possible that—for the above reasons discussing the difference between recognition and recall in aphantasia—Berger and Gaunitz were unable to replicate Gur and Hilgard because the recognition task did not *require* mental image generation. Importantly, the median split design used for both of those experiments almost certainly did not result in a group of participants that would be classified today as having aphantasia. VVIQ scores range from 16 to 80, and people reporting aphantasia generally score, on average, in the teen range (e.g., 19.0 in this series of studies; 17.9 in Dawes et al., 2020; 19.2 in Keogh & Pearson, 2018; 17.1 in Zeman et al., 2020). In Gur and Hilgard, the average VVIQ score for the “good” imagers was 70.6 (*SD* = 3.5), while the average for the “poor” imagers was 50.3 (*SD* = 12.5). In Berger and Gaunitz, the average VVIQ score for the “good” imagers was 56.5, while the average for the “poor” imagers was 41.8 (no *SD* was reported). Therefore, my design used a different group of interest: people who report no subjective experience of voluntary visual mental imagery.

Method

Participants

The participants were carried over from Study 1. As mentioned, one colorblind person was excluded from the Study 4 analyses because the trials depended on color discrimination. This reduced the sample to 94 participants in the aphantasia group, and 120 in the control group. In Gur and Hilgard's (1975) study, they detected a difference between groups with only 15 sequential trials and 10 participants in each group. Both my sample size and number of trials greatly exceeded their numbers, although the study design was substantially different.

Materials

Previous experiments that have used categorical/coordinate changes in visual stimuli have used several types of categorical/coordinate changes, including position, orientation, size, and form (e.g., Arnold, 2018; Casner, 2006). However, position, orientation, and size are all mostly *spatial* changes while form is a mostly *visual* change. In order to more evenly distribute the changes made to the visual stimuli, size was not used. Instead, color was used as another visual change. Therefore, this experiment was a 2 (aphantasia x control) x 2 (categorical x coordinate) x 4 (color x form x orientation x position) design. These four trial types were intermixed such that participants did not know which features of the image were important for later recall. The materials for this experiment were prepared using image editing software that allowed precise manipulation of color, form, orientation, and position.

Each stimulus was made up of three shapes and two lines arranged on a white 1000 by 1000 pixel square. Shapes could be rectangles, ovals, or triangles. Lines were always colored black. Each line was a minimum of 100 pixels long (>10% of the image width) and was always 4 pixels wide (4% of image width). Shapes and lines could be connected or disconnected from one another (i.e., touching), but in most cases were disconnected.

For color trials, one of the shapes changed in color. For categorical color trials, this color was a disparate color in name from the original color (e.g., red to green, blue to red, brown to pink, etc.) such that a visual inspection revealed an obvious difference in color. For coordinate color trials, the change was more subtle such that the color was lightened or darked into a different tint/shade of the original color (e.g., light green to dark green, dark blue to light blue, etc.). An equal number of coordinate trials featured lightening and darkening.

For form trials, one shape would change its form while the remaining shapes and lines remained unchanged. For categorical form trials, a rectangle, oval, or triangle would become one of the other two possible shapes (e.g., a triangle could become a rectangle or an oval), but with the same relative length as the initial shape. For coordinate form trials, the shape could become shorter or it could become elongated. A roughly equal number of trials featured each possible initial shape that was changed. For categorical form trials, there was a roughly even split as to what the initial shape changed into (e.g., a triangle changed into a rectangle a roughly equal number of times as a triangle changed into an oval). An equal number of coordinate form trials featured the shape becoming shorter or becoming longer.

For orientation trials, a line attached to a shape or an unattached line was rotated. For categorical orientation trials an oblique line (e.g., not horizontal or vertical) was modified to become horizontal or vertical (i.e., parallel to an edge of the computer screen). Categorical orientation trials were evenly split between horizontal and vertical lines, in terms of the resultant “changed” versions of the stimuli. For coordinate orientation trials, an oblique line was modified to become a different oblique line, such that its angle of “categorical” orientation relative to the picture remained unchanged (e.g., a line pointing up and to the left was changed to continue to point that same direction but at a different angle). Orientation trials were evenly split such that an equal number of lines started with each of the four relative pointing positions (i.e., top-left, top-right, bottom-left, and bottom-right). Angle changes were constant between the categorical and coordinate versions of each stimulus such that the distance rotated was the same relative to the original stimulus' orientation. These angle changes were evenly split between 10, 15, 25, and 30 degree-change trials.

For position trials, relationships of above/below were manipulated. For categorical position trials, a relationship of above was changed to below, or vice versa. For categorical position trials, an equal number of trials featured moves from "above to below" as "below to above." For coordinate position trials, manipulation of position (up, down, or sideways) produced a version of the stimulus that was visibly changed but wherein no categorical positional relationships (e.g., above/below) were changed. Lines, shapes, or one line connected to one shape, could potentially move, and they always moved by at least 100 pixels (>10% of the image width) when moved.

Example trials are presented as Figures 26 (color), 27 (form), 28 (orientation), and 29 (position) and are further elaborated on in Appendix G.

Procedure

Participants first completed four "example trials" to become acquainted with the task. First, participants were exposed to an example of one of the four changes: in order, these were color, form, orientation, and position. They viewed a *simultaneous presentation* of the original stimulus and two versions of a changed stimulus. An example categorical change and an example coordinate change were both presented along with the original stimulus, but the changed stimulus versions were not identified as categorical/coordinate, only as changes. While all three versions of the stimulus were presented and clearly labeled, they pressed the key to indicate the type of change. If incorrect, they repeated that example trial. Participants responded to all trials using a set of four keys on the keyboard, one for each type of change (“C,” “F,” “O,” and “P”). Note that the participants’ goal was to identify which of the trial types occurred (color, form, orientation, or position), but no attention was drawn to the difference between categorical and coordinate trials. The keys to be pressed and their definitions were displayed on the screen for the duration of all example, practice, and experimental trials.

Next, they completed “practice trials” in which *sequential presentations* were used, similarly to the following experimental trials. A total of eight practice trials were used, using two each of the four types of possible stimulus changes (i.e., color, form, orientation, position). The eight practice stimuli were, for each participant, randomly divided into *sets of two* stimuli presented in sequence. First, participants initiated the sequence with a keypress, and viewed two practice stimuli presented in sequence, spaced by numbered fixation crosses (i.e., “+1+” and “+2+”). Fixation crosses appeared for one second, and the stimuli to be studied each appeared for eight seconds. Further, in between each stimulus disappearing and the next screen appearing, there was a one second pause during which nothing was displayed. Next, the stimulus that had been displayed first of the set of two re-appeared, but with an added change. For each practice trial, the change was randomly determined by participant to be a categorical or coordinate change. The participants viewed the changed stimulus until they pressed one of four identification keys (corresponding to color, form, orientation, or position) on their keyboard, and if the correct key was pressed, the stimulus vanished. If the incorrect key was pressed, a red “X” appeared until they pressed the correct key. A blank screen was displayed for 500 ms following each correct key press response. Next, the stimulus that had been displayed second appeared, but with an added change, and the participant also responded to this stimulus. After each set of two, an instruction screen appeared that asked the participant to take a short break if needed before continuing by initiating the next set of two trials with a keypress. The number of completed trials, as well as the number of yet-to-be-completed trials was also presented on this rest screen. The sequence of "sets of two" repeated until all eight practice trials were correctly completed. None of the images used in the example or practice trials were used in the experimental trials.

After the practice trials, participants completed 128 experimental (criterion) trials, split into 64 sets of two stimuli each. The presentation was identical to the practice trials, except that corrective feedback was not provided. This design included 32 trials of each type (form, color, position, orientation) which were presented in a random order for each participant. Two versions of changed stimuli were prepared for each trial, one categorical and the other coordinate, and were randomly assigned throughout the trials such that each participant saw only one randomly-selected version. Both reaction time and accuracy were recorded during the experimental trials.

Study 4 Results

All datasets used for this study have been anonymized and can be accessed through the OSF project page, and are usable with attribution (i.e., cite this project: Toftness, 2022; https://osf.io/u3bxj/).

Overall Performance on Study 4

Raw performance between groups was examined before attempting to control for age and sex. Average accuracy in the aphantasia group (n = 94, M = 43.2%, SD = 9.29%) was significantly higher than for the control group (n = 120, M = 39.8%, SD = 9.52%) across all Study 4 trials ($t(212) = 2.62$, p = .010, d = .360). Average response times in the aphantasia group were significantly longer (*M* = 5782, *SD* = 2243) than for the control group (*M* = 4886, *SD* = 1759) across all Study 4 trials (*t*(212) = 3.27, *p* = .001, *d* = .449). Performance measures after controlling for age and sex are discussed as part of the ANCOVA analyses below.

ANCOVA Investigation into Accuracy

Because accuracy was the main measure of interest, it was investigated as the dependent measure using an ANCOVA approach. A 2 x 2 x 4 mixed ANCOVA was run using group, degree changed, and trial type ((aphantasia x control) x (categorical x coordinate) x (color x form x orientation x position)) while entering age and sex assigned at birth as covariates.

Covariates

There was not a significant main effect of age (F (1, 210) = 2.20, p = .140, η_p^2 = .010, observed power = .315), nor was there a significant main effect of sex (F (1, 210) = .022, p = .882, η_p^2 = .000, observed power = .053). For the purposes of holding age constant, a value of 25.45 was used. For the purposes of holding sex constant, a value of 1.6 was used (where 1 = male and 2 = female). After the Greenhouse-Geisser correction, there was a significant three-way interaction between sex, degree changed, and trial type, such that there were some different slopes between degrees changed for each sex for some, but not all, of the trial types (F (2.97, 622.65) = 2.96, p = .032, η_p^2 = .014, observed power = .698). In fact, the relationship between sex and degree changed looked visually different when examining separate plots for the four trial types, and the most striking slope difference between the sexes occurred for position trials (see Figure 30). No other interactions with covariates were significant.

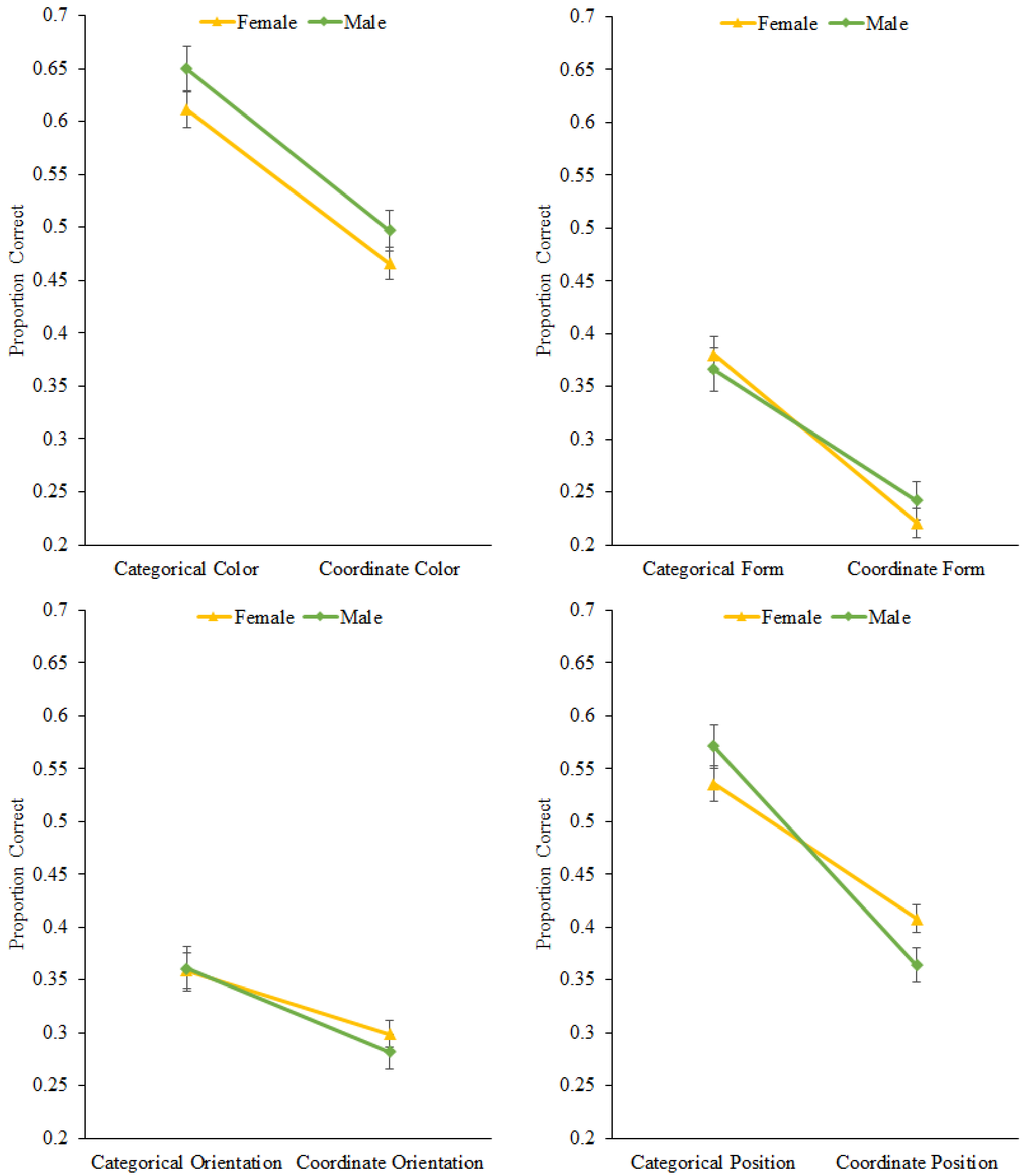


Figure 30. The significant three-way sex by degree changed by trial type (color, form, orientation, or position) interaction on proportion correct for Study 4, plotted separately for all four trial types. The main effect for sex was not significant. Proportion correct appears on the vertical axis. Mean accuracies have been adjusted for age, and the estimated marginal means are depicted. Error bars are standard error.

Main Effects

There was a main effect of group such that the estimated marginal means for the aphantasia group (*M* = 44.1%, *SE* = 1.2%) were significantly higher than those for the control group (*M* = 39.0%, *SE* = 1.0%; *F* (1, 210) = 8.58, *p* = .004, η_p^2 = .039, observed power = .830), indicating that the aphantasia group was more accurate during the trials. There was also a main effect of "degree changed" with the estimated marginal means for categorical trial accuracy (*M* = 48.2%, *SE* = 0.9%) significantly higher than for coordinate trial accuracy (*M* = 35.0%, *SE* = 0.6%; *F* (1, 210) = 22.35, *p* < .001, η_p^2 = .096, observed power = .997). After the Greenhouse-Geisser correction, there was also a main effect of trial type (color x form x orientation x position; *F* (2.94, 617.80) = 6.68, *p* < .001, η_p^2 = .031, observed power = .973). Estimated marginal means for accuracy was highest on the color trials (55.6%), followed by the position trials (47.1%), then the orientation trials (33.1%), and finally the form trials (30.6%). Each of these trial types will be examined separately below.

Two-Way Interactions

There was a significant interaction between group and degree changed (see Figure 31) such that the negative slope from categorical trials to coordinate trials was steeper for the aphantasia group than it was for the control group (*F* (1, 210) = 4.11, *p* = .044, η_p^2 = .019, observed power = .523). In terms of simple effects, accuracy was significantly different between the categorical and coordinate trials when considering only participants in the aphantasia group (*F* (1, 210) = 123.47, *p* < .001, η_p^2 = .370, observed power = 1) and when considering only participants in the control group (*F* (1, 210) = 89.63, *p* < .001, η_p^2 = .299, observed power = 1). When considering only categorical trials, the difference in accuracy between the aphantasia group and the control group was significant (*F* (1, 210) = 8.83, *p* = .003, η_p^2 = .040, observed power = .841), but this was less true for coordinate trials, with the difference in accuracy between the groups falling just out of significance (F (1, 210) = 3.85, p = .051, η_p^2 = .018, observed power = .497). Considering that this interaction pools all four trial types, it was important to consider the four trial types separately, as well (see below).

There was no interaction detected between group and trial type (F (2.94, 617.80) = 1.05, p = .371, η_p^2 = .005, observed power = .28). There was a significant interaction detected between degree changed and trial type (F (2.97, 622.65) = 3.13, p = .026, η_p^2 = .015, observed power = .725). Thus, the degree changes (categorical versus coordinate) within the four trial types (color, form, orientation, and position) need to be further analyzed, and are considered separately below.

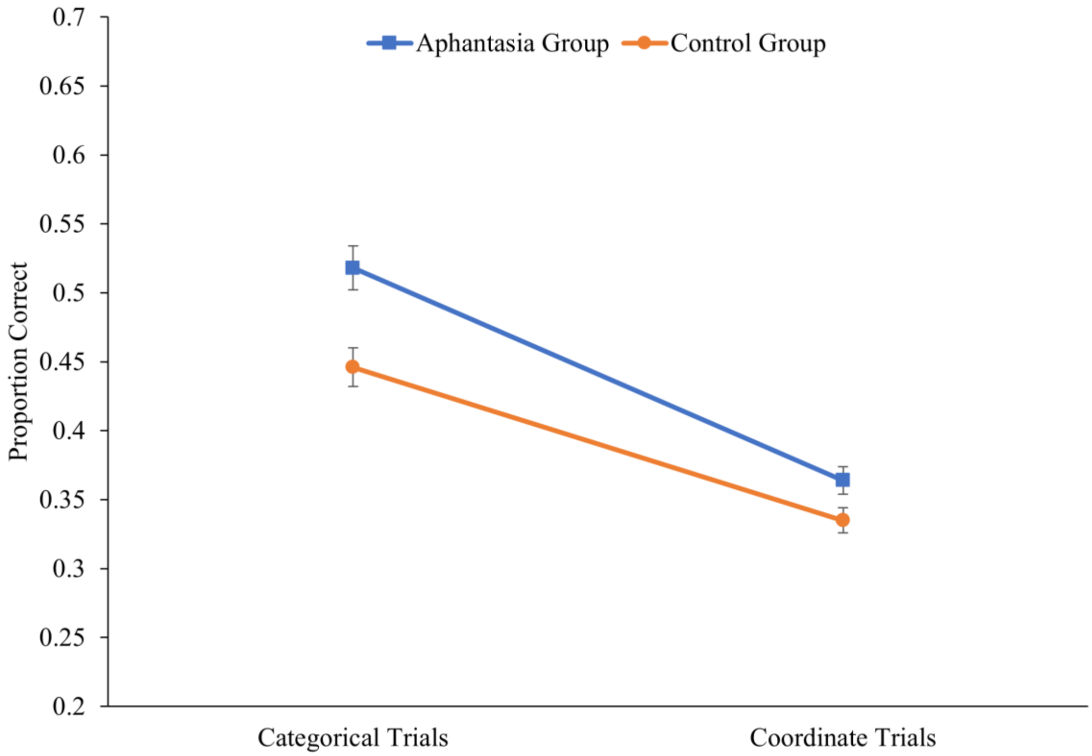


Figure 31. The significant group x degree changed interaction on accuracy from Study 4, collapsed across all four trial types. The main effects of group and degree changed are both significant. Accuracy appears on the y-axis. Mean accuracies have been adjusted for age and sex, and estimated marginal means are depicted. Error bars are standard error.

Three-Way Interaction

No three-way interaction was detected between group, degree changed, and trial type (*F* (2.97, 622.65) = 1.91, *p* = .128, η_p^2 = .009, observed power = .491).

Accuracy on Color Trials

To parcel out effects within the color trials, a 2 x 2 mixed ANCOVA was run using group and degree changed ((aphantasia x control) x (categorical x coordinate)) with accuracy on color trials as the dependent measure, and with age and sex as covariates (see Figure 32). For color trials, there were no main effects of age (*F* (1, 210) = .865, *p* = .353, η_p^2 = .004, observed power = .152) or sex (*F* (1, 210) = 2.04, *p* = .154, η_p^2 = .010, observed power = .296), or any significant interactions with the covariates. There was also no main effect of group, although it approached significance—the estimated marginal mean for the aphantasia group was 58.2% (*SE* = 1.9%) and for the control group it was 53.0% (*SE* = 1.6%; *F* (1, 210) = 3.49, *p* = .063, η_p^2 = .016, observed power = .460). There was a main effect of degree changed, such that participants tended to be more accurate on categorical trials (*M* = 62.9%, *SE* = 1.3%) than on coordinate trials (*M* = 48.2%, *SE* = 1.2%; *F* (1, 210) = 4.58, *p* = .033, η_p^2 = .021, observed power = .568). No significant interaction between group and degree changed was detected (*F* (1, 210) = .099, *p* = .753, η_p^2 = .000, observed power = .061).

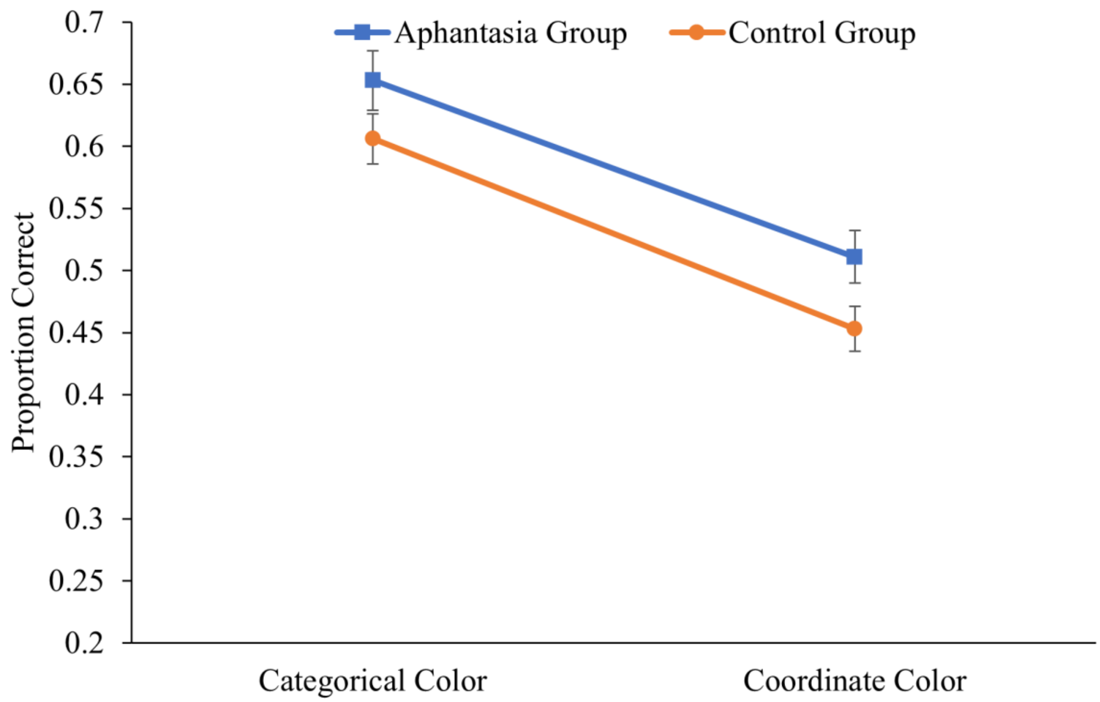


Figure 32. The non-significant group x degree changed interaction on accuracy for color trials in Study 4. The main effect of group was not significant. The main effect for degree changed was significant. Accuracy appears on the y-axis. Mean accuracies have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

Accuracy on Form Trials

A 2 x 2 mixed ANCOVA was run using group and degree changed ((aphantasia x control) x (categorical x coordinate)) with accuracy on form trials as the dependent measure, and with age and sex as covariates (see Figure 33). For form trials, there were no main effects of age (F (1, 210) = 1.18, p = .278, η_p^2 = .006, observed power = .191) or sex (F (1, 210) = .013, p = .909, η_p^2 = .000, observed power = .051), or any significant interactions with covariates. There was a main effect of group: the estimated marginal mean was higher for the aphantasia group at 33.7% (SE = 1.8%) than the control group at 27.5% (SE = 1.6%; F (1, 210) = 5.31, p = .022, η_p^2 = .025, observed power = .631). No main effect was detected for degree changed: participants performed similarly on categorical trials (M = 38.0%, SE = 1.3%) and coordinate trials (M = 23.2%, SE = 1.1%; F (1, 210) = 2.30, p = .131, η_p^2 = .011, observed power = .327). The interaction was not significant (F (1, 210) = 2.14, p = .145, η_p^2 = .010, observed power = .308). There seems to be a floor effect here for coordinate trials, such that average accuracy on coordinate form trials was not significantly different from random responding—a 25% average accuracy rate would be expected based on chance alone due to there being four trial types, but participants attained only 22.9% correct on average on coordinate form trials (SD = 15.9%; t(214) = 1.89, p = .061).

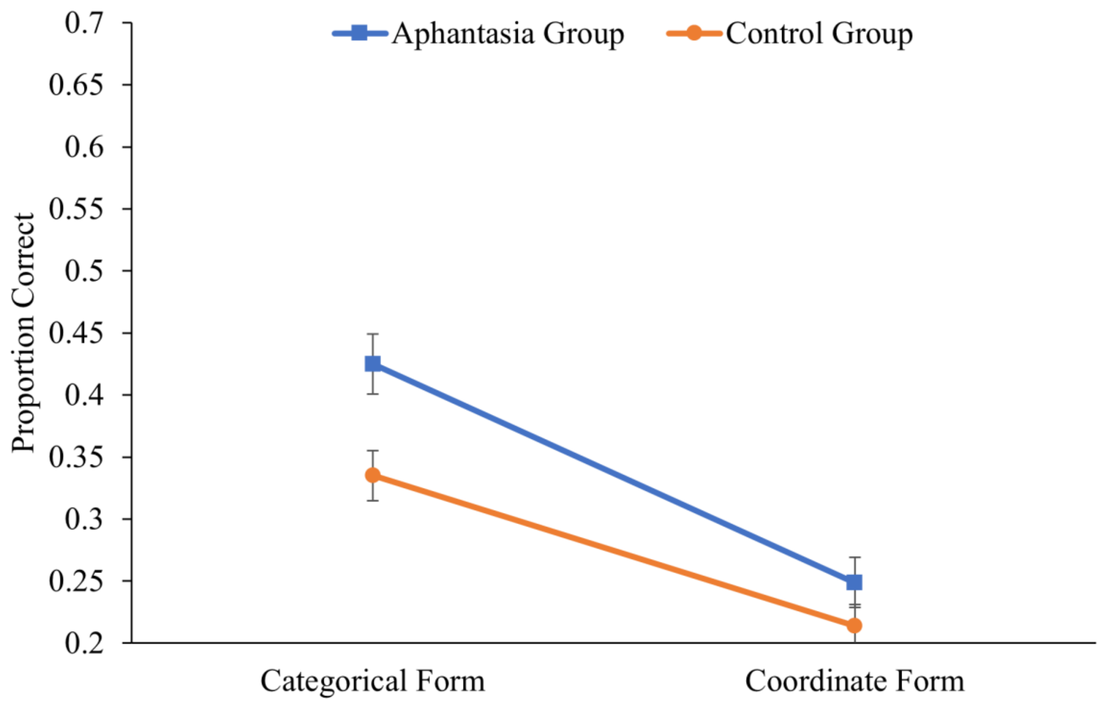


Figure 33. The non-significant group x degree changed interaction on accuracy for form trials in Study 4. The main effect of group was significant. The main effect for degree changed was not significant. Accuracy appears on the y-axis. Mean accuracies have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

Accuracy on Orientation Trials

A 2 x 2 mixed ANCOVA was run using group and degree changed ((aphantasia x control) x (categorical x coordinate)) with accuracy on orientation trials as the dependent measure, and with age and sex as covariates (see Figure 34). For orientation trials, there was a main effect of age (F (1, 210) = 5.19, p = .024, η_p^2 = .024, observed power = .621). There was no detected main effect of sex (*F* (1, 210) = .516, *p* = .473, η_p^2 = .002, observed power = .110). There were no significant interactions with the covariates. There was a main effect of group: the estimated marginal mean was higher for the aphantasia group at 36.7% (*SE* = 1.7%) than the control group at 29.5% (*SE* = 1.5%; *F* (1, 210) = 7.87, *p* = .006, η_p^2 = .036, observed power = .797). There was no main effect detected for degree changed, although the difference approached significance such that participants performed numerically but not significantly better on categorical trials (*M* = 36.5%, *SE* = 1.3%) than coordinate trials (*M* = 29.6%, *SE* = 1.0%; *F* (1, 210) = 3.67, *p* = .057, η_p^2 = .017, observed power = .479). The interaction was not significant (*F* (1, 210) = .518, *p* = .473, η_p^2 = .002, observed power = .111).

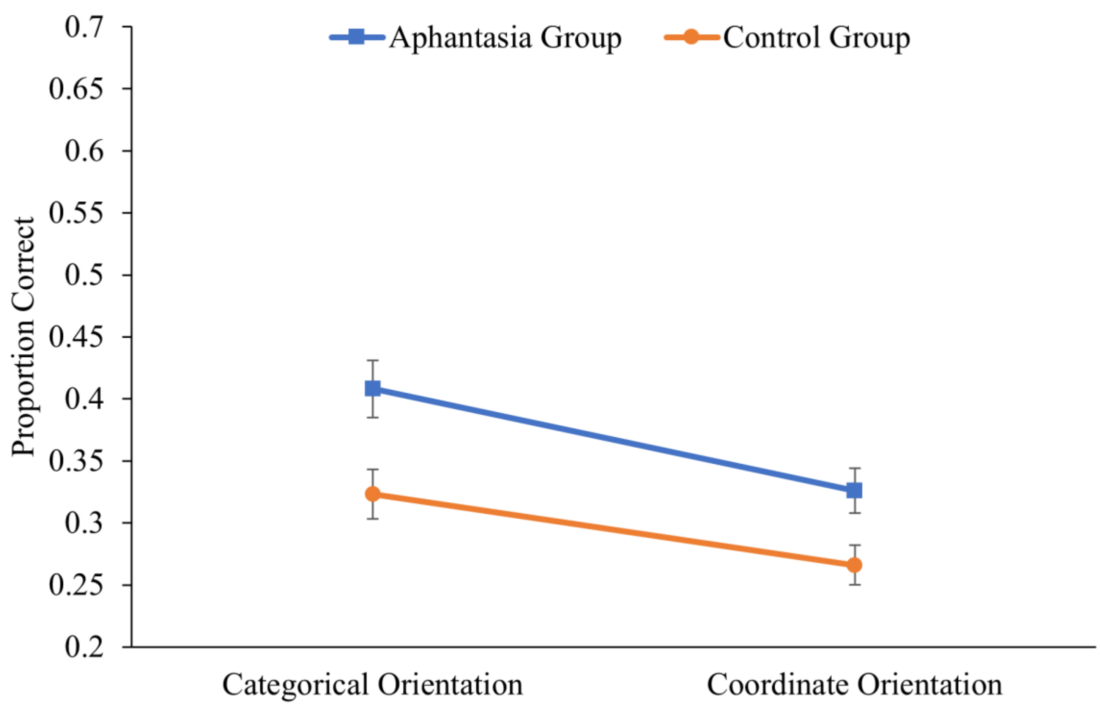


Figure 34. The non-significant group x degree changed interaction in accuracy for orientation trials in Study 4. The main effect of group was significant. The main effect for degree changed was not significant. Accuracy appears on the y-axis. Mean accuracies have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

Accuracy on Position Trials

A 2 x 2 mixed ANCOVA was run using group and degree changed ((aphantasia x control) x (categorical x coordinate)) with accuracy on position trials as the dependent measure, and with age and sex as covariates (see Figure 35). For position trials, there were no main effects of age (*F* (1, 210) = 2.12, *p* = .147, η_p^2 = .010, observed power = .305) or sex (*F* (1, 210) = .082, *p* = .775, η_p^2 = .000, observed power = .082). There was a significant interaction between sex and degree changed, which was already mentioned (see Figure 30; (*F* (1, 210) = 6.58, *p* = .011, η_p^2 = .030, observed power = .723). The main effect of group was not significant: the estimated marginal mean was similar for the aphantasia group at 47.9% (*SE* = 1.9%) and the control group at 46.3% (*SE* = 1.6%; *F* (1, 210) = .325, *p* = .570, η_p^2 = .002, observed power = .088). There was a significant main effect for degree changed, such that participants performed more accurately on categorical trials (*M* = 55.4%, *SE* = 1.3%) than coordinate trials (*M* = 38.9%, *SE* = 1.1%; *F* (1, 210) = 27.32, *p* < .001, η_p^2 = .115, observed power = .999). There was also a significant interaction, such that there was a steeper slope for the aphantasia group than for the control group when considering the reduction in accuracy between the categorical and coordinate trials (*F* (1, 210) = 7.87, *p* = .005, η_p^2 = .036, observed power = .798). The simple effects were then examined. When considering only the categorical trials, there was not a significant difference between the accuracy of the aphantasia group and the control group (*F* (1, 210) = 3.54, *p* = .061, η_p^2 = .017, observed power = .466). When considering only the coordinate trials, there was not a significant difference between the aphantasia group and the control group (*F* (1, 210) = 1.25, *p* = .265, η_p^2 = .006, observed power = .199). The difference in accuracy between the categorical and coordinate trials was significant when looking only at the aphantasia group (*F* (1, 210) = 80.27, *p*

< .001, η_p^2 = .277, observed power = 1) and when looking only at the control group (*F* (1, 210) = 31.28, *p* < .001, η_p^2 = .130, observed power = 1).

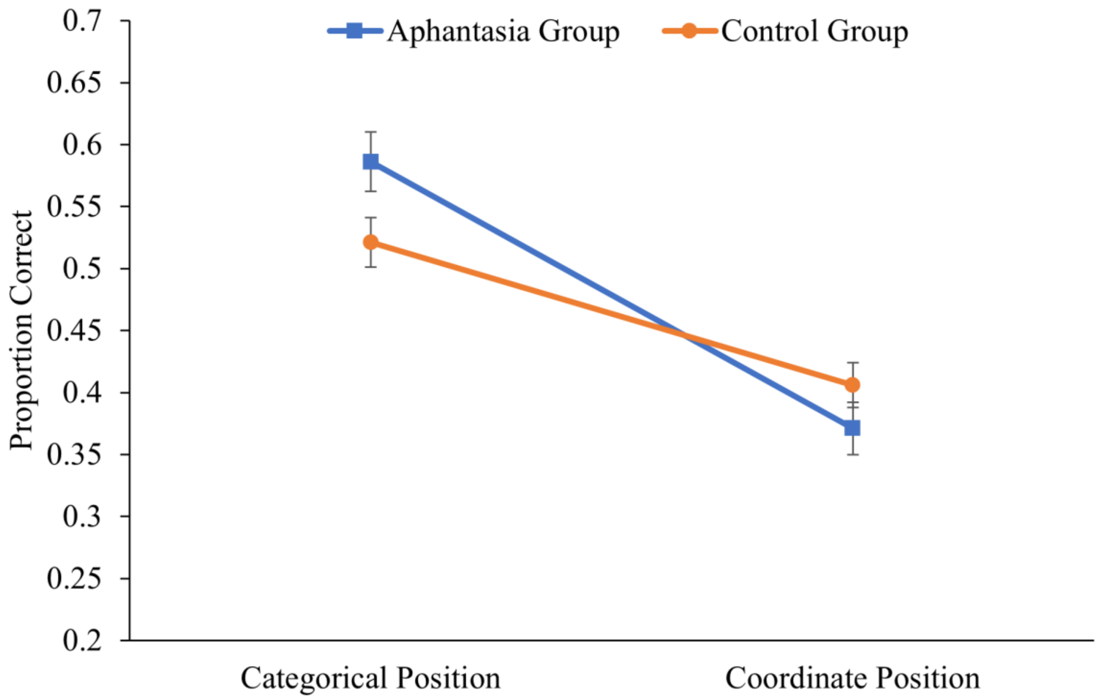


Figure 35. The significant group x degree changed interaction on accuracy for position trials in Study 4. The main effect of group was not significant. The main effect for degree changed was significant. Accuracy appears on the y-axis. Mean accuracies have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

ANCOVA Investigation into Response Time

Response time was also investigated as a dependent measure using an ANCOVA approach. A 2 x 2 x 4 mixed ANCOVA was run using group, categorical/coordinate, and trial type ((aphantasia x control) x (categorical x coordinate) x (color x form x orientation x position)) while entering age and sex assigned at birth as covariates.

Covariates

There was not a significant main effect of age (*F* (1, 210) = 1.91, *p* = .168, η_p^2 = .009, observed power = .280), nor was there a significant main effect of sex (*F* (1, 210) = .170, *p* = .681, η_p^2 = .001, observed power = .069). For the purposes of holding age constant, a value of

25.45 was used. For the purposes of holding sex constant, a value of 1.6 was used (where 1 = male and 2 = female). There was a significant interaction between sex and degree changed, such that the slope of the increased response time between categorical and coordinate trials was steeper for males than it was for females (see Figure 36; F (1, 210) = 4.29, p = .040, η_p^2 = .020, observed power = .541). To further investigate the sex x degree changed interaction, sex was entered into the model as a between-subjects variable instead of as a covariate. The simple effects were then investigated. The difference in response time between males and females was not significant when looking at only the categorical trials (F (1, 209) = .074, p = .786, η_p^2 = .000, observed power = .058) nor was it significant when looking only at the coordinate trials (F (1, 209) = .762, p = .384, η_p^2 = .004, observed power = .140). The difference in accuracy between the categorical and coordinate trials was significant when looking only at males (F (1, 209) = 57.51, p < .001, η_p^2 = .216, observed power = 1) as well as when looking only at females (F (1, 209) = 34.75, p < .001, η_p^2 = .143, observed power = 1). No other interactions with covariates were significant.

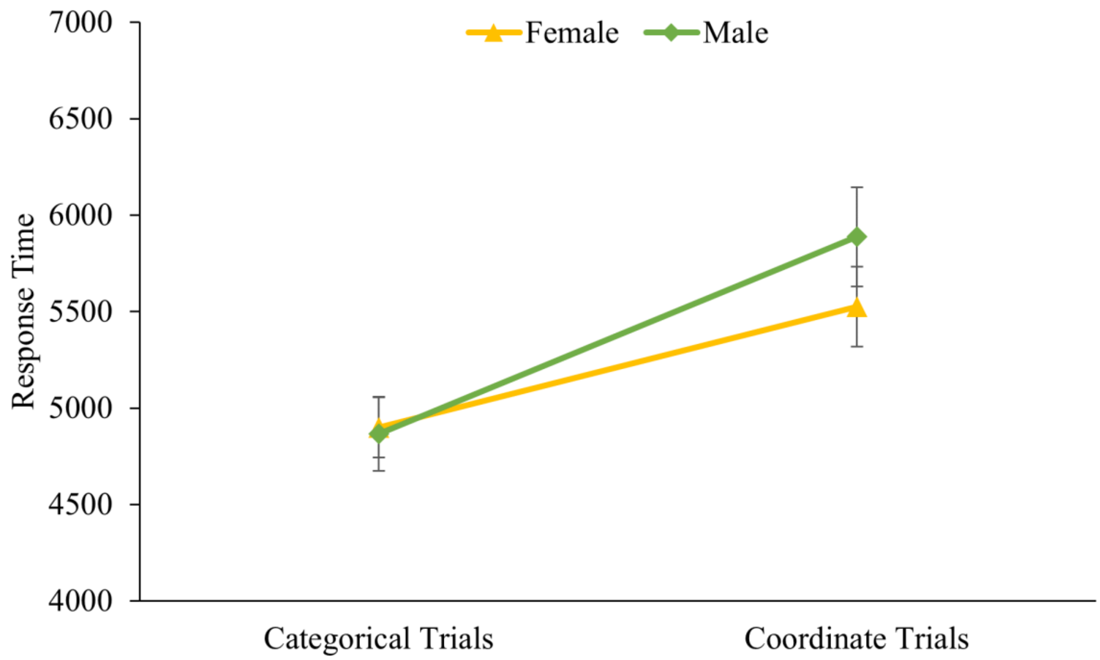


Figure 36. The significant sex x degree changed interaction on response times for Study 4, collapsed across all four trial types. Response times in ms are on the y-axis. Mean response times were adjusted for age, and estimated marginal means are depicted. Error bars are standard error.

Main Effects

No main effect was detected for group: marginal means of response times for the aphantasia group (*M* = 5574 ms, *SE* = 246) were similar to the control group (*M* = 5045 ms, *SE* = 210; *F* (1, 210) = 2.09, *p* = .150, η_p^2 = .010, observed power = .301). There was a main effect of "degree changed" with marginal means of response times for categorical trials (*M* = 4908, *SE* = 122) significantly lower than for coordinate trials (*M* = 5711, *SE* = 162; *F* (1, 210) = 14.78, *p* < .001, η_p^2 = .066, observed power = .969). After the Greenhouse-Geisser correction, there was no detected main effect of trial type (color x form x orientation x position; *F* (2.63, 553.14) = 1.04, *p* = .369, η_p^2 = .005, observed power = .265). Estimated marginal means for response times were numerically highest for the orientation trials (5651 ms), followed by the form trials (5504 ms), then the position trials (5119 ms), and finally the color trials (4964 ms). Each of these trial types are examined separately below.

Two-Way Interactions

There was not a significant interaction between group and degree changed (see Figure 37; F (1, 210) = 2.08, p = .151, η_p^2 = .010, observed power = .300). There was no interaction detected between group and trial type (F (2.63, 553.14) = .785, p = .488, η_p^2 = .004, observed power = .207). There was no significant interaction detected between degree changed and trial type (F (2.85, 599.28) = .247, p = .854, η_p^2 = .001, observed power = .096).

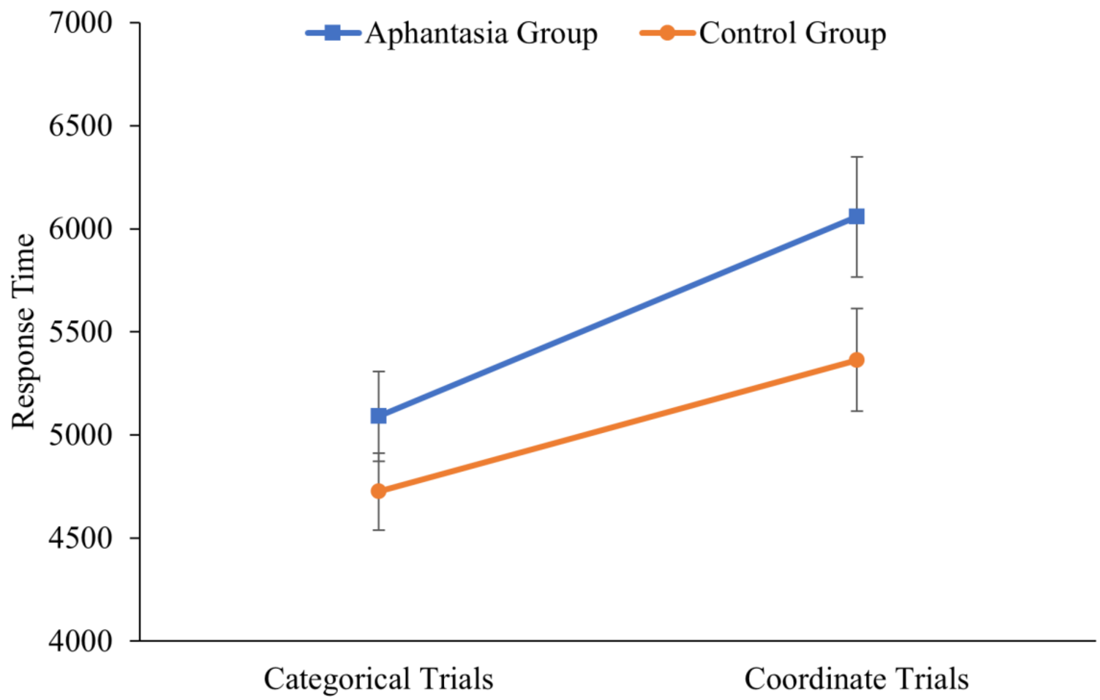


Figure 37. The non-significant group x degree changed interaction on response times for Study 4, collapsed across all four trial types. The main effect for group was not significant. The main effect of degree changed was significant. Response times appear in terms of ms on the y-axis. Mean response times have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

Three-Way Interaction

There was not a significant interaction between group, degree changed, and trial type (F (2.85, 599.28) = 1.18, p = .316, η_p^2 = .006, observed power = .310).

Response Time on Color Trials

A 2 x 2 mixed ANCOVA was run using group and degree changed ((aphantasia x control) x (categorical x coordinate)) with response time on color trials as the dependent measure, and with age and sex as covariates (see Figure 38). There were no main effects of age (F (1, 210) = .309, p = .579, η_p^2 = .001, observed power = .086) or sex (F (1, 210) = .056, p = .813, η_p^2 = .000, observed power = .056), or any significant covariate interactions. No main effect of group was detected: marginal means for the aphantasia group (M = 5230 ms, SE = 253) and the control group (M = 4699 ms, SE = 217) were similar (F (1, 210) = 1.97, p = .162, η_p^2 = .009, observed power = .287). There was a main effect of degree changed: responses were faster on categorical trials (M = 4500, SE = 147) than on coordinate trials (M = 5429, SE = 181; F (1, 210) = 5.52, p = .020, η_p^2 = .026, observed power = .647). No significant group x degree changed interaction was detected (F (1, 210) = .208, p = .649, η_p^2 = .001, observed power = .074).

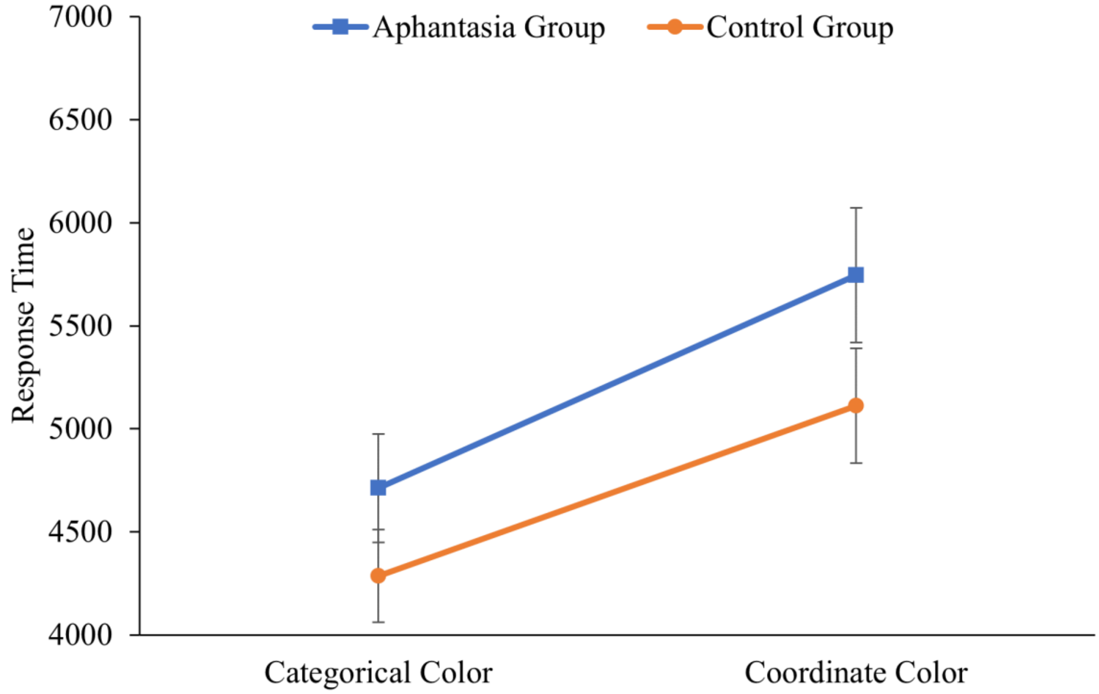


Figure 38. The non-significant group x degree changed interaction on response times for color trials in Study 4. The main effect for group was not significant. The main effect of degree changed was significant. Response times appear in terms of ms on the y-axis. Mean response times have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

Response Time on Form Trials

A 2 x 2 mixed ANCOVA was run using group and degree changed ((aphantasia x control) x (categorical x coordinate)) with response time on form trials as the dependent measure, and with age and sex as covariates (see Figure 39). There were no main effects of age (F (1, 210) = 3.22, p = .074, η_p^2 = .015, observed power = .431) or sex (F (1, 210) = .152, p = .697, η_p^2 = .001, observed power = .067), or any significant covariate interactions. No main effect of group was detected: marginal means were similar for the aphantasia group at 5753 ms (SE = 263) and the control group at 5255 ms (SE = 225; F (1, 210) = 1.61, p = .206, η_p^2 = .008, observed power = .244). There was a main effect of degree changed: participants responded faster to categorical trials (M = 5085, SE = 136) than to coordinate trials (M = 5923, SE = 178; F

(1, 210) = 9.01, p = .003, η_p^2 = .041, observed power = .848). There was also a significant interaction between group and degree changed such that the slope of the increase in response times between the categorical and coordinate trials was steeper for the aphantasia group than it was for the control group (F (1, 210) = 6.32, p = .013, η_p^2 = .029, observed power = .706). Simple effects were then investigated. When considering only categorical trials, the difference in response times between the aphantasia and control groups were non-significant (F (1, 210) = .066, p = .797, η_p^2 = .000, observed power = .058). This was also true when considering only coordinate trials, although much more marginally so (F (1, 210) = 3.58, p = .060, η_p^2 = .017, observed power = .469). The difference in response times between the categorical and coordinate trials was significant when considering only the aphantasia group (F (1, 210) = 33.16, p < .001, η_p^2 = .136, observed power = 1) and when considering only the control group (F (1, 210) = 5.55, p = .019, η_p^2 = .026, observed power = .650).

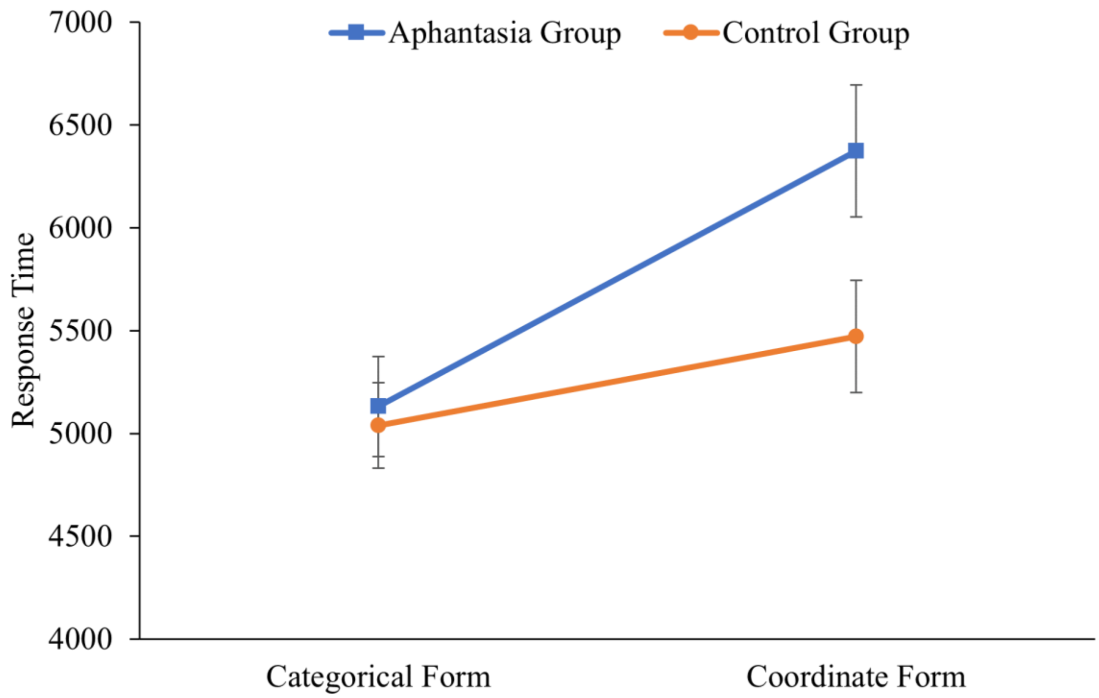


Figure 39. The significant group x degree changed interaction on response times for form trials in Study 4. The main effect for group was not significant. The main effect of degree changed was significant. Response times appear in ms on the y-axis. Mean response times have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

Response Time on Orientation Trials

A 2 x 2 mixed ANCOVA was run using group and degree changed ((aphantasia x control) x (categorical x coordinate)) with response time on orientation trials as the dependent measure, and with age and sex as covariates (see Figure 40). No main effect was detected for age (F (1, 210) = 1.32, p = .253, η_p^2 = .006, observed power = .208), or sex (F (1, 210) = .251, p = .617, η_p^2 = .001, observed power = .079), and there were no significant covariate interactions. No main effect was detected for group: marginal means were similar for the aphantasia group at 6015 ms (SE = 296) and the control group at 5287 ms (SE = 253; F (1, 210) = 2.73, p = .100, η_p^2 = .013, observed power = .376). There was a main effect for degree changed: participants responded faster on categorical trials (M = 5418, SE = 173) than coordinate trials (M = 5884, SE = 179; F (1, 210) = 4.46, p = .036, η_p^2 = .021, observed power = .557). No significant interaction between group and degree changed was detected (F (1, 210) = 1.40, p = .239, η_p^2 = .007, observed power = .218).

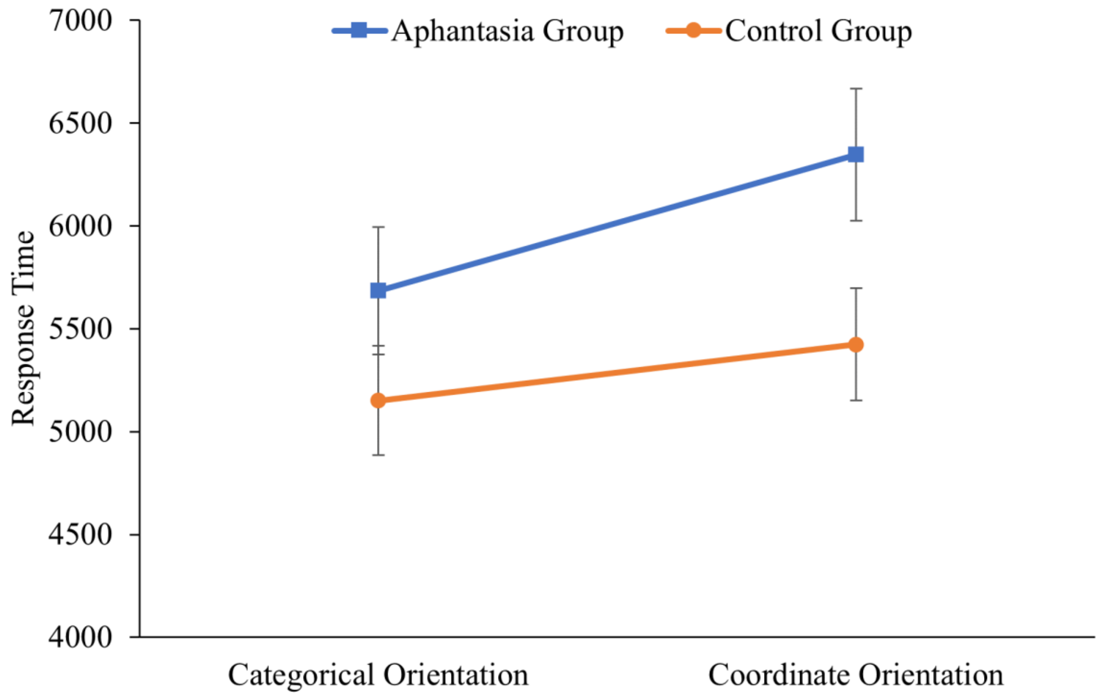


Figure 40. The non-significant group x degree changed interaction on response times for orientation trials in Study 4. The main effect for group was not significant. The main effect of degree changed was significant. Response times appear in terms of ms on the y-axis. Mean response times have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

Response Time on Position Trials

A 2 x 2 mixed ANCOVA was run using group and degree changed ((aphantasia x control) x (categorical x coordinate)) with accuracy on position trials as the dependent measure, and with age and sex as covariates (see Figure 41). For position trials, there were no main effects of age (F (1, 210) = 2.74, p = .099, η_p^2 = .013, observed power = .378) or sex (F (1, 210) = .147, p = .702, η_p^2 = .001, observed power = .067). There were no significant covariate interactions. There was no detected main effect of group: marginal means were similar for the aphantasia group at 5299 ms (SE = 245) and the control group at 4939 ms (SE = 209; F (1, 210) = .971, p =

.326, η_p^2 = .005, observed power = .165). The main effect for degree changed was significant: participants responded faster on categorical trials (*M* = 4630, *SE* = 119) than on coordinate trials (*M* = 5608, *SE* = 182; *F* (1, 210) = 4.34, *p* = .039, η_p^2 = .020, observed power = .545). An interaction between group and degree changed was not detected (*F* (1, 210) = .048, *p* = .827, η_p^2 = .000, observed power = .055).

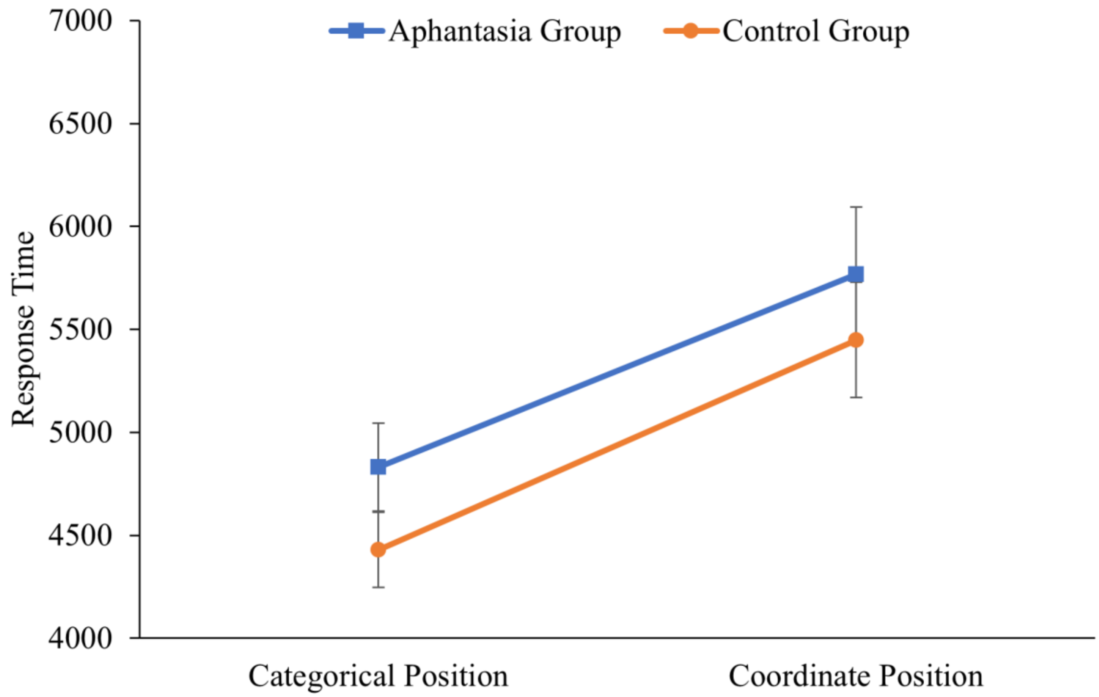


Figure 41. The non-significant group x degree changed interaction on response times for position trials in Study 4. The main effect for group was not significant. The main effect of degree changed was significant. Response times appear in terms of ms on the y-axis. Mean response times have been adjusted for age and sex, and the estimated marginal means are depicted. Error bars are standard error.

ANCOVA Investigation into Rate Correct Scores

Because it could be argued that different participants could prioritize accuracy and response time differently, and there was no clear prediction in terms of whether accuracy or response time would show a crucial effect, I also created a measure that captured the variance from both accuracy and response time by combining those two measures in a linear fashion: rate correct score (RCS). This additional approach is justified in the Supplemental Analyses.

Rate correct score was investigated as the dependent measure using an ANCOVA approach. Four 2 x 2 mixed ANCOVAs were run, one for each trial type, using group and degree changed ((aphantasia x control) x (categorical x coordinate)) while entering age and sex assigned at birth as covariates. The trial types are reported below, and depicted in Figure 42.

Rate Correct Scores on Color Trials

There were no main effects of age (F (1, 210) = .016, p = .899, η_p^2 = .000, observed power = .052) or sex (F (1, 210) = 2.03, p = .156, η_p^2 = .010, observed power = .294), or any significant covariate interactions. No main effect of group was detected: marginal means for the aphantasia group (M = .135, SE = .008) and the control group (M = .137, SE = .007) were similar (F (1, 210) = .036, p = .850, η_p^2 = .000, observed power = .054). There was a main effect of degree changed: the average correct responses per second were higher for categorical trials (M = .166, SE = .006) than for coordinate trials (M = .107, SE = .004; F (1, 210) = , p = .007, η_p^2 = .034, observed power = .777). No significant group x degree changed interaction was detected (see Figure 42; F (1, 210) = .164, p = .686, η_p^2 = .001, observed power = .069).

Rate Correct Scores on Form Trials

There were no main effects of age (F (1, 210) = 2.31, p = .130, η_p^2 = .011, observed power = .328) or sex (F (1, 210) = .045, p = .833, η_p^2 = .000, observed power = .055), or any significant covariate interactions. No main effect of group was detected: marginal means for the aphantasia group (M = .064, SE = .004) and the control group (M = .062, SE = .004) were similar (F (1, 210) = .205, p = .651, η_p^2 = .001, observed power = .074). There was a main effect of degree changed: the average correct responses per second were higher for categorical trials (M =

.081, *SE* = .003) than for coordinate trials (*M* = .045, *SE* = .003; *F* (1, 210) = 8.41, *p* = .004, η_p^2 = .038, observed power = .823). No significant group x degree changed interaction was detected (see Figure 42; *F* (1, 210) = 2.63, *p* = .107, η_p^2 = .012, observed power = .364).

Rate Correct Scores on Orientation Trials

There was a significant main effect of age (*F* (1, 210) = 6.95, *p* = .009, η_p^2 = .032, observed power = .747). There was not a significant main effect of sex (*F* (1, 210) = .377, *p* = .540, η_p^2 = .002, observed power = .094), or any significant covariate interactions. No main effect of group was detected: marginal means for the aphantasia group (*M* = .071, *SE* = .004) and the control group (*M* = .064, *SE* = .004) were similar (*F* (1, 210) = 1.25, *p* = .265, η_p^2 = .006, observed power = .199). Unlike the other trial types, there was no main effect of degree changed: the average correct responses per second were similar for categorical trials (*M* = .076, *SE* = .003) and coordinate trials (*M* = .059, *SE* = .003; *F* (1, 210) = 3.51, *p* = .062, η_p^2 = .016, observed power = .463). No significant group x degree changed interaction was detected (see Figure 42; *F* (1, 210) = .640, *p* = .425, η_p^2 = .003, observed power = .125).

Rate Correct Scores on Position Trials

There was a main effect of age (*F* (1, 210) = 4.40, *p* = .037, η_p^2 = .021, observed power = .551). There was not a significant main effect of sex (*F* (1, 210) = .101, *p* = .750, η_p^2 = .000, observed power = .062), or any significant covariate interactions. No main effect of group was detected: marginal means for the aphantasia group (*M* = .111, *SE* = .005) and the control group (*M* = .106, *SE* = .006) were similar (*F* (1, 210) = .255, *p* = .614, η_p^2 = .001, observed power = .079). There was a main effect of degree changed: the average correct responses per second were higher for categorical trials (*M* = , *SE* =) than for coordinate trials (*M* = , *SE* = ; *F* (1, 210) = 27.73, *p* < .001, η_p^2 = .117, observed power = .999). There was a significant group x degree changed interaction (see Figure 42; F (1, 210) = 4.104, p = .044, η_p^2 = .019, observed power = .523).

The simple effects were then examined. When considering only the categorical trials, there was not a significant difference between the RCS of the aphantasia group and the control group (F (1, 210) = .449, p = .503, η_p^2 = .002, observed power = .102). When considering only the coordinate trials, there was not a significant difference between the aphantasia group and the control group, albeit the effect was just outside of significance (F (1, 210) = 3.57, p = .060, η_p^2 = .017, observed power = .468). The difference in accuracy between the categorical and coordinate trials was significant when looking only at the aphantasia group (F (1, 210) = 61.54, $p < .001$, η_p^2 = .227, observed power = 1) and when looking only at the control group (F (1, 210) = 31.89, $p < .001$, η_p^2 = .132, observed power = 1).

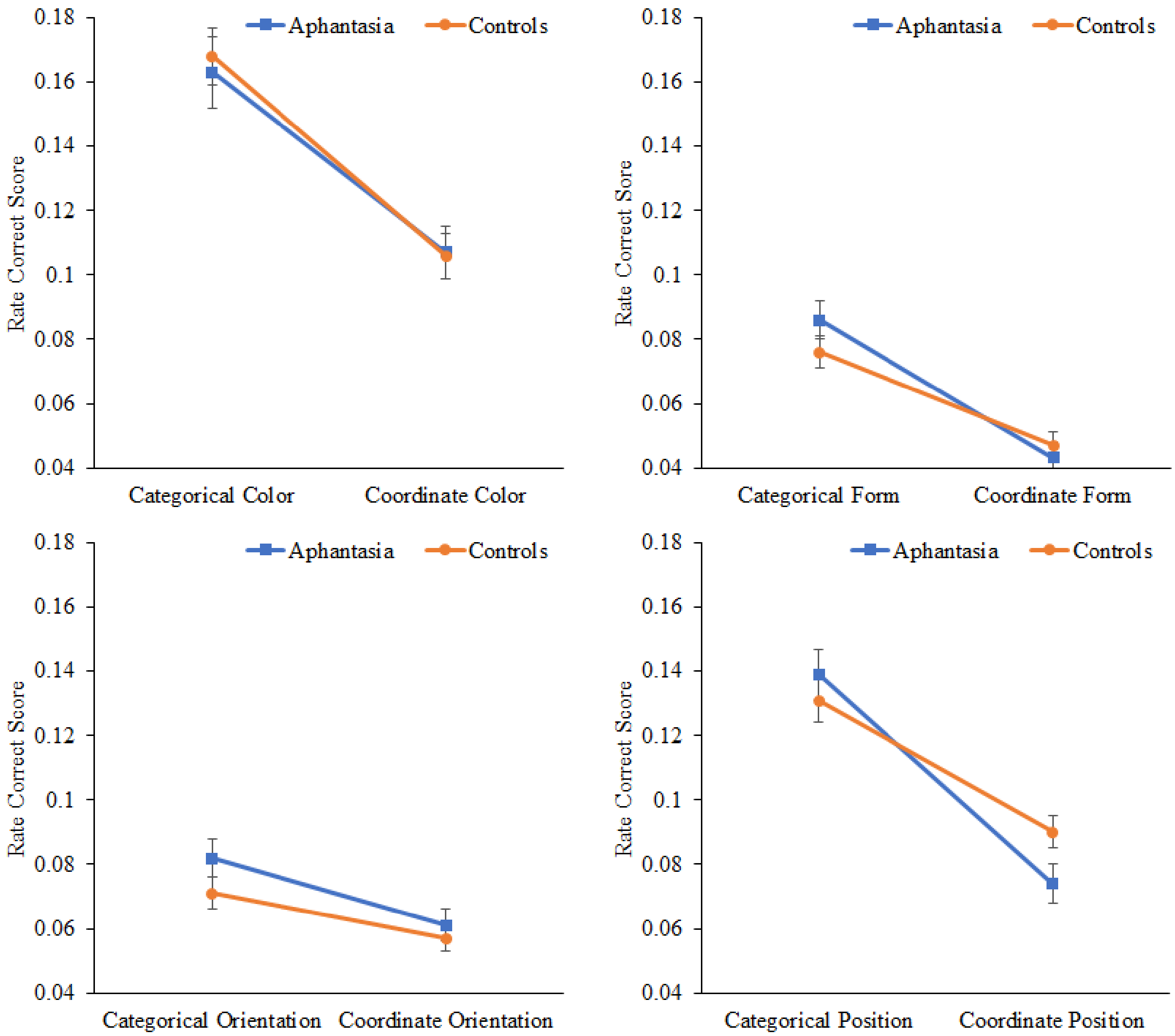


Figure 42. The four trial types from Study 4 shown as group x degree changed interactions, with rate correct score (correct responses per second) on the vertical axis. Means have been adjusted for age and sex, and estimated marginal means appear on the vertical axis. The only interaction that was significant was for position trials.

Self-Reported Imagery Use and VVIQ During Study 4

The aphantasia group reported significantly less use of visual mental imagery during Study 4 (n = 94, M = 1.12, SD = .355) than did the control group (n = 120, M = 2.57, SD = .847) after adjusting the degrees of freedom for a significant Levene's test (F = 102.98, p < .001; t(167.50) = 16.94, p < .001, d = 2.62). Within the control group, participants who were randomly assigned to watch the imagery video had significantly lower SIU for Study 4 (n = 64, M = 2.38, *SD* = .807) when compared to the control participants who watched the color perception video (*n* = 56, *M* = 2.79, *SD* = .847; $t(118) = 2.72$, $p = .008$, $d = .501$).

When considering all participants, there was a significant negative correlation between VVIQ scores and accuracy on the change identification task such that participants with lower VVIQ scores had higher accuracy ($r(212) = -.156$, $p = .022$). The correlation remained significant even when age and sex were controlled for (partial $r(210) = -.161$, $p = .019$). There was also a significant negative correlation between VVIQ score and response times on the change identification task such that participants with lower VVIQ scores had longer response times ($r(212) = -.235$, $p = .001$). The correlation remained *barely* significant even after controlling for age and sex (partial $r(210) = -.135$, $p = .050$). Clearly, participants in the aphantasia group sacrificed response time in order to enhance accuracy. A better measure that accounts for both is RCS. When VVIQ scores were correlated with RCS, the correlation was small and positive ($r(212) = .141$, $p = .039$) such that participants with higher VVIQ scores had more correct responses per second. However, after controlling for age and sex, this correlation between VVIQ scores and RCS vanished (partial $r(210) = .063$, $p = .361$). Thus, evidence is weak that VVIQ scores were related to performance on the change identification task.

SIU on the change identification task used in Study 4 had a strong positive correlation with VVIQ scores ($r(212) = .719$, $p < .001$). This was true even after controlling for age and sex (partial $r(210) = .607$, $p < .001$). When considering only participants with aphantasia, the correlation between SIU and VVIQ scores was strong and positive ($r(92) = .563$, $p < .001$). However, when considering only participants in the control condition, the correlation is not significant ($r(118) = .153$, $p = .096$). When considering only participants in the control condition who watched the imagery video, that relationship is not significant ($r(62) = .168$, $p = .184$).

When only participants in the control condition who watched the color video are considered, the relationship is not significant ($r(54) = .017$, $p = .901$).

When considering all participants, and controlling for age and sex, correlations between SIU for Study 4 and task performance as measured by rate correct score (RCS) were significant only for categorical position trials and coordinate position trials, but not for the other three trial types for either degree changed condition. See Table 6 for a depiction of these correlations.

Because the position trial type showed a significant relationship with SIU scores, it was further investigated. Within the imagery video condition, control participants showed a significant positive relationship between RCS on categorical position trials and SIU ($r(62) = .379$, $p = .002$). Within the color video condition, control participants did not show a significant relationship between RCS on categorical position trials and SIU ($r(54) = .137$, $p = .315$). This pattern is evidence that watching the imagery video assisted control participants in better calibrating their subjective estimates of their use of mental imagery during the task. In contrast, VVIQ scores did not correlate with RCS for the categorical position trials for the imagery video group ($r(62) = .087$, $p = .494$) or for the color perception video group ($r(54) = .191$, $p = .158$), showcasing again that VVIQ scores missed the video manipulation effect because they were a poor subjective measurement for the purposes of comparing to performance on cognitive tasks.

Within the imagery video condition, control participants did not show a significant relationship between RCS on coordinate position trials and SIU ($r(62) = .151$, $p = .232$). Within the color video condition, control participants did not show a significant relationship between RCS on coordinate position trials and SIU ($r(54) = -.003$, $p = .983$). Similarly, VVIQ scores did not correlate with RCS for the coordinate position trials for the imagery video group ($r(62) = .016$, $p = .899$) or for the color perception video group ($r(54) = -.020$, $p = .884$).

Table 6. *Intercorrelations of SIU Scores and Rate Correct Scores for each Trial Type*

Variable	1	2	3	4	5	6	7	8	9	10
1. SIU	—	.607†	.069	-.008	.003	.103	.046	.075	.151*	.157*
2. VVIQ		—	.027	.026	-.003	.073	-.038	.080	.017	.129
3. ColorCat			—	.485†	.168*	.059	.198†	.097	.291†	.118
4. ColorCoord				—	.046	.115	.157*	.127	.328†	.248†
5. FormCat					—	.324†	.248†	.097	.238†	.041
6. FormCoord						—	-.015	.059	.127	.058
7. OrientCat							—	.285†	.400†	.080
8. OrientCoord								—	.220†	.079
9. PositCat									—	.334†
10. PositCoord										—

Note. These are partial correlations controlling for age and sex. All participants from Study 4 are depicted here ($N = 214$). The performance measure used here was rate correct score (RCS). SIU = Self-reported imagery usage for Study 4. VVIQ = Vividness of Visual Imagery Questionnaire. ColorCat = Categorical color trials. ColorCoord = Coordinate color trials. FormCat = Categorical form trials. FormCoord = Coordinate form trials. OrientCat = Categorical orientation trials. OrientCoord = Coordinate orientation trials. PositCat = Categorical position trials. PositCoord = Coordinate position trials. $^*p < .05$. $^†p < .01$

Visual Versus Spatial Task Performance

Before collecting the data for Study 4, I had intended to compare the results of the mostly-visual trials (color and form changes) to the results of the mostly-spatial trials (orientation and position changes). However, binning the trials in this fashion is less useful than simply reporting the four types of trials, especially considering that doing so would dilute effects and does not have a strong theoretical basis for why one task should count as spatial while another should count as visual. Thus, I have reported the four trial types separately, and as can be seen above, the pattern of results for each is discrete.

Dream Imagery

Similar to Study 3, results were also analyzed after splitting the aphantasia group into two groups based on the frequency of their dreams with imagery content (see Study 1). However, these additional analyses did not reveal any additional findings worth discussing in the main body of this study, and so discussion of these have been relegated to the Supplemental Analyses.

Study 4 Discussion

Overall, the results of Study 4 provided some information about what kinds of stimuli may be appropriate for future aphantasia experiments, as well as some evidence for the perceptual/conceptual theory, and some possible evidence for the implicit/explicit theory. The video manipulation within the control group also played a role in the results of Study 4.

Evaluating the Trial Types from Study 4

Of the four trial types that were used for Study 4, two of them ended up working well, one of them was so-so, and one of them was poor. The measure of "working well" is whether there was a main effect of the degree changed (categorical vs. coordinate) manipulation, because one would expect the coordinate trials to be more difficult than the categorical trials, but these specific stimuli have never been used in a study before. Thus, the trial types must be evaluated.

The two trial types that worked best were color and position, wherein the categorical/coordinate manipulation was clearly effective as indicated by a main effect of "degree changed" for all three dependent measures of accuracy, response time, and RCS. Form trials, on the other hand, showed a main effect of degree changed for response time, but not accuracy. This result appears to have been driven by a floor effect in the coordinate form trials, where accuracy was extremely low and not significantly different from chance responding. Form trials remain somewhat valid, however, as when RCS was the dependent measure, the main effect of degree changed is significant. Because RCS is a linear combination of accuracy and response time, it captures more variance in a single dependent measure than either accuracy or response times alone, and is therefore ideal for determining whether the categorical/coordinate manipulation "worked." Thus, the manipulations of color, position, and form appeared to perform as expected, although perhaps the form trials were too difficult for the participants, at times. In contrast, the main effect of degree changed was *never* significant for the orientation trials: not for accuracy, response time, or RCS. That is, the categorical/coordinate manipulation of orientation appeared to have been too subtle for drawing firm conclusions in this study. Thus, the orientation trials will be avoided in the discussion below.

Support for Perceptual/Conceptual Theory from Study 4

Turning our attention to the theories of aphantasia, the perceptual/conceptual theory was supported by the existence of group x degree changed interactions, such that the difference in performance measures between the categorical and coordinate trials (the slopes) were at times significantly different between the aphantasia and control groups.

Significant group x degree changed interactions showed up in four specific places: for the overall group x degree changed interaction for accuracy collapsed across all four trial types (see Figure 31), for the dependent measure of accuracy on position trials (see Figure 35), for the dependent measure of response times on form trials (see Figure 39), and for the dependent measure of rate correct score on position trials (see bottom right of Figure 42). In all four cases, the relationship was in the direction that conceptual/perceptual theory would predict, such that participants with aphantasia had a steeper slope between the categorical and coordinate conditions than did the control group.

The clearest example of this interaction, and the strongest evidence for perceptual/conceptual theory, comes from the RCS measure for position trials. Rate correct score (RCS) is best conceptualized as “correct responses per second.” As can be seen in the lower right panel of Figure 42, there is a “crossed” interaction between the aphantasia and control groups when it comes to correct responses per second. This difference in slope between the two groups, which accounts for both accuracy and response time, is significant. This pattern indicates that whatever strategy the aphantasia group was using, it was different from the control group, and the strategy was less able to handle the coordinate version of the task relative to the categorical version.

Also, for all three significant group x degree changed interactions, the main effect of “degree changed” was significant (see Figures 35, 39, and bottom right of 42), such that for both groups, the coordinate condition was more difficult than the categorical condition. That is, coordinate trials were more difficult than categorical trials for all participants, as would be expected, as the degree changed manipulation was intended to create more cognitively demanding trials. However, crucially for perceptual/conceptual theory, the increase in difficulty from categorical to coordinate was even steeper for participants with aphantasia than it was for the control group!

The interaction patterns suggest that the perceptual/conceptual theory is correct in its prediction that the participants in the aphantasia group were using propositional strategies rather than depictive strategies to encode information about form and position during the change detection task used in Study 4 (see Figure 43). This is because propositional strategies (e.g., remembering “BLUE OVAL ABOVE RED BOX NEXT TO PURPLE OVAL”) allow for faster and more accurate detection of categorical changes (e.g., moving the red box so that the blue oval is now BELOW the red box, a categorical position change, is mismatched to the encoded proposition and is therefore easier to identify as the change) when compared to attempts to detect coordinate changes (e.g., moving the red box farther away from the blue oval, a coordinate position change, is not mismatched to the proposition, and is therefore more difficult to identify as the change). The pattern of the results from Study 4 also fits with the idea that the control participants were using a more depictive strategy on average. A depictive strategy that makes use of mental imagery should have more power to detect coordinate changes relative to a propositional strategy due to having an easier time finding the mismatch in the encoded mental image. That is, comparing two pictures that preserve relative distance between the shapes will allow the identification of coordinate changes more often that comparing a proposition that did not preserve relative distance to the newly presented picture. For both strategies, the coordinate change is more difficult to detect than the categorical change, but this difficulty is greater for the propositional strategy. This relationship between strategies and categorical/coordinate changes fits the relationship shown in the bottom right panel of Figure 42, assuming that the aphantasia group was more likely to use a propositional strategy and the control group was more likely to use a depictive strategy.

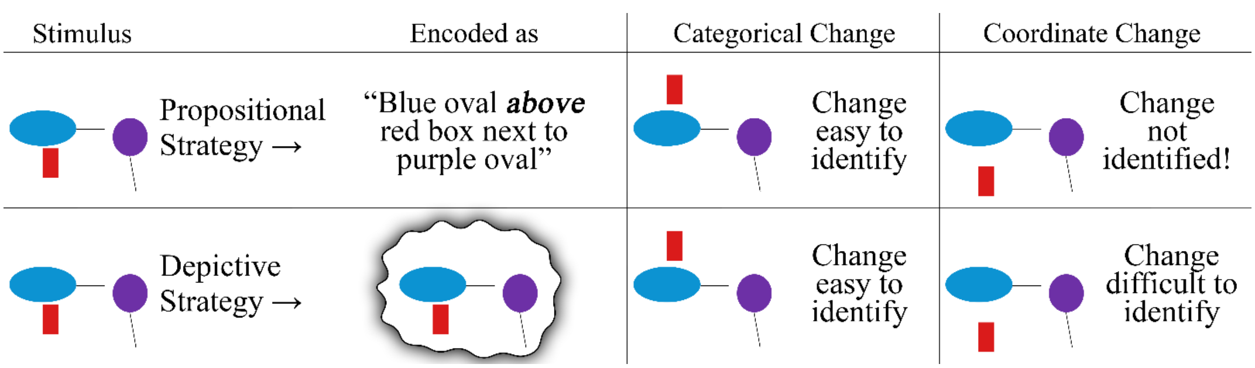


Figure 43. Theoretical results of different encoding strategies on a propositional trial of the change identification task.

Additionally, self-reported imagery use (SIU) on Study 4 supports perceptual/conceptual theory, albeit on a weaker basis, because there was a highly significant difference between how participants in the aphantasia group and participants in the control group responded to the question about how much imagery they used during the task. The aphantasia group reported rock-bottom imagery use, averaging a score of 1.12 on a scale that ranged from 1 to 4. The control group, on the other hand, averaged 2.57, slightly above the midpoint of the scale, indicating more use of imagery. Those SIU scores correlated significantly and positively with rate correct scores on the categorical position trials and with rate correct scores on the coordinate position trials (see Table 5). Granted, SIU is a subjective measurement, but it is an important detail: whether or not they are correct about what their brain is actually computing, the people in the aphantasia group at the very least *believe* that they are not making use of imagery during a task explicitly designed to benefit from the use of imagery. Their actual performance reflected that belief.

Support for Implicit/Explicit Theory from Study 4

Implicit/explicit theory was tenuously supported by the lack of significant group x degree changed interactions in the following seven situations: accuracy on color trials (see Figure 32), accuracy on form trials (see Figure 33), response time on color trials (see Figure 38), response time on position trials (see Figure 41), rate correct score on color trials (see upper left of Figure 42), and rate correct score on form trials (see upper right of Figure 42). With the exception of “accuracy on form trials” where performance appeared to hit floor for coordinate trials (see Figure 33), the main effect of degree changed (categorical versus coordinate trials) was significant for these measurements, indicating that the degree changed manipulation was impactful. So then, if the manipulation worked as intended, why didn’t the supposedly different cognitive strategies of the aphantasia group (propositional strategy) and the control group (depictive strategy) show up as interactions such that the slopes for the two groups had different shapes between categorical and coordinate trials, as perceptual/conceptual theory would predict?

The implicit/explicit theory would postulate that the reason no interactions showed up for those investigations is that the aphantasia group was using a depictive strategy, even though they were unaware of it. That would explain why, in many cases, there wasn't a group difference for the slopes between categorical and coordinate trials. The clearest argument for this interpretation comes from accuracy, response times, and RCS for color trials (See Figures 32, 38, and upper left of Figure 42). As mentioned, rate correct score (RCS) captures variance in both response times and accuracy because it is a linear combination of the two. If there was an interaction between group and degree changed, you would expect it to show up *somewhere* between these three dependent measures. And yet, there is no hint of a group x degree changed interaction for color trials: the aphantasia group and the control group were both equally affected by the degree changed manipulation, as you would expect if both groups were using the same cognitive strategy to complete the task.

Implicit/explicit theory would also postulate that the group difference in self-reported imagery use (SIU) scores were due to experiencing or not experiencing explicit imagery, but that imagery had no bearing on the implicit imagery that was actually used to solve the change identification task for all participants. Therefore, the subjective nature of the SIU scores makes them unreliable as evidence for either theory, because either theory can plausibly explain that group difference.

Position Trials Versus Color Trials

Because the position trials and the color trials always performed as anticipated when it came to the main effect of degree changed (i.e., for accuracy, response times, and RCS), it makes sense to use these two trial types to examine the theories of aphantasia. Unfortunately, position trials tell one story (i.e., they support perceptual/conceptual theory and refute implicit/explicit theory), while color trials tell a different story (i.e., they fit with implicit/explicit theory). Which is stronger evidence?

The position trials are stronger evidence because they refute implicit/explicit theory in favor of perceptual/conceptual theory. That is, implicit/explicit theory is falsified if the interaction effects detected in Study 4, as depicted in Figures 35, 39, and the lower right panel of Figure 42, are real effects; the interaction effects cannot be explained by implicit/explicit theory. The color trials are weaker evidence because we must use null findings (i.e., the fact that interactions were *not* detected) in order to support the implicit/explicit theory. Null findings are never as strong as are significant findings when using null hypothesis significance testing. That is, perceptual/conceptual theory could also plausibly explain the null findings, simply as a limitation of power due to sample size or sub-optimal study design (e.g., observed power for many examined relationships was quite low due to smaller-than-expected effect sizes). Thus, if only one theory is the correct explanation for developmental aphantasia, then perceptual/conceptual theory is best supported by the findings of Study 4.

Impact of Imagery Video Manipulation

Like Study 3, the results of Study 4 make a compelling case for the need of an instructional video explaining individual differences in mental imagery ability before administering imagery-related measures to control participants who may be naïve to how their mental imagery compares to the mental imagery of other people.

The video manipulation significantly affected SIU scores for Study 4, and one measure of performance (RCS on categorical position trials) correlated significantly with SIU scores only for control participants who watched the imagery video, but not for control participants who watched the irrelevant color perception video. This result is evidence that watching the mental imagery video helped the control participants calibrate themselves such that they better understood mental imagery and whether or not they were making use of it during the change identification task. Paralleling Studies 2 and 3, VVIQ scores did not detect such a difference, again showcasing the weakness of VVIQ scores as a subjective measurement of mental imagery when it comes to whether or not that imagery is used or is useful (e.g., related to accuracy) on a cognitive task.

CHAPTER 6. GENERAL CONCLUSION

These findings extend our knowledge about developmental aphantasia. Most importantly, this series of studies found differences in performance on cognitive tasks between participants with aphantasia and control participants. These differences were relative to the manipulated internal demands of the tasks (i.e., they were interaction effects), indicating that participants with aphantasia and the control group were, at least on average, using different cognitive strategies to complete the cognitive tasks, and that the fit of those strategies depended on the demands of the task (e.g., task complexity). That is, significant results were in the direction that would be expected if participants with aphantasia *were not* using depictive mental imagery—as is claimed by people with aphantasia—and instead were using propositional representations (e.g., encoding a phrase such as "GREEN OVAL ABOVE ORANGE RECTANGLE"), while the control group *was* using depictive mental imagery (i.e., encoding an image). This result is encouraging for explaining developmental aphantasia as a valid phenomenon, as both the subjective and objective measures in the latter two studies aligned as would be expected, such that participants with aphantasia reported not using as much mental imagery as the control group, and the performance measures reflected those subjective reports. However, this series of studies also raised some points of caution concerning the current state of aphantasia research: namely, overreliance on VVIQ scores from participants naïve about individual differences in mental imagery ability.

The VVIQ Problem and a Potential Video Solution

Study 1 noted that there were group differences between the aphantasia group and the control group. Novel differences were detected on an exploratory basis, including the findings that participants with aphantasia appeared to be less likely to identify as cisgender, less likely to experience enjoyment when reading fictional stories, and less likely to be religious. Because my sample was not matched (e.g., control and aphantasia groups were recruited in different ways), these exploratory differences should be replicated. Agreeing with the findings of some recent survey research, my sample of participants with aphantasia seem to have a decreased frequency of dreaming, and a reduced amount of imagery content within their dreams. These dream-related findings fit with the findings of Dawes et al. (2020).

Importantly, there was an experimental manipulation within the control group that had a significant impact. Participants in the control group were randomly assigned to either watch a short educational video about individual differences in mental imagery vividness, or they were randomly assigned to watch a short educational video about the irrelevant topic of individual differences in color perception. The average Vividness of Visual Imagery Questionnaire (VVIQ) score was significantly impacted by this manipulation, and so was a novel measure of total self-reported imagery use (TSIU) which added up the self-reported imagery use (SIU) scores for each participant from all five tasks used within these studies. That is, watching the imagery video produced a lower average VVIQ score and a lower average TSIU score relative to watching the color perception video. Participants, after learning about what mental imagery may look like for different people, reduced their estimates concerning their own mental imagery, both when it came to reporting their general ability to experience imagery vividness (VVIQ), and when reporting whether or not they made use of mental imagery during the cognitive tasks (TSIU).

VVIQ scores correlated strongly and positively with TSIU scores. VVIQ scores also correlated strongly and positively with each of the five SIU scores for each individual task. This is somewhat compelling evidence that the beliefs of participants about the limits of their mental imagery vividness are aligned with their beliefs about the amount of imagery that they “used” (experienced) during each of the five tasks. However, because both VVIQ and TSIU scores were pushed around by the video manipulation used within the control group, these self-reports may not accurately reflect the true amount of mental image generation that was occurring for each participant. The remaining studies investigated whether there was evidence that the control participants who watched the imagery video were better calibrated (e.g., between their SIU scores and performance measures) than were the control participants who watched the video about color perception. Findings from Studies 2, 3, and 4 indicated that those who watched the imagery video were indeed better calibrated in terms of the relationship between their beliefs about their imagery and their performance on tasks, at least for some cognitive tasks. But this was only true when SIU scores were considered, and not when VVIQ scores were considered. Thus, VVIQ scores don't seem to capture enough information about *when* participants make *use* of imagery to be useful in situations where that question is important, such as within this series of studies.

The issue of being unable to directly compare one's own imagery abilities to other people's imagery abilities may be partly responsible for the mismatches in the literature such that tasks that seem to depend on, or at least make use of, visual mental imagery, do not seem to correlate well with VVIQ scores (e.g., McAvinue & Robertson, 2007; Reisberg, 2013). If the VVIQ continues to be used in future studies, the instructions coupled with the measure must be improved in order to be certain that participants understand their own mental imagery vividness relative to other people. Or, perhaps it is time for the researchers in the sphere of aphantasia literature to develop their own measure specifically for aphantasia research, one with carefully crafted instructions and questions that leave no doubt in the mind of the participants as to the possible individual differences of mental imagery ability. Such a measure is particularly important considering the track record of the VVIQ as a measure that is easy to "push around" in order to produce different scores depending on how it is administered and the demands of the researcher (e.g., e.g., Ahsen, 1993; Di Vesta et al., 1971; Farah, 1988; J. T. E. Richardson, 1980; McKelvie, 1995; Winograd et al., 1998).

In this study, watching a video about individual differences in mental imagery helped participants better calibrate their own estimates of self-reported imagery use (SIU) with their actual "objective" performance on tasks that theoretically benefit from the use of mental imagery. To my knowledge, this is the first time that such an effect has been demonstrated, and it may be useful for other mental imagery researchers in the future.

A video may be the ideal form of delivering VVIQ instructions—or instructions for any subjective measure of mental imagery—to participants, when compared to having them read the instructions themselves or having the instructions read to them by an experimenter. With a video, the instructions are presented to each participant in an identical manner, preventing individual variations. With a video, the pace of the instruction presentation is controlled such that participants (or experimenters) cannot skip steps intentionally or accidentally. People in the 21st century are becoming increasingly used to have instructions delivered in video form, including on airplanes, workplace safety trainings, and in virtual classrooms. With a video, standardizing the administration of instructions and replicating research is made easier for future studies. My video files used for this series of studies are available for any researchers interested in using them or improving upon them (Toftness, 2021a; Toftness, 2021b).

Evidence for Theories of Aphantasia

There are three prominent theories of developmental aphantasia in the literature. The first, the perceptual/conceptual theory (Faw, 1997), posits that people with aphantasia use propositional strategies to accomplish cognitive tasks because they are not able to use depictive strategies that rely on mental imagery. The second, the implicit/explicit theory (e.g., Botez et al., 1985; McAvinue & Robertson, 2007), posits that people with aphantasia are still using mental images to complete cognitive tasks, but they are simply not consciously aware of the images, unlike people without aphantasia who are aware of the mental images. The third, the mistaken/malingering theory (e.g., Smith, 1978), posits that people who claim to have aphantasia still use mental images, but they are either mistaken about what mental images are and so believe that they are not using them, or are maliciously pretending to not use mental images—thus, mistaken/malingering theory is not really a theory explaining developmental aphantasia, because it posits that no such disorder exists. Overall, this series of studies found evidence for the perceptual/conceptual theory of aphantasia thanks to between-within interaction effects.

In Study 3, a three-way interaction between group, scanning distance, and number of squares revealed the relationship that would be expected if the perceptual/conceptual theory was correct. That is, the scanning distance manipulation affected the performance of the control group more than it affected the performance of the aphantasia group, while the squares manipulation affected the aphantasia group more than the control group (see Study 3 Discussion). In Study 4, the position trials featured an interaction between group and categorical/coordinate trials (i.e., the “degree changed” manipulation) such that the performance of the aphantasia group was more impacted by the degree changed manipulation than was the control group (see Study 4 Discussion). Those effects are consistent with the perceptual/conceptual theory’s claim that people with aphantasia only make use of propositional representations while people without aphantasia are free to make use of depictive representations.

In contrast, the implicit/explicit theory of aphantasia was not well supported by this series of studies because it was refuted by the existence of interaction effects in Studies 3 and 4. That is, if the underlying mental calculations of people with developmental aphantasia rely upon mental imagery, as the implicit/explicit theory posits, then we would not expect to be able to detect such interaction effects that indicate a difference in cognitive strategies between the groups. Meanwhile, the significant effect of scanning distance in Study 3 for both participants with aphantasia and control participants that was predicted by the implicit/explicit theory (e.g., it could potentially be explained by participants with aphantasia making use of mental images) could alternatively be explained as a consequence of the increase in task difficulty when the arrow was further away from the target square, due to the difficulty of identifying where the arrow was pointing in “far” trials relative to “near” trials. Thus, an effect of scanning distance alone is not enough to support the implicit/explicit theory, especially in the face of the evidence to the contrary thanks to the interaction effects that support the perceptual/conceptual theory.

These results cannot completely rule out the mistaken/malingering theory of aphantasia if the intent of the participants with aphantasia was to deceive. That is, we could go further down the rabbit hole and speculate that *all* participants are capable of using depictive strategies (mental imagery) and are also capable of using propositional strategies, and the significant differences resulted from the aphantasia group *choosing* not to make use of mental imagery in favor of using propositional representations, even when propositions were not ideal for the task at hand. It remains to be seen whether participants *instructed to pretend* to have no access to mental imagery would show a similar pattern of results as the aphantasia group did in Studies 3 and 4—but it would certainly make for an interesting study design!

It is also useful to note that main effects between the aphantasia group and the control group were not significant in this series of studies when considering rate correct score (which accounts for both accuracy and response time) as the dependent measure.

In Study 2, there were no main effects of performance (as measured by rate correct score) on three tasks when comparing the aphantasia group and the control group after controlling for age and sex. The backwards spelling task (BST), snowy pictures task (SPT), and the tail length task (TLT) thus do not appear to be useful as indicators of aphantasia moving forward. There was a slight relationship between SIU and accuracy for two of the tasks (SPT and TLT). Otherwise, it was not clear that performance on any of those three tasks benefitted from the use of mental imagery, although the control group did consistently report using significantly more mental imagery than did the aphantasia group as measured by SIU. Take the TLT for example. Participants with aphantasia overwhelmingly reported not using mental imagery, while control participants overwhelmingly reported using mental imagery (see Figure 12). However, after controlling for age and sex, there were no significant differences in accuracy, response time, or RCS. Perhaps using images versus propositions for certain cognitive tasks does not grant easily detectable advantages or disadvantages.

While it might seem straightforward to predict that people with aphantasia should perform faster on some cognitive tasks because they do not need to retrieve/generate a complex image, this view is overly simplistic. For example, it has been written that people seem to have longer “information retrieval times,” which can be measured as response time during a cognitive task, when they are using mental imagery versus when they are using other forms of mental representations (Kosslyn, 1976). According to this view, if people with aphantasia are not using mental images, perhaps we would expect their response times to be shorter than the control group. But as we saw in Study 3, the demands of the task matter! That is, the pattern of response times between the control group and the aphantasia group looked completely different depending on the scanning distance and the number of squares (see Figure 17). It is not a simple matter of predicting a main effect of group difference between people with aphantasia and the control group: the task matters. Just because a person is or is not using mental imagery doesn't mean that their performance on a task will be overall affected in a detectible way, as suggested by the null findings in Study 2 on the BST, SPT, and TLT. It is *inter-task manipulations* that are the key to detecting the differences between people who are using a more propositional approach (e.g., people with aphantasia) and people who are using a more depictive approach (e.g., people without aphantasia).

Task selection is important when attempting to investigate mental imagery. For any given task, we must ask ourselves (before drawing any conclusions): did this task make use of mental imagery in the first place? What is the typical pattern of results for people without aphantasia? We also have to look at both accuracy and response time, separately and together, in order to see the entire picture, because sacrificing accuracy can make response times faster and vice versa. Using a combined measure such as rate correct score can thus capture the variance from both accuracy and response time and show if there is any underlying significant effect between people with aphantasia and controls even when both measures are accounted for, such as was done in Study 4 with position trials. The mental imagery literature, as a whole, has not historically done this. It is my hope that modern-day aphantasia research will begin to incorporate such advice before attempting to administer cognitive tasks, and then wonder why they didn't find an effect or assume that accuracy and response times are unrelated to one another and should only be examined separately.

Cognitive research on developmental aphantasia is becoming more common. Most notably, a recent study by Pounder et al. (2022) found that on a mental rotation task that used three different degrees of rotated stimuli (40, 85, and 220 degrees), participants with aphantasia (when defined as a VVIQ score of less than 25) and control participants did not significantly differ in terms of the main effect of group on response times. Additionally, there was no interaction between group and degrees rotated. Unfortunately for those conclusions, their study only investigated accuracy and response times separately, and was limited in power considering a small sample size of 20 participants per group. I obtained the dataset and ran an analysis on rate correct score using the formula from Woltz & Was (2006; see Figure 53), and found that the reported lack of a main effect was almost certainly a Type II error (F (1, 38) = 3.36, p = .075, η_p^2 = .081, observed power = .431), while the attempt to detect an interaction effect was extremely underpowered (F (1.495, 56.795) = .424, p = .598, η_p^2 = .011, observed power = .107).

My conclusions agree with Pounder et al., in the sense that, in general, there do not appear to be main effects of performance between the aphantasia and control groups on imagery-related cognitive tasks (although there were main effects of accuracy for some of my tasks). However, my studies looked deeper, and revealed significant interactions based on within-task manipulations that changed how appropriate propositional or depictive strategies were for that given task. Perhaps the general lack of main effects is an encouraging finding, in the sense that people with aphantasia can complete cognitive tasks indistinguishably from people without aphantasia, even though they appear to be using a different encoding and recall strategy that avoids generating and using mental imagery to solve such tasks! Perhaps aphantasia is not a very disruptive condition to a person's cognitive life, at least as measured by cognitive tasks. Thus, cognitive strategies of various kinds seem to be useful, just as James (1890a) described different cognitive strategies as different paths to the same destination.

Limitations

In terms of limitations, it could be argued that the aphantasia group was more motivated to complete the cognitive tasks than was the control group. For example, in Study 4, before controlling for age and sex, the accuracy of the aphantasia group was significantly higher than the control group, and the response times of the aphantasia group were significantly higher than the control group. Such a result may be interpreted as the aphantasia group “trying harder” at the task, or as a speed/accuracy trade-off. Indeed, differences in motivation to try at the task may have impacted the main effects of performance between groups. Fortunately, testing the theories of aphantasia depended upon interaction effects based on inter-task manipulations of difficulty (Studies 3 and 4), and testing the impact of watching a video about mental imagery was done using random assignment within the control group (Study 1). Therefore, a difference in motivation would not have impacted those effects. Rate correct score (RCS) was also used to combine accuracy and response times in order to create a “correct responses per second” dependent measure that made trading of response time for accuracy irrelevant. Neuropsychological research using pre-existing group differences is, by definition, quasi-experimental, and so arguments against such possible differences in motivation (or other factors) are impossible to completely counter. Hopefully this series of studies was convincing despite the built-in limitations of being unable to randomly assign people to either have or not have aphantasia.

It was a limitation that this series of studies did not explicitly ask participants to report the nature of their cognitive strategies beyond whether their selected strategy featured the use of mental images. For example, I did not ask them to report the nature of their propositional attempts to represent stimuli in Studies 3 or 4. Advice on how to measure mnemonic strategies of people with aphantasia has been briefly discussed in the literature (J. Pearson & Keogh, 2019), but there are not yet clearly established ways of administering such investigations into the nature of the selected cognitive strategies used by participants with aphantasia. J. Pearson & Keogh (2019) suggest irrelevant retinal stimulation, transcranial magnetic stimulation, instructing participants to intentionally use different strategies, and other techniques such as EEG, fMRI, and neuromodulation. I would suggest the possibility of using a protocol analysis (think aloud) paradigm rather than asking the participant to report their strategy following the cognitive task, because strategies used may change throughout the task.

Additionally, a possible limitation was that the VVIQ was always administered *before* the administration of the cognitive tasks. This may have made a difference in SIU scores or in task performance, as thinking about mental imagery before a task that might benefit from the use of mental imagery may bias participants to either 1) change their strategy in terms of how much mental imagery to use, 2) change their self-reported amount of mental imagery use because they focused on it differently than they would have if it had not been previously mentioned (e.g., they became more or less conscious of their use of it), or 3) experimenter demands influenced their willingness to report the use of imagery. In order to determine if administering the VVIQ before the cognitive tasks affected SIU scores or task performance, a manipulation would need to be added to future studies in which the VVIQ (or an improved questionnaire) is randomly placed before or after the bulk of the study. It is true that McKelvie (1995, p. 27) argued that placing the VVIQ at the end of experimental tasks may bias participants to respond more leniently and to report more vivid imagery, but this has not been tested with SIU scores or my specific cognitive tasks used in Studies 3 and 4 which showed the interaction effects between the control group and the aphantasia group.

Because these studies were conducted during the COVID-19 pandemic, the research was conducted entirely online, even for participants at the same academic institution where I reside. The geographic distribution of the participants with aphantasia was also wide, and a further reason for the remote study. Thus, no researchers were monitoring the participants as they completed the tasks. Even though participants were instructed to complete the cognitive tasks in a single series with no other distractions (e.g., such as people talking to them, browsing the Internet, listening to music, etc.), I cannot guarantee that such instructions were followed. I did exclude participants with extremely long response times to stimuli during criterion phases of the cognitive tasks, and there was evidence of excluded participants responding unnaturally quickly or slowly to the tasks suggesting distraction, but this does not rule out multitasking entirely. I also did not ask participants to report whether they followed the instructions of the online aphantasia study (e.g., Königsmark et al., 2021). Thus, future research should more explicitly monitor participants.

Conclusion

While developmental aphantasia has been known about within the fringes of cognitive psychology since at least the 1980s, and acquired aphantasia has been known about since at least the 1880s, it is only just recently that the widespread nature of the disorder has been acknowledged by the literature thanks to renewed scientific investigations and popular media (Zeman et al., 2010; Zeman et al., 2015; Zimmer, 2010). Getting enough participants for a high-powered study of developmental aphantasia was not possible until recently, and so investigations of performance on cognitive tasks such as those contained in this set of studies have been slow going. Hopefully, these studies serve as a groundbreaking in a few key ways.

First, these studies should serve as a cautionary tale of overreliance on VVIQ scores, and prompt the creation of a new measure specifically tailored for aphantasia (and not retrofitted from decades-old research; Marks, 1973). Second, these studies should raise the question of whether participants are too naïve about their own mental imagery abilities relative to the abilities of others to accurately judge their own mental imagery in the first place, such as required by any subjective measurement of mental imagery (e.g., J. T. E. Richardson, 1980). More education about individual differences in mental imagery, namely watching a video about imagery as was done in this series of studies, could be key to obtaining accurately calibrated estimates of imagery. After all, many people with aphantasia do not realize that the way that they imagine is different from the average person until they are already adults—we should not assume that people have an inherent grasp on what their mind is like relative to the minds of others. Finally, these studies demonstrate clear evidence that the cognitive strategies used by people with aphantasia differ from the cognitive strategies used by the average person, as measured by performance on cognitive tasks.

The mental imagery literature has shown time and again that people are not readily willing to accept that the inner workings of one’s mind can be so very different from the inner workings of another’s mind (e.g., Dennet, 1978; Reisberg et al., 2003). Incredulity, doubt, and downright accusations of fabrication riddle the philosophical and psychological history of individual differences in mental imagery ability (e.g., Smith, 1978). Hopefully, this series of studies will help put a few of those ghosts to rest by demonstrating that aphantasia is a measurable phenomenon experienced by many people.

REFERENCES

Ajina, S., & Bridge, H. (2017). Blindsight and unconscious vision: What they teach us about the human visual system. *The Neuroscientist*, *23*(5), 529–541. https://doi.org/gb2wr4

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & De Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology, 23*(15), 1427–1431. https://doi.org/f46xm5

Amedi, A., Malach, R., & Pascual-Leone, A. (2005). Negative BOLD differentiates visual imagery and perception. *Neuron, 48*(5), 859–872. https://doi.org/dc3rpt

Angell, J. R. (1910). Methods for the determination of mental imagery. *The Psychological Monographs*, 13(53), 61–107.

Arnold, L. (2018). *Training reduces the drawing detriments associated with categorical perception* [Doctoral dissertation, Iowa State University]. ProQuest Dissertations & Theses Global.

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

Bartolomeo, P. (2002). The relationship between visual perception and visual mental imagery: A reappraisal of the neuropsychological evidence. *Cortex, 38*(3), 357–378. https://doi.org/d5892j

Bartolomeo, P., Bachoud-Lévi, A. C., De Gelder, B., Denes, G., Dalla Barba, G., Brugières, P., & Degos, J. D. (1998). Multiple-domain dissociation between impaired visual perception and preserved mental imagery in a patient with bilateral extrastriate lesions. *Neuropsychologia, 36*(3), 239–249. https://doi.org/bfsmpz

Basso, A., Bisiach, E., & Luzzatti, C. (1980). Loss of mental imagery: A case study. *Neuropsychologia, 18*(4–5), 435–442. https://doi.org/dfzd2j

Behrmann, M., Moscovitch, M., & Winocur, G. (1994). Intact visual imagery and impaired visual perception in a patient with visual agnosia. *Journal of Experimental Psychology: Human Perception and Performance, 20*(5), 1068–1087. https://doi.org/d7rd9r

Berger, G. H., & Gaunitz, S. C. (1977). Self-rated imagery and vividness of task pictures in relation to visual memory. *British Journal of Psychology, 68*(3), 283 –288. https://doi.org/dj6rjn

Bergmann, J., Genç, E., Kohler, A., Singer, W., & Pearson, J. (2016). Smaller primary visual cortex is associated with stronger, but less precise mental imagery. *Cerebral Cortex, 26*(9), 3838–3850. https://doi.org/g485

Betts, G. H. (1909). *The distribution and functions of mental imagery*. Teachers College Press.

Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex, 14*(1), 129–133. https://doi.org/ggbwbw

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology, 20*(2), 239–263. https://doi.org/bfk6x9

Blazhenkova, O. (2016). Vividness of object and spatial imagery. *Perceptual and Motor Skills, 122*(2), 490–508. https://doi.org/g486

Blazhenkova, O., & Pechenkova, E. (2019). The two eyes of the blind mind: Object vs. spatial aphantasia? *The Russian Journal of Cognitive Science, 6*(4), 51–62. https://doi.org/gs43

Borst, G., Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2012). Representations in mental imagery and working memory: Evidence from different types of visual masks. *Memory & Cognition, 40*(2), 204–217. https://doi.org/b9fvtv

Borst, G., & Kosslyn, S. M. (2010). Individual differences in spatial mental imagery. *Quarterly Journal of Experimental Psychology*, *63*(10), 2031–2050. https://doi.org/ckcwr3

Botez, M. I., Olivier, M., Vézina, J. L., Botez, T., & Kaufman, B. (1985). Defective revisualization: Dissociation between cognitive and imagistic thought case report and short review of the literature. *Cortex*, *21*(3), 375–389. https://doi.org/g487

Brain, R. (1954). Loss of visualization. *Proceedings of the Royal Society of Medicine*, *47*(4), 288–290.

Bridge, H., Harrold, S., Holmes, E. A., Stokes, M., & Kennard, C. (2012). Vivid visual mental imagery in the absence of the primary visual cortex. *Journal of Neurology, 259*(6), 1062–1070. https://doi.org/dc4pgq

Brooks, B. E., & Cooper, E. E. (2006). What types of visual recognition tasks are mediated by the neural subsystem that subserves face recognition? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 684–698. https://doi.org/dfqcvz

Brown, J. W. (1972). *Aphasia, apraxia, and agnosia: Clinical and theoretical aspects*. CC Thomas.

Bulmer, M. G. (1979). *Principles of statistics*. Courier Corporation.

Burton, L. J., & Fogarty, G. J. (2003). The factor structure of visual imagery and spatial abilities. *Intelligence, 31*(3), 289–318. https://doi.org/c3pztj

Cairns, E., & Coll, P. (1977). The role of visual imagery in visual, tactual and cross-modal matching. *British Journal of Psychology*, *68*(2), 213–214. https://doi.org/dxhfq4

Campos, A., & Pérez-Fabello, M. J. (2009). Psychometric quality of a revised version vividness of visual imagery questionnaire. *Perceptual and Motor Skills, 108*(3), 798–802. https://doi.org/cxvhkf

Chatterjee, A., & Southwood, M. H. (1995). Cortical blindness and visual imagery. *Neurology, 45*(12), 2189–2195. https://doi.org/g488

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., Bookheimer, S. Y., Rosen, B. R., & Belliveau, J. W. (1996). Changes in cortical activity during mental rotation: A mapping study using functional MRI. *Brain, 119*(1), 89–100. https://doi.org/fvhcwz

Coltheart, M., Hull, E., & Slater, D. (1975). Sex differences in imagery and reading. *Nature, 253*(5491), 438–440. https://doi.org/bvdb8x

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75-176). Academic Press. https://doi.org/g489

Crowder, A. (2018). *Differences in spatial visualization ability and vividness of spatial imagery between people with and without aphantasia* [Doctoral dissertation, Virginia Commonwealth University]. ProQuest Dissertations & Theses Global. https://doi.org/g49b

Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Research, 47*(4), 474–478. https://doi.org/cmbxcj

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*(1), Article 10022. https://doi.org/gg3fkf

Dean, G. M., & Morris, P. E. (2003). The relationship between self-reports of imagery and spatial ability. *British Journal of Psychology, 94*(2), 245–273. https://doi.org/dgrxzm

de Gelder, B., Tamietto, M., Pegna, A. J., & Van den Stock, J. (2015). Visual imagery influences brain responses to visual stimulation in bilateral cortical blindness. *Cortex, 72*, 15–26. https://doi.org/f7zscr

Devinsky, O., & Lai, G. (2008). Spirituality and religion in epilepsy. *Epilepsy and Behavior, 12*(4), 636–643. https://doi.org/c568hj

de Vreese, L. P. (1991). Two systems for colour-naming defects: Verbal disconnection vs colour imagery disorder. *Neuropsychologia*, *29*(1), 1–18. https://doi.org/dq4mgf

Diedenhofen, B. & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE, 10*(4), Article e0121945. https://doi.org/f7df6t

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual areas. *Journal of Neuroscience, 37*(5), 1367–1373. https://doi.org/f9q7rd

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. (2019). Shared neural mechanisms of visual perception and imagery. *Trends in Cognitive Sciences, 23*(5), 423–434. https://doi.org/gg7jtw

Dobson, M., & Markham, R. (1993). Imagery ability and source monitoring: Implications for eyewitness memory. *British Journal of Psychology, 84*(1), 111–118. https://doi.org/fm22mb

Dror, I. E., & Kosslyn, S. M. (1994). Mental imagery and aging. *Psychology and Aging, 9*(1), 90–102. https://doi.org/d6nhx3

Dror, I. E., Kosslyn, S. M., & Waag, W. L. (1993). Visual-spatial abilities of pilots. *Journal of Applied Psychology, 78*(5), 763–773. https://doi.org/dthb62

Eberman, C., & McKelvie, S. J. (2002). Vividness of visual imagery and source memory for audio and text. *Applied Cognitive Psychology, 16*(1), 87–95. https://doi.org/bgq7zr

Farah, M. J. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition, 18*(1–3), 245–272. https://doi.org/cb7rf2

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95*(3), 307–317. https://doi.org/djx8kc

Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology, 20*(4), 439–462. https://doi.org/bzms6f

Farah, M. J., Levine, D. N., & Calvanio, R. (1988). A case study of mental imagery deficit. *Brain and Cognition, 8*(2), 147–164. https://doi.org/fjhhgv

Farah, M. J., & Peronnet, F. (1989). Event-related potentials in the study of mental imagery. *Journal of Psychophysiology, 3*(2), 99–109.

Faul, F., Erdfelder, E., Buchner, A. and Lang, A.G. (2009) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149 –1160. https://doi.org/b22kn7

Faw, B. (1997). Outlining a brain model of mental imaging abilities. *Neuroscience & Biobehavioral Reviews, 21*(3), 283–288. https://doi.org/bgkpw2

Faw, B. (2009). Conflicting intuitions may be based on differing abilities: Evidence from mental imaging research. *Journal of Consciousness Studies*, *16*(4), 45–68.

Fechner, G. T. (1966). *Elements of psychophysics* (H. E. Adler, D. H. Howes, & E. G. Boring, Trans). Holt, Rinehart and Winston. (Original work published 1860)

Fernald, M. R. (1912). The diagnosis of mental imagery. *The Psychological Monographs, 14*(58), 1–169.

Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE edge.

Finke, R. A., Johnson, M. K., & Shyi, G. C. W. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory & Cognition*, *16*(2), 133–137. https://doi.org/b27cvg

Finke, R. A., & Pinker, S. (1983). Directional scanning of remembered visual patterns. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(3), 398–410. https://doi.org/brb3ts

French, J. W., Eustrom, R. B., & Price, L. A. (1963). *Manual for kit of reference tests for cognitive factors*. Educational Testing Services.

Fulford, J., Milton, F., Salas, D., Smith, A., Simler, A., Winlove, C., & Zeman, A. (2018). The neural correlates of visual imagery vividness–An fMRI study and literature review. *Cortex*, *105*, 26–40. https://doi.org/gd5mn2

Galton, F. (1880). Statistics of mental imagery. *Mind, 5*(19), 301–318.

Galton, F. (1883). *Inquiries into human faculty and its development*. Macmillan.

Ganis, G., Thompson, W. L., Mast, F. W., & Kosslyn, S. M. (2003). Visual imagery in cerebral visual dysfunction. *Neurologic Clinics*, *21*(3), 631–646. https://doi.org/fkpgkh

Gastinger, M. J., Tian, N., Horvath, T., & Marshak, D. W. (2006). Retinopetal axons in mammals: Emphasis on histamine and serotonin. *Current Eye Research, 31*(7-8), 655–667. https://doi.org/c4wcxj

Goldenberg, G. (1992). Loss of visual imagery and loss of visual knowledge—A case study. *Neuropsychologia, 30*(12), 1081–1099. https://doi.org/csgq22

Gordon, R. (1949). An investigation into some of the factors that favour the formation of stereotyped images. *British Journal of Psychology, 39*(3), 156–167. https://doi.org/bpvf55

Guo, P. J., Kim, J., & Rubin, R. (2014, March). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the first ACM conference on Learning@scale*, 41–50.

Gur, R. C., & Hilgard, E. R. (1975). Visual imagery and the discrimination of differences between altered pictures simultaneously and successively presented. *British Journal of Psychology, 66*(3), 341–345. https://doi.org/bcqfcs

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology, 82*(6), 903–918. https://doi.org/fhc3h6

Heuer, F., Fischman, D., & Reisberg, D. (1986). Why does vivid imagery hurt colour memory? *Canadian Journal of Psychology, 40*(2), 161–175. https://doi.org/dd8n28

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*(3), 199–236. https://doi.org/ctfsg5

Holt, R. R. (1972). On the nature and generality of mental imagery. In P. W. Sheehan (Ed.), *The function and nurture of imagery* (pp. 3–33). Academic Press.

Humphrey, M. E., & Zangwill, O. L. (1951). Cessation of dreaming after brain injury. *Journal of Neurology, Neurosurgery & Psychiatry, 14*(4), 322–325. https://doi.org/c2rqz8

Intraub, H., & Hoffman, J. E. (1992). Reading and visual memory: Remembering scenes that were never seen. *The American Journal of Psychology, 105*(1), 101–114. https://doi.org/d2phz4

Jacobs, C., Schwarzkopf, D. S., & Silvanto, J. (2018). Visual working memory performance in aphantasia. *Cortex, 105*, 61–73. https://doi.org/gd5njh

James, W. (1890a). *The principles of psychology, Vol. 1*. Henry Holt and Company.

James, W. (1890b). *The principles of psychology, Vol. 2*. Henry Holt and Company.

Johnson, M. K., Raye, C. L., Wang, A. Y., & Taylor, T. H. (1979). Fact and fantasy: The roles of accuracy and variability in confusing imaginations with perceptual experiences. *Journal of Experimental Psychology: Human Learning and Memory, 5*(3), 229–240. https://doi.org/b2ghnh

Kaczmarek, B. L. J., Code, C., & Wallesch, C. (2003). The fractionation of mental life: Luria’s study of Lieutenant Zasetsky. In C. Code, C. Wallesch, Y. Joanette, & A. R. Lecours (Eds.), *Classic cases in neuropsychology* (vol. 2, pp. 31–44). Psychology Press.

Katz, A. N. (1983/2014). What does it mean to be a high imager? In J. C. Yuille (Ed.), *Imagery, memory and cognition (PLE: memory): Essays in honor of Allan Paivio* (reprint edition, pp. 39–63).

Kaufmann, G. (1981). What is wrong with imagery questionnaires? *Scandinavian Journal of Psychology*, *22*(1), 59–64. https://doi.org/cdtdpw

Kendle, A. (2017). *Aphantasia: Experiences, perceptions, and insights*. Dark River.

Keogh, R., & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, *105*, 53–60. https://doi.org/gd5m7v

Keogh, R., & Pearson, J. (2021). Attention driven phantom vision: Measuring the sensory strength of attentional templates and their relation to visual mental imagery and aphantasia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1817). https://doi.org/ghqcrn

Kerr, M., & Pear, T. H. (1931). Unseen drama and imagery; Some experimental observations.
British Journal of Psychology, *22*(1), 43–54. https://doi.org/ccr55t

Kerr, N. H. (1983). The role of vision in "visual imagery" experiments: Evidence from the congenitally blind. *Journal of Experimental Psychology: General*, *112*(2), 265–277. https://doi.org/b275t2

Kerr, N. H., & Neisser, U. (1983). Mental images of concealed objects: New evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(2), 212–221. https://doi.org/dp2wvr

Königsmark, V. T., Bergmann, J., & Reeder, R. R. (2021). The Ganzflicker experience: High probability of seeing vivid and complex pseudo-hallucinations with imagery but not aphantasia. *Cortex*, *141*, 522–534. https://doi.org/gkb3h3

Kosslyn, S. M. (1976). Can imagery be distinguished from other forms of internal representation? Evidence from studies of information retrieval times. *Memory & Cognition, 4*(3), 291–297. https://doi.org/cqw6x2

Kosslyn, S. M. (1983). *Ghosts in the mind's machine: Creating and using images in the brain*. Norton.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. MIT Press.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4*(1), 47–60. https://doi.org/c8z6r3

Kosslyn, S. M., Brunn, J., Cave, K. R., & Wallach, R. W. (1984). Individual differences in mental imagery ability: A computational analysis. *Cognition, 18*(1-3), 195–243. https://doi.org/dghbpp

Kosslyn, S. M., & Jolicoeur, P. (1980). A theory-based approach to the study of individual differences in mental imagery. In R. E. Snow, P. Federico, & W. E. Montaque (Eds.), *Aptitude, learning, and instruction. Cognitive processes analyses* (pp. 139–175). Erlbaum.

Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Ganis, G., Sukel, K. E., & Alpert, N. M. (1999). The role of area 17 in visual imagery: Convergent evidence from PET and rTMS. *Science, 284*(5411), 167–170. https://doi.org/ddh9t4

Kosslyn, S. M., Thompson, W. L., & Alpert, N. M. (1997). Neural systems shared by visual imagery and visual perception: A positron emission tomography study. *Neuroimage, 6*(4), 320–334. https://doi.org/fbw3s3

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. Oxford University Press. https://doi.org/ct8kh5

Kosslyn, S. M., Thompson, W. L., Shephard, J. M., Ganis, G., Bell, D., Danovitch, J., Wittenberg, L. A., & Alpert, N. M. (2004). Brain rCBF and performance in visual imagery tasks: Common and distinct processes. *European Journal of Cognitive Psychology*, *16*(5), 696–716. https://doi.org/djgr9c

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/dgfsz3

Kunzendorf, R. (1984). Centrifugal effects of eidetic imaging on flash electroretinograms and autonomic responses. *Journal of Mental Imagery, 8*(4), 67–75.

Kunzendorf, R., & Hall, S. (2001). Electroretinographic after-effects of visual imaging: Individual differences in imagery vividness and reality testing. *Journal of Mental Imagery, 25*(3&4), 79–92.

Laeng, B., & Sulutvedt, U. (2014). The eye pupil adjusts to imaginary light. *Psychological Science, 25*(1), 188–197. https://doi.org/p9x

Lequerica, A., Rapport, L., Axelrod, B. N., Telmet, K., & Whitman, R. D. (2002). Subjective and objective assessment methods of mental imagery control: Construct validations of self-report measures. *Journal of Clinical and Experimental Neuropsychology, 24*(8), 1103–1116. https://doi.org/dzrg7t

Levine, D. N., Warach, J., & Farah, M. (1985). Two visual systems in mental imagery: Dissociation of “what” and “where” in imagery disorders due to bilateral posterior cerebral lesions. *Neurology, 35*(7), 1010–1018. https://doi.org/gtv6

Likert, R. & Quasha, W. H. (1970). *Revised Minnesota paper form board test*. The Psychological Corporation.

Luria, A. R. (1972). *The man with a shattered world: The history of a brain wound*. Basic Books.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*(1), 17–24. https://doi.org/b6v5wv

Marks, D. F. (1977). Imagery and consciousness: A theoretical review from an individual differences perspective. *Journal of Mental Imagery, 1*(2), 275–290.

Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery, 19*(3–4), 153–167.

Marks, D. F. (1999). Consciousness, mental imagery and action. *British Journal of Psychology, 90*(4), 567–585. https://doi.org/bfdnkz

Mazard, A., Tzourio-Mazoyer, N., Crivello, F., Mazoyer, B., & Mellet, E. (2004). A PET meta-analysis of object and spatial mental imagery. *European Journal of Cognitive Psychology, 16*(5), 673–695. https://doi.org/ckqcqr

McAvinue, L. P., & Robertson, I. H. (2007). Measuring visual imagery ability: A review. *Imagination, Cognition and Personality, 26*(3), 191–211. https://doi.org/bs38nw

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.

McKelvie, S. J., & Rohrberg, M. M. (1978). Individual differences in reported visual imagery and cognitive performance. *Perceptual and Motor Skills, 46*(2), 451–458. https://doi.org/cvnvrd

McLemore, C. W. (1976). Factorial validity of imagery measures. *Behaviour Research and Therapy, 14*(6), 399–408. https://doi.org/c8s7vr

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*(1), 40-48. https://doi.org/bn992m

Milton, F., Fulford, J., Dance, C., Gaddum, J., Heuerman-Williamson, B., Jones, K., Knight, K. F., MacKisack, M., Winlove, C., & Zeman, A. (2020). Behavioral and neural signatures of visual imagery vividness extremes: Aphantasia vs. hyperphantasia. *psyArchiv*. https://doi.org/g49d

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research, 6*(1), 57–77. https://doi.org/cbm64d

Monzel, M., Mitchell, D., Macpherson, F., Pearson, J., & Zeman, A. (2022). Aphantasia, dysikonesia, anauralia: Call for a single term for the lack of mental imagery – Commentary on Dance et al. (2021) and Hinwar and Lambert (2021). *Cortex*. https://doi.org/hjkw

Móró, L., Simon, K., Bárd, I., & Rácz, J. (2011). Voice of the psychonauts: Coping, life purpose, and spirituality in psychedelic drug users. *Journal of Psychoactive Drugs, 43*(3), 188–198. https://doi.org/d49j6t

Nanay, B. (2021). Unconscious mental imagery. *Philosophical Transactions of the Royal Society B, 376,* Article 20190689. https://doi.org/ghr4f5

Nielsen, J. (1946). *Agnosia, apraxia, aphasia: Their value in cerebral localization* (2nd ed.). Hoeber.

Ogden, J. A. (1993). Visual object agnosia, prosopagnosia, achromatopsia, loss of visual imagery, and autobiographical amnesia following recovery from cortical blindness: Case MH. *Neuropsychologia, 31*(6), 571–589. https://doi.org/bkf3k6

Paivio, A. (1971). *Imagery and verbal processes*. Holt, Rinehart & Winston.

Palombo, D. J., Sheldon, S., & Levine, B. (2018). Individual differences in autobiographical memory. *Trends in Cognitive Sciences*, *22*(7), 583–597. https://doi.org/gds8sh

Pearson, D. G., Deeprose, C., Wallace-Hadrill, S. M., Heyes, S. B., & Holmes, E. A. (2013). Assessing mental imagery in clinical psychology: A review of imagery measures and a guiding framework. *Clinical Psychology Review, 33*(1), 1–23. https://doi.org/f3pdzh

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/gf3c8x

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences, 112*(33), 10089–10092. https://doi.org/f7pcvd

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences, 19*(10), 590–602. https://doi.org/f3pzg6

Pearson, J., Rademaker, R. L., & Tong, F. (2011). Evaluating the mind’s eye: The metacognition of visual imagery. *Psychological Science*, *22*(12), 1535–1542. https://doi.org/d5z83q

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience, 20*(10), 624–634. https://doi.org/ghfj64

Perky, C. W. (1910). An experimental study of imagination. *The American Journal of Psychology, 21*(3), 422–452. https://doi.org/dt5n7d

Phillips, I. (2014). Lack of imagination: Individual differences in mental imagery and the significance of consciousness. In J. Kallestrup & M. Sprevak (Eds.), *New waves in philosophy of mind* (pp. 278–300). Palgrave Macmillan.

Podgorny, P., & Shepard, R. N. (1978). Functional representations common to visual perception and imagination. *Journal of Experimental Psychology: Human Perception and Performance, 4*(1), 21–35. https://doi.org/dk9qq7

Policardi, E., Perani, D., Zago, S., Grassi, F., Fazio, F., & Ladavas, E. (1996). Failure to evoke visual images in a case of long-lasting cortical blindness. *Neurocase, 2*(5), 381–394. https://doi.org/b7g3th

Poltrock, S. E., & Brown, P. (1984). Individual differences in visual imagery and spatial ability. *Intelligence, 8*(2), 93–138. https://doi.org/dnsrjs

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*. https://doi.org/hgnw

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80*(1), 1–24. https://doi.org/dbrvzx

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25*(2), 157–238. https://doi.org/dsk54w

Pylyshyn, Z. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences, 7*(3), 113–118. https://doi.org/dpqqpc

Rademaker, R. L., & Pearson, J. (2012). Training visual imagery: Improvements of metacognition, but not imagery strength. *Frontiers in Psychology, 3*(224). https://doi.org/ghqjhf

Reeder, R. [@kerblooee] (2022, January 12). *Validating a new mental imagery questionnaire against VVIQ, OSIQ, PSI-Q* [Thumbnail with link attached] [Tweet]. Twitter. https://twitter.com/kerblooee/status/1481209544588115970

Reisberg, D. (2013). Mental images. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology*, 374–387.

Reisberg, D., Culver, L. C., Heuer, F., & Fischman, D. (1986). Visual memory: When imagery vividness makes a difference. *Journal of Mental Imagery, 10*(4), 51–74.

Reisberg, D., & Leak, S. (1987). Visual imagery and memory for appearance: Does Clark Gable or George C. Scott have bushier eyebrows? *Canadian Journal of Psychology, 41*(4), 521–526. https://doi.org/b3v7b4

Reisberg, D., Pearson, D. G., & Kosslyn, S. M. (2003). Intuitions and introspections about imagery: The role of imagery experience in shaping an investigator's theoretical views. *Applied Cognitive Psychology, 17*(2), 147–160.

Richardson, A. (1969). *Mental imagery*. Springer.

Richardson, A. (1983). Imagery: Definition and types. In A. A. Sheikh (Ed.), *Imagery: Current theory, research and application* (pp. 3–41). Wiley.

Richardson, A. (1994). *Individual differences in imaging: Their measurement, origins, and consequences*. Routledge.

Richardson, J. T. E. (1980). *Mental imagery and human memory*. MacMillan Press.

Richardson, J. T. E. (1999). *Imagery*. Psychology Press, Ltd.

Riddoch, M. J. (1990). Loss of visual imagery: A generation deficit. *Cognitive Neuropsychology, 7*(4), 249–273. https://doi.org/fs8bx8

Riddoch, M. J., & Humphreys, G. W. (1987). A case of integrative visual agnosia. *Brain, 110*(6), 1431–1462. https://doi.org/c578x9

Riske, M. L., Wallace, B., & Allen, P. A. (2000). Imaging ability and eyewitness accuracy. *Journal of Mental Imagery*, *24*(1-2), 137–148.

Roeckelein, J. E. (2004). *Imagery in psychology: A reference guide*. Greenwood Publishing Group.

Roldan, S. M. (2017). Object recognition in mental representations: Directions for exploring diagnostic features through visual mental imagery. *Frontiers in Psychology, 8*(833). https://doi.org/f99hcz

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*(3), 169 – 188. https://doi.org/dp7dhk

Sacks, O. (2010). *The mind's eye*. Knopf.

Sacks, O. J., & Wasserman, R. (1987). The case of the colorblind painter. *The New York Review of Books, 34*, 25–33.

Schwitzgebel, E. (2011). *Perplexities of consciousness*. MIT press. https://doi.org/g49f

Schwitzgebel, E. (2002). How well do we know our own conscious experience? The case of visual imagery. *Journal of Consciousness Studies, 9*(5–6), 35–53.

Segal, S. J. (1971). Processing of the stimulus in imagery and perception. In S. J. Segal (Ed.), *Imagery* (pp. 69–100). Academic Press. https://doi.org/g49g

Segal, S. J., & Nathan, S. (1964). The Perky effect: Incorporation of an external stimulus into an imagery experience under placebo and control conditions. *Perceptual and Motor Skills, 18*(2), 385–395. https://doi.org/b8wqxd

Sheehan, P. W. (1967). A shortened form of Betts' questionnaire upon mental imagery. *Journal of Clinical Psychology*, *23*(3), 386–389. https://doi.org/fcfk5f

Sheehan, P. W., & Neisser, U. (1969). Some variables affecting the vividness of imagery in recall. *British Journal of Psychology, 60*(1), 71–80. https://doi.org/fgtw8r

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

Shuttleworth Jr, E. C., Syring, V., & Allen, N. (1982). Further observations on the nature of prosopagnosia. *Brain and Cognition, 1*(3), 307–322. https://doi.org/dqm6vn

Slee, J. A. (1988). Vividness as a descriptor and index of imagery. *Journal of Mental Imagery, 12*(3-4), 123–132.

Smith, B. (1978) *Memory*. Humanities Press.

Solms, M., Kaplan-Solms, K., & Brown, J. W. (1996). Wilbrand's case of "mind-blindness". In C. Code, C. Wallesch, Y. Joanette, & A. R. Lecours (Eds.), *Classic cases in neuropsychology* (vol. 1, pp. 89–110). Psychology Press.

Spalding, J. M. K., & Zangwill, O. L. (1950). Disturbance of number-form in a case of brain injury. *Journal of Neurology, Neurosurgery, and Psychiatry, 13*(1), 24–29. https://doi.org/dgtmdt

Stokes, D. (2019). Mental imagery and fiction. *Canadian Journal of Philosophy, 49*(6), 731–754. https://doi.org/g49j

Thorudottir, S., Sigurdardottir, H. M., Rice, G. E., Kerry, S. J., Robotham, R. J., Leff, A. P., & Starrfelt, R. (2020). The architect who lost the ability to imagine: The cerebral basis of visual imagery. *Brain Sciences, 10*(59), 1–15. https://doi.org/gmb7gr

Toftness, A. R. (2021a, October 4). Visual mental imagery [Video]. YouTube. https://youtu.be/U9Rqtwbr7Sk

Toftness, A. R. (2021b, October 4). Color perception qualia [Video]. YouTube. https://youtu.be/qcupdMuYYyY

Tomes, J. L., & Katz, A. N. (1997). Habitual susceptibility to misinformation and individual differences in eyewitness memory. *Applied Cognitive Psychology, 11*(3), 233–251. https://doi.org/ddt8kp

Tong, F. (2013). Imagery and visual working memory: One and the same? *Trends in Cognitive Sciences, 17*(10), 489–490. https://doi.org/gf3jd3

Tootell, R. B., Switkes, E., Silverman, M. S., & Hamilton, S. L. (1988). Functional anatomy of macaque striate cortex. II. Retinotopic organization. *Journal of Neuroscience, 8*(5), 1531–1568. https://doi.org/gfz33z

Trojano, L., & Grossi, D. (1994). A critical review of mental imagery defects. *Brain and Cognition, 24*(2), 213–243. https://doi.org/dx82wr

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods, 49*(2), 653 –673. https://doi.org/f96whp

Walczyk, J. J., & Taylor, R. W. (2000). Reverse-spelling, the VVIQ, and mental imagery. *Journal of Mental Imagery, 24*(1–2), 177–188.

Wantz, A. L., Borst, G., Mast, F. W., & Lobmaier, J. S. (2015). Colors in mind: A novel paradigm to investigate pure color imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(4), 1152–1161. https://doi.org/gk4n2g

Watkins, N. W. (2018). (A)phantasia and severely deficient autobiographical memory: Scientific and personal perspectives. *Cortex, 105*, 41–52. https://doi.org/gd5m58

Wicken, M., Keogh, R., & Pearson, J. (2021). The critical role of mental imagery in human emotion: insights from fear-based imagery and aphantasia. *Proceedings of the Royal Society B: Biological Sciences, 288*(1946), Article 20210267. https://doi.org/gjgf8t

Winch, W. H. (1908). The function of images. *The Journal of Philosophy, Psychology and Scientific Methods, 5*(13), 337–352. https://doi.org/fbzxxr

Winograd, E., Peluso, J. P., & Glover, T. A. (1998). Individual differences in susceptibility to memory illusions. *Applied Cognitive Psychology, 12*(7), S5–S27. https://doi.org/dtp8np

Woltz, D. J., & Was, C. A. (2006). Availability of related long-term memory during and after attention focus in working memory. *Memory and Cognition, 34*(3), 668–684. https://doi.org/bvh9r3

Woodworth, R. S. (1938). *Experimental psychology*. Holt.

Young, A. W., & van de Wal, C. (1996). Charcot's case of impaired imagery. In C. Code, C. Wallesch, Y. Joanette, & A. R. Lecours (Eds.), *Classic cases in neuropsychology* (vol. 1, pp. 31–44). Psychology Press.

Zago, S., Corti, S., Bersano, A., Baron, P., Conti, G., Ballabio, E., Lanfranconi, S., Cinnante, C., Costa, A., Cappellari, A., & Bresolin, N. (2010). A cortically blind patient with preserved visual imagery. *Cognitive and Behavioral Neurology, 23*(1), 44–48. https://doi.org/b3pd73

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

Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery—Congenital aphantasia. *Cortex*, *73*. 378–380. https://doi.org/gdf5hc

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, 130*, 426–440. https://doi.org/ghdxvf

APPENDIX A. VIVIDNESS OF VISUAL IMAGERY QUESTIONNAIRE

This version of the VVIQ was adapted from Marks (1973, 1995) and Crowder (2018).

INSTRUCTIONS

For each item on this questionnaire, try to form a visual image, and consider your experience carefully. For any image that you do experience, rate how vivid it is using the five-point scale described below. Please note that there are no right or wrong answers to the questions, and that it is not necessarily desirable to experience imagery or, if you do, to have more vivid imagery.

5 – Perfectly clear and vivid as real seeing

4 – Clear and reasonably vivid

3 – Moderately clear and lively

2 – Vague and dim

1 – No image at all, you only “know” that you are thinking of the object

For items 1-4, think of some relative or friend whom you frequently see (but who is not with you at present) and consider carefully the picture that comes before your mind’s eye.

1. The exact contour of face, head, shoulders, and body
2. Characteristic poses of head, attitudes of body, etc.
3. The precise carriage, length of step, etc., in walking
4. The different colors worn in some familiar clothes

Visualize a rising sun. Consider carefully the picture that comes before your mind’s eye.

5. The sun rising above the horizon into a hazy sky
6. The sky clears and surrounds the sun with blueness

7. Clouds. A storm blows up with flashes of lightning
8. A rainbow appears

Think of the front of a shop to which you often go. Consider the picture that comes before your mind’s eye.

9. The overall appearance of the shop from the opposite side of the road
10. A window display including colors, shapes, and details of individual items for sale
11. You are near the entrance. The color, shape, and details of the door.
12. You enter the shop and go to the counter. The counter assistant serves you. Money changes hands.

Finally, think of a country scene which involves trees, mountains, and a lake. Consider the picture that comes before your mind’s eye.

13. The contours of the landscape
14. The color and shape of the trees
15. The color and shape of the lake
16. A strong wind blows on the trees and on the lake, causing waves in the water

APPENDIX B. BACKWARDS SPELLING TASK

Words that appeared in my version of the task have been bolded and underlined. These are the twenty words used in the Spell-Back task from Walczyk & Taylor (2000).

steak	**brand**	**small**	sold	**peel**
salt	lisp	**cusp**	**bold**	**hard**
happy	**garner**	**bellow**	farther	**buster**
swallow	**power**	**backward**	**kicking**	**harbor**

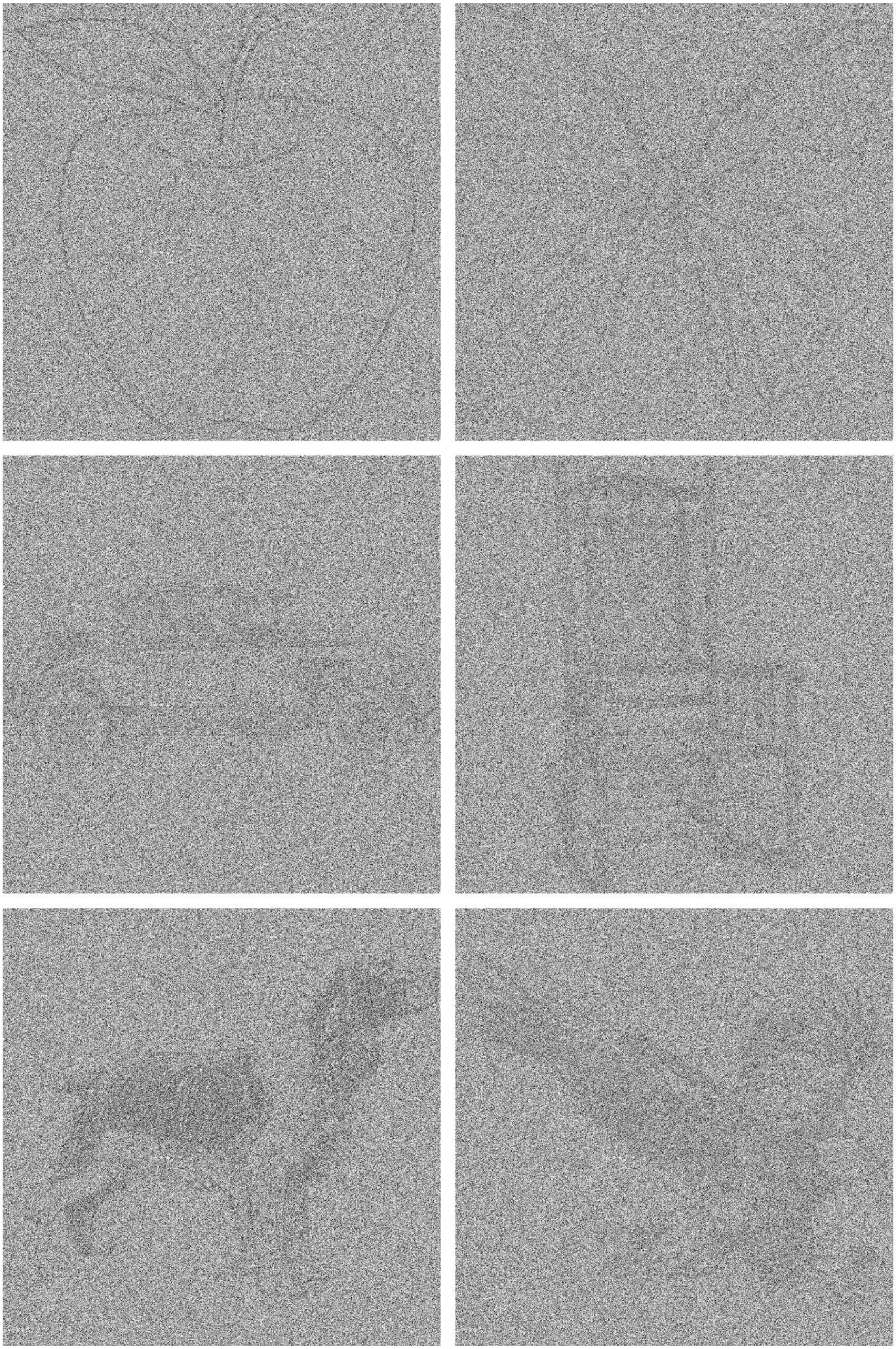
These are the twenty-two words that were used in Fernald's Spelling Backwards—Written task (1912).

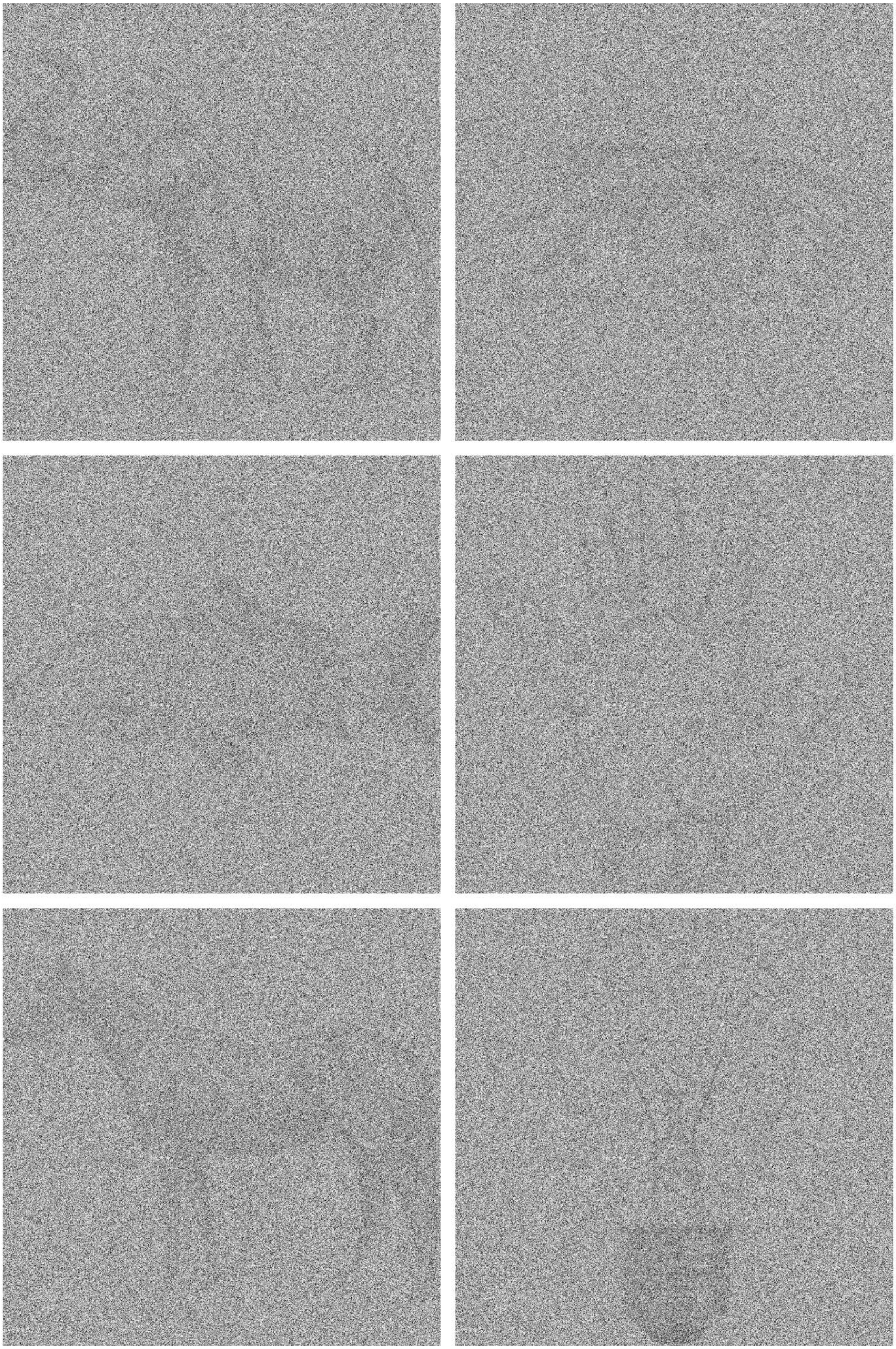
sequestered	friendliness	utterance	witchcraft	chivalrous
simplicity	equivalent	temperament	tumultuous	fictitious
assurance	credibility	reverence	manuscript	**intersperse**
heterogeneous	**cylindrical**	**publicity**	insatiable	promiscuous
substitute	vivacious			

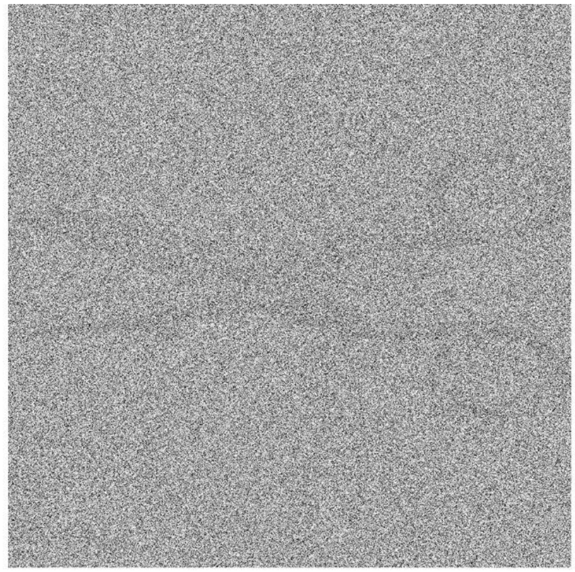
APPENDIX C. SNOWY PICTURES TASK

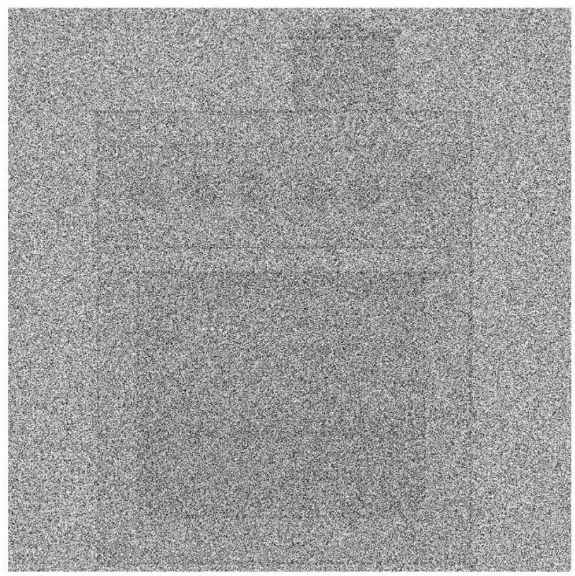
The following are the objects used in the Snowy Pictures Task as part of Study 2. Each object was presented as a black and white line drawing covered with a snowfield so as to make the object difficult to perceive. One object, the apple, was used as a practice trial. The remaining sixteen objects comprised the task proper. These materials were created especially for this study, but were inspired by similar sets of materials from existing studies (e.g., Ekstrom et al., 1976; Kozhevnikov et al., 2005). Each object was coded for accuracy by hand; sample correct answers are shown.

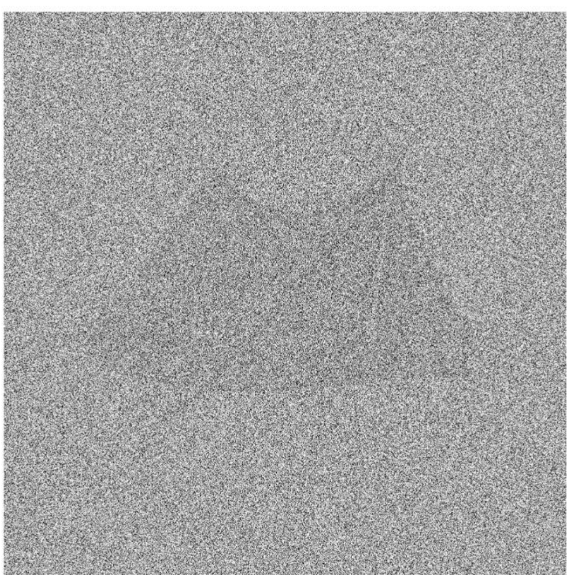
Object	Correct Answer(s)	Object	Correct Answer(s)
Apple (practice)	apple	Horse	horse; equidae
Butterfly	butterfly; moth	Lightbulb	lightbulb; bulb
Car	car; automobile	Plane	plane; airplane
Chair	chair; seat	Scissors	scissors; scizzors
Dog	dog; beagle	Ship	ship; boat
Eagle	eagle; bird	Stove	stove; oven
Elephant	elephant	Tent	tent; canopy
Eye	eye; eyeball		
Fish	fish; trout		
Hand	hand; fingers		











APPENDIX D. TAIL LENGTH TASK

The following are the twenty animals and their correct answers used in Policardi et al. (1996) for the tail length test. The question asked for each was "Does this animal generally have a short or long tail?"

Animal	Correct Answer	Animal	Correct Answer
rabbit	short	kangaroo	long
fox	long	donkey	short
mole	short	cat	long
deer	short	sheep	short
tiger	long	lizard	long
crocodile	long	rooster	long
goat	short	cow	long
mouse	long	hamster	short
pig	short	turtle	short
squirrel	long	hippopotamus	short

APPENDIX E. APPROVAL FOR RESEARCH (IRB)

IOWA STATE UNIVERSITY

collapse

Study:	**21-214**	**Sponsor(s):**	
Committee:	IRB #1	**Sponsor Id:**	
Category:		**Grants:**	
Department:	Psychology		
Agent Types:	SBER	**CRO:**	
Title:	Clarifying Aphantasia	**Year:**	2021
2018 Common Rule Date:	07/01/2021	**HIPAA:**	No
Exempt Categories:	2018 - 2 (i): Research that only includes interactions involving educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) when the information obtained is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained, directly or through identifiers linked to the subjects., 2018 - 3 (i.A): Research involving benign behavioral interventions in conjunction with the collection of information from an adult subject through verbal or written responses or audiovisual recording when the subject prospectively agrees to the intervention and information collection and the information obtained is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained, directly or through identifiers linked to the subjects. - 3 (ii) If research involves deception, it is prospectively authorized by the subject.	**FDA Study:**	No

Comments: This study is intended to investigate aphantasia, which is a term used to describe people who have little or no visual mental imagery abilities. Aphantasia, and individual differences in mental imagery abilities as a whole, are currently poorly understood in the general population when it comes to how such differences impact a person's ability to process different types of information. For example, does having no visual mental imagery lead to poorer memory for visual details such as color? The main aim of this project is to investigate the performance of people with aphantasia on a few behavioral tasks, such as identifying changes in colors, in order to determine whether their performance is different from the performance of people who do not have aphantasia.

Study-Site

Site(s):	**00 - Unspecified**	**PI:**	**Toftness, Alexander**
Status:	Active	**Additional:**	N
Approval:	July 1, 2021	**Expiration:**	**Exempt**
Initial Approval:	July 1, 2021	**Other Expirations:**	Exempt Determination Expiration - 06/29/2024
Tags:	Exempt		
Comments:			

Study-Site Contacts (1) collapse

Name	Role
Cooper, Eric	Supervising Investigator

Reference xForms (1) collapse

Form	Identifier	Stage	As Of	Ref Active	Inactivated
IRB Application	No answer provided. Clarifying Aphantasia	Complete	07/01/2021 5:36 PM ET	07/01/2021 5:36 PM ET	

Events (1) collapse

Event	Att	Instance/UDF	Start	Complete	Last Mtg
Initial Submission	5		05/25/2021	08/18/2021	08/17/2021

Copyright ©2000-2022 Tech Software. All Rights Reserved.
2022.3.6445.0/Release/14aac27 | GCWAWS1 | 2022-04-04 14:14:572 | 0.177s
Powered By IRBManager

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office of Research Ethics
Vice President for Research
2420 Lincoln Way, Suite 202
Ames, Iowa 50014
515 294-4566

Date:	07/01/2021		
To:	Alexander Toftness	Eric Cooper	
From:	Office of Research Ethics		
Title:	**Clarifying Aphantasia**		
IRB ID:	**21-214**		
Submission Type:	Initial Submission	**Exemption Date:**	07/01/2021

The project referenced above has been declared exempt from most requirements of the human subject protections regulations as described in 45 CFR 46.104 or 21 CFR 56.104 because it meets the following federal requirements for exemption:

2018 - 2 (i): Research that only includes interactions involving educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) when the information obtained is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained, directly or through identifiers linked to the subjects.
2018 - 3 (i.A): Research involving benign behavioral interventions in conjunction with the collection of information from an adult subject through verbal or written responses or audiovisual recording when the subject prospectively agrees to the intervention and information collection and the information obtained is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained, directly or through identifiers linked to the subjects. - 3 (ii) If research involves deception, it is prospectively authorized by the subject.

The determination of exemption means that:

- **You do not need to submit an application for continuing review. Instead, you will receive a request for a brief status update every three years. The status update is intended to verify that the study is still ongoing.**
- **You must carry out the research as described in the IRB application.** Review by IRB staff is required prior to implementing modifications that may change the exempt status of the research. In general, review is required for any *modifications to the research procedures* (e.g., method of data collection, nature or scope of information to be collected, nature or duration of behavioral interventions, use of deception, etc.), any change in *privacy or confidentiality protections*, modifications that result in the *inclusion of participants from vulnerable populations*, removing plans for informing participants about the study, any *change that may increase the risk or discomfort to participants, and/or* any change such that the revised procedures do not fall into one or more of the regulatory exemption categories. The purpose of review is to determine if the project still meets the federal criteria for exemption.

- All ***changes to key personnel*** must receive prior approval.

- **Promptly inform the IRB of any addition of or change in federal funding for this study.** Approval of the protocol referenced above applies only to funding sources that are specifically identified in the corresponding IRB application.

  Detailed information about requirements for submitting modifications for exempt research can be found on our website. For modifications that require prior approval, an amendment to the most recent IRB application must be submitted in IRBManager. A determination of exemption or approval from the IRB must be granted before implementing the proposed changes.

  Non-exempt research is subject to many regulatory requirements that must be addressed prior to implementation of the study. Conducting non-exempt research without IRB review and approval may constitute non-compliance with federal regulations and/or academic misconduct according to ISU policy.

Additionally:

- All research involving human participants must be submitted for IRB review. **Only the IRB or its designees may make the determination of exemption**, even if you conduct a study in the future that is exactly like this study.

- **Please inform the IRB if the Principal Investigator and/or Supervising Investigator end their role or involvement with the project** with sufficient time to allow an alternate PI/Supervising Investigator to assume oversight responsibility. Projects must have an eligible PI to remain open.

- **Immediately inform the IRB of** (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.

- **Approval from other entities may also be needed**. For example, access to data from private records (e.g., student, medical, or employment records, etc.) that are protected by FERPA, HIPAA or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **An IRB determination of exemption in no way implies or guarantees that permission from these other entities will be granted**.

- Your research study may be subject to **post-approval monitoring by Iowa State University's Office for Responsible Research**. In some cases, it may also be subject to formal audit or inspection by federal agencies and study sponsors.

- Upon completion of the project, transfer of IRB oversight to another IRB, or departure of the PI and/or Supervising Investigator, please initiate a Project Closure in IRBManager to officially close the project. For information on instances when a study may be closed, please refer to the IRB Study Closure Policy.

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.

APPENDIX F. SQUARE DONUT SCANNING TASK

The Square Donut Scanning Task was adapted from the tasks used by Dror et al. (1993) and Dror & Kosslyn (1994) to assess individual differences in mental imagery ability between groups such as pilots, elderly people, and so on. The full set of stimuli used for Study 3 are available on the OSF repository for this project, and are usable with attribution (i.e., cite this project: Toftness, 2022; https://osf.io/u3bxj/).

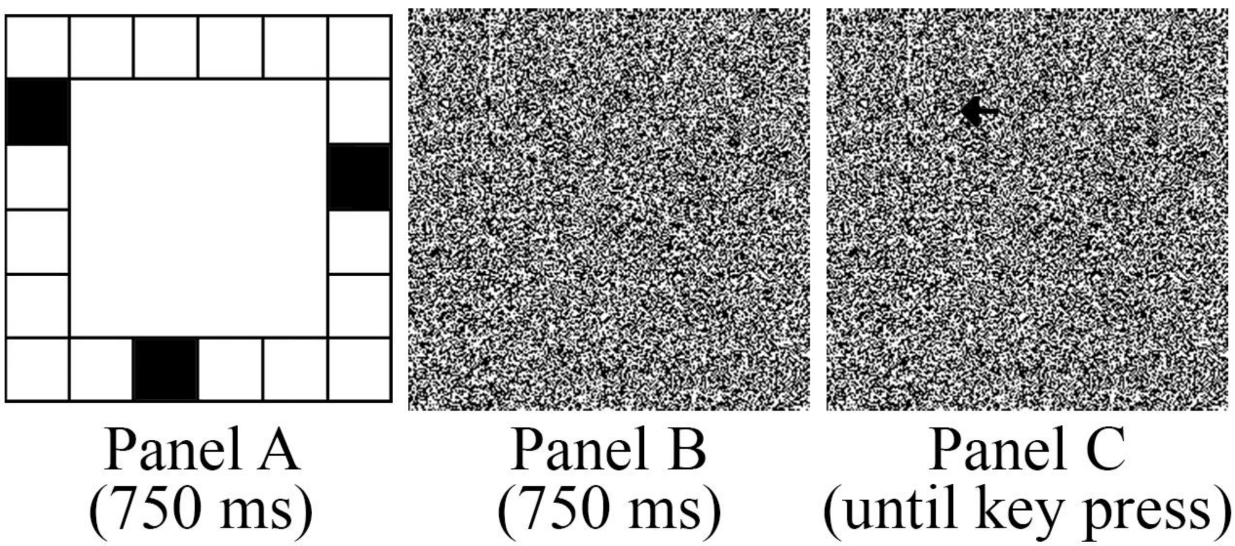


Figure 44. The general procedure for Study 3. Upon the beginning of a trial, Panel A was displayed for 750 ms, then Panel B was displayed for 750, and finally Panel C remained on the screen until the participant pressed a response key. Additionally, trials were spaced apart by 500 ms, such that there was a blank screen displayed between each trial.

The key differences between the version of the task from Dror & Kosslyn (1994) and my version of the task are:

1) The original procedure allowed participants to study the stimulus in Panel A for as long as they wanted before pressing the spacebar to continue. In my version, Panel A was displayed for 750 ms.

2) The original procedure did not use a visual mask as in Panel B. In my version, a visual mask as seen in Panel B was displayed for 750 ms.

3) The original procedure displayed the arrow for only 50 ms, the stimulus in Panel A remained on the screen during this period, and then following this period all stimuli disappeared from the screen and the participant was asked to press their response key while the screen was blank. In my version, the visual mask and the arrow were displayed together for as long as it took the participant to press a response key.

4) The original procedure feature 3 arrow conditions: near, medium, and far (i.e., the distance from the square that the arrow was pointing at). In my version, there were only two arrow conditions: near and far.

5) The original procedure featured 3 black squares for all stimuli used. My version featured a 3-square condition, as well as a 6-square condition.

APPENDIX G. CHANGE IDENTIFICATION TASK

The color, form, orientation, and position manipulations used for Study 4 are illustrated with sample stimuli below. Note that the original stimuli always appear on the left, the categorically changed versions of the stimuli always appear in the middle, and the coordinately changed versions of the stimuli always appear on the right.

The full set of stimuli used for Study 4 are available on the OSF repository for this project, and are usable with attribution (i.e., cite this project: Toftness, 2022; https://osf.io/u3bxj/).

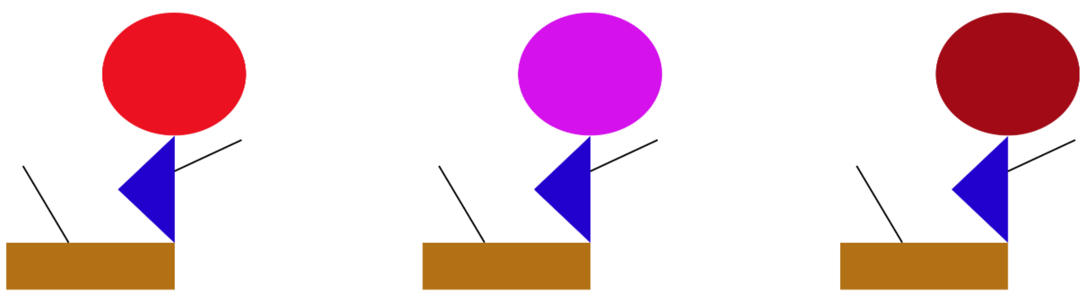


Figure 45. Sample color trial.

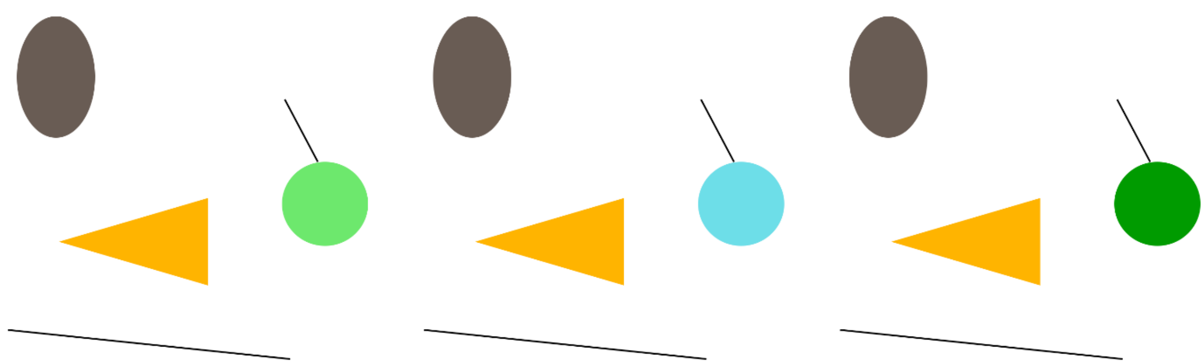


Figure 46. Sample color trial.

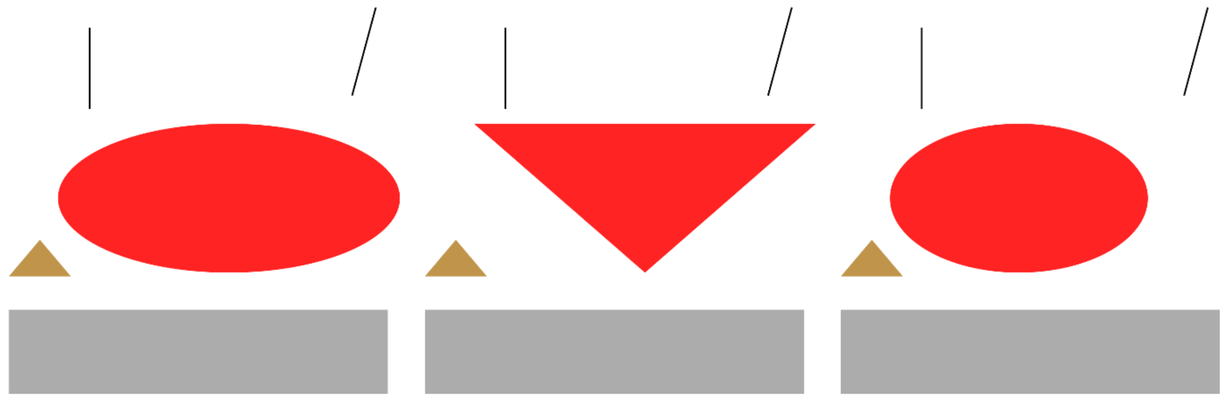


Figure 47. Sample form trial.

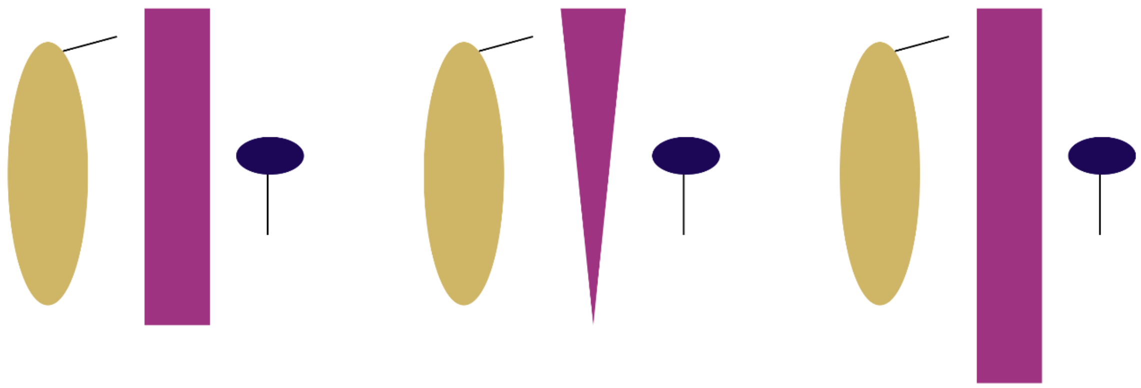


Figure 48. Sample form trial.

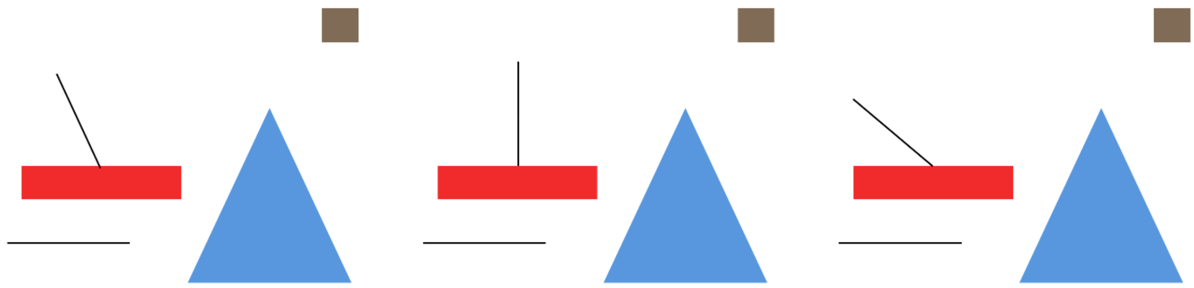


Figure 49. Sample orientation trial.

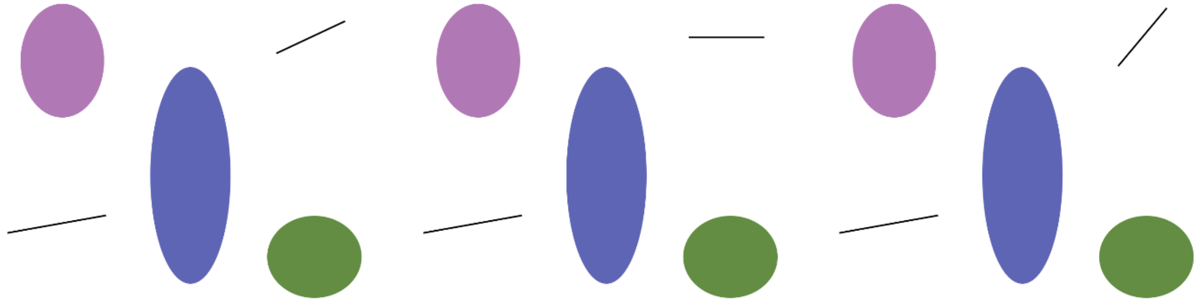


Figure 50. Sample orientation trial.

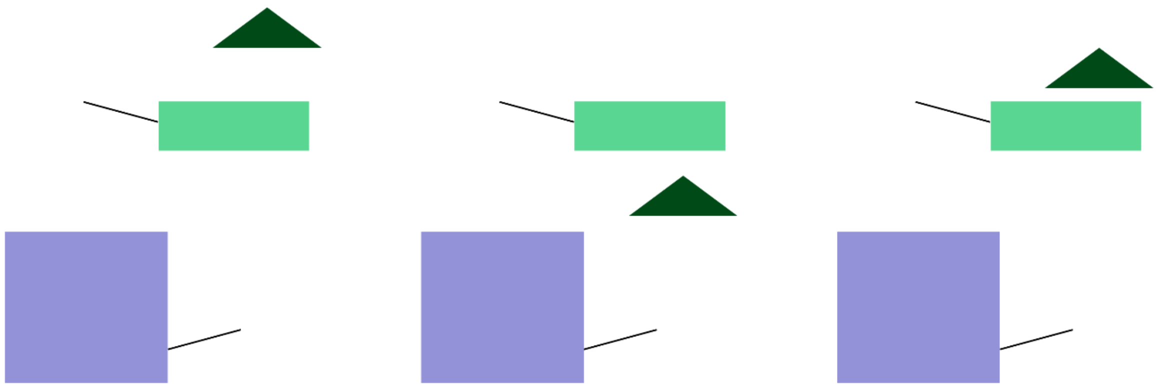


Figure 51. Sample position trial.

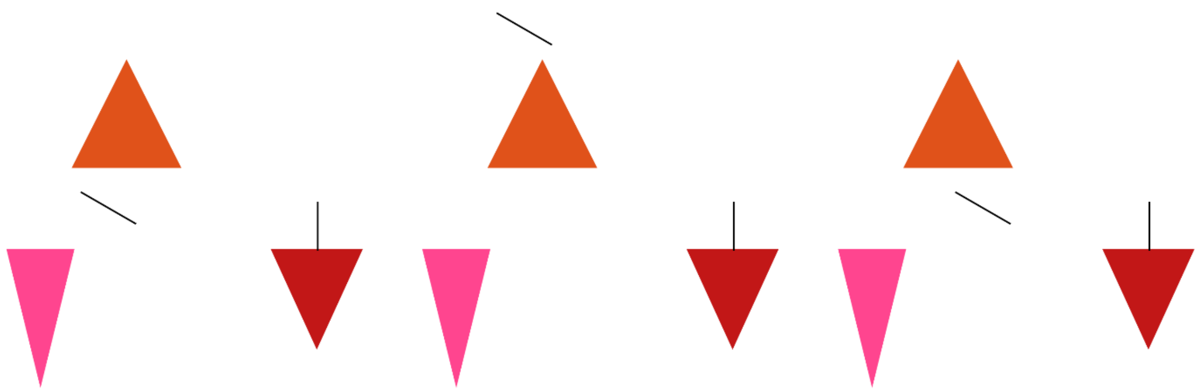


Figure 52. Sample position trial.

APPENDIX H. SUPPLEMENTAL ANALYSES

Here are the analyses and discussions that did not fit with the main body of the text or were otherwise supplemental, so placed in order to save space between the more key analyses. All datasets used for this study have been anonymized and can be accessed through the OSF project page, and are usable with attribution (i.e., cite this project: Toftness, 2022; https://osf.io/u3bxj/).

Propensity Score Approach

The propensity score approach was suggested to me during the creation of this set of studies as an alternative to creating a matched control group with the ages and sexes of the aphantasia group taken into consideration. I was encouraged, instead, to use the university's participant pool. That pool unfortunately consisted almost entirely of teenagers, as well as a disproportionate number of women. The mismatch between the undergraduate participant pool and the aphantasia group in terms of age was so pronounced, that no conventional propensity score approach was viable to correct for it without pruning out the majority of my participants, or setting an inadvisable caliper size, or pretending that one of my participants was several people. Because several of my planned analyses looked within groups, and a large number of participants had already been excluded, reducing the sample size any further was inadvisable. For example, one approach would have been to use 1:1 propensity matching without replacement using age and sex, resulting in throwing out 25 poorly-matched participants from the control group (i.e., those who were young and female—100% of exclusions with this method were 18 years old and female). However, I was not willing to trade-off the power that came from having those additional 25 participants for my analyses that looked *within* the control group as I had already instigated strict inclusion rules such as implausibly quick response times during the cognitive tasks (see the Participants section of Study 1), and the resulting balance for the covariate of age would not have been remedied by this particular propensity score approach anyway (See Table 5). I am reporting the attempt to correct for the age mismatch here in the supplemental analyses in order to save space within the study descriptions themselves.

Propensity score matching was attempted using MatchIt, an R package (see Table 7). Using the MatchIt documentation bundled with the package, I followed the recommended steps to check the initial imbalance of my dataset for the variables of age and sex. As recommended by Rubin (2001), I examined the variance ratio (the ratio of the variances of the propensity score in the two groups). A variance ratio close to 1 is ideal, with scores of .5 or 2 being “far too extreme” (Rubin, 2001, p. 6). By that metric, the initial (raw data) values for age (41.160) and sex (1.158) were respectively catastrophically poor and marginally poor. I used a logistic regression approach to estimate propensity scores for the covariates of age and sex, with 1:1 nearest neighbor matching without replacement, without caliper width. The pruned data had 25 fewer control participants—removed participants skewed young and female, as expected. This resulted in improved variance ratios for age (36.240) and sex (1.042), although the variance ratio for age remained catastrophically poor. The overall balance of the data as indexed by variance ratio for distance improved from 3.479 to 2.978, but the dataset was still poorly balanced. Additionally, as advised by Ho et al. (2007), I next considered the balance achieved by propensity score matching by “directly assessing differences” in the means for each covariate across groups while avoiding hypothesis tests (2007, p. 221). The imbalance here was not “small,” with the mean age in the aphantasia group (33.461) continuing to eclipse the mean age in the control group (19.463) even after pruning (*ns* = 95). Thus, the propensity score matching approach was discarded in favor of the approaches seen within the Studies in the main body of the manuscript.

Table 7. *Propensity Score Matching Attempt*

Variable	x̄ Aphantasia	x̄ Control	Std. x̄ Diff.	Var. Ratio
Raw Data Balance				
Distance	0.804	0.155	2.126	3.479
Age	33.461	19.158	1.198	41.160
Sex Assignment	1.495	1.683	-0.375	1.158
Matched Data Balance				
Distance	0.804	0.179	2.050	2.978
Age	33.461	19.463	1.173	36.240
Sex Assignment	1.495	1.600	-0.209	1.042

Note. For the variable of sex, a score of 1 represented a male participant and a score of 2 represented a female participant, and thus, the ideal mean was 1.5 for sex. Std. x̄ Diff. = standardized mean difference (values closer to 0 indicate more balance of the covariate across groups). Var. Ratio = variance ratios (values closer to 1 indicate more balance of the covariate across groups).

Additional VVIQ Score Correlations for Study 2

In Study 2, I looked for correlations within the aphantasia group in conjunction with the measured outcome variables (accuracy and response time) on the BST, SPT, and TLT. The VVIQ did not correlate with these measures, regardless of whether all people with aphantasia were considered, whether only people with aphantasia reporting few imagery dreams were considered (self-reported dream imagery frequency scores of 1–2), or whether only people with aphantasia reporting some imagery dreams were considered (scores of 3–6). Additionally, when looking at the control group overall, VVIQ scores did not correlate with any measured outcomes from Study 2. For completeness, these correlations are reported in the following tables.

Table 8. *Intercorrelations of Study 2 Measures and VVIQ Disaggregated by Group*

Variable	1	2	3	4	5	6	7
1. VVIQ	—	-.031	-.009	.015	-.126	-.173	.157
2. BST-ACC	-.051	—	-.358†	.152	-.154	.067	-.121
3. BST-RT	.172	-.090	—	-.385†	.558†	.042	.444†
4. SPT-ACC	.014	.299†	.012	—	-.506†	.056	-.179
5. SPT-RT	.045	-.011	.582†	-.221*	—	.140	.478†
6. TLT-ACC	.108	.292†	-.033	.173	-.023	—	-.121
7. TLT-RT	.073	-.091	.507†	.062	.540†	-.115	—

Note. The results for the aphantasia participants ($n = 94$) are shown above the diagonal. The results for the control participants ($n = 120$) are shown below the diagonal.
*$p < .05$. †$p < .01$

Table 9. *Intercorrelations of Study 2 Measures and VVIQ for Aphantasia Group Disaggregated by Imagery Dream Frequency*

Variable	1	2	3	4	5	6	7
1. VVIQ	—	-.112	-.005	.119	.068	-.233	.001
2. BST-ACC	-.016	—	-.358*	-.027	.008	.103	-.165
3. BST-RT	.012	-.352†	—	-.284	.505†	-.049	.344
4. SPT-ACC	-.050	.272*	-.440†	—	-.650†	.036	-.263
5. SPT-RT	-.168	-.230	.582†	-.442†	—	-.017	.622†
6. TLT-ACC	-.150	.057	.071	.081	.191	—	-.156
7. TLT-RT	.198	-.103	.499†	-.144	.428†	-.106	—

Note. The results for the aphantasia participants with few imagery dreams (self-reported dream imagery frequency scores of 1–2; $n = 32$) are shown above the diagonal. The results for the aphantasia participants with some imagery dreams (self-reported dream imagery frequency scores of 3–6; $n = 62$) are shown below the diagonal.
*$p < .05$. †$p < .01$

Thus, there does not appear to be a definitive linear relationship between VVIQ scores and performance on the BST, SPT, or TLT, and the significant overall correlations seen in Study 2 seems instead to be driven by a categorical difference between the aphantasia group and the control group, that can potentially be explained as a group-selection based difference due to the quasi-experimental nature of this series of studies (e.g., control participants were not paid money for participation unlike the aphantasia group, and were not interested in their mental imagery abilities unlike the control group).

Moderation Models for Study 2 Using PROCESS

I used PROCESS to test whether video condition acted as a moderator in Study 2.

Table 10. *Fit of Moderation Models for Extracted Factors from Study 2.*

Model	coefficient	*SE*	*t*	*p*	LLCI	ULCI
RT Factor						
Constant	-.926	1.26	-.725	.470	-3.42	1.59
VVIQ	.014	.021	.654	.515	-.028	.056
Video Condition	.319	.921	.346	.730	-1.50	.026
Interaction	-.004	.015	-.279	.781	-.034	.026
ACC Factor						
Constant	.687	1.08	.636	.526	-1.45	2.83
VVIQ	-.016	.018	-.830	.408	-.0511	.0209
Video Condition	-.521	.787	-.662	.509	-2.08	1.04
Interaction	.011	.013	.852	.396	-.015	.037

Note. LLCI = lower 95% confidence interval. ULCI = upper 95% confidence interval. For each model, the model name represents the outcome variable (*Y*). VVIQ was always entered as the predictor variable (*X*), and Video Condition was always entered as the Moderator (*W*) variable. *N* = 120 for each model, representing the entire control group. The six moderation analyses not shown showed a similar pattern of universal non-significance (i.e., when BST-ACC, BST-RT, SPT-ACC, SPT-RT, TLT-ACC, or TLT-RT were separately entered as *Y*, the interaction was not significant in any case, nor was VVIQ alone or Video Condition alone significant in any case).

Specifically, I wanted to know if video condition mattered for the relationship between VVIQ scores and the six performance measures as well as the two extracted factors (Model 1; Hayes, 2018). In short, video condition was never a significant moderator, and the models were poor fits. Table 8 depicts the results of the PROCESS model for the extracted factors (see Study 2) to give you an idea of how poor the fit was for all tested models.

The Imagery Dream Frequency Analyses

Planned analyses for this project included investigating the aphantasia group as two sub-groups: those who infrequently experienced visual mental imagery during periods of dreaming (*n* = 32), and those who at least sometimes experienced visual mental imagery during periods of dreaming (*n* = 62; see Study 1). This was done because these two groups may be heterogeneous at the level of this recognition task due to differences in their deficits at the level of their brains. This was completed for both Study 3 and Study 4, but no significant results were found. The entire set of analyses is not reported here, as it does not reveal new insights beyond the analyses reported in the Results sections for Studies 3 and 4.

For Study 3, 2 x 2 x 2 mixed ANCOVAs compared group, scanning distance, and number of black squares ((few imagery dreams aphantasia x some imagery dreams aphantasia) x (near x far) x (three-square x six-square)) while entering covariates of age and sex, first using response times as the dependent measure, and then using accuracy as the dependent measure. These analyses did not reveal any new information beyond what was already gleaned from the analyses that treated the aphantasia group as a whole. Whether response times or accuracy was used as the dependent measure did not make a difference: no significant effects of group (or interactions with group) emerged when the aphantasia group was split in two. Most crucially, there was no significant interaction between group and scanning distance for response times (F (1, 91) = .136, p = .713, η_p^2 = .001, observed power = .065), and there was no significant interaction between group and number of squares for response times (F (1, 91) = 3.94, p = .074, η_p^2 = .035, observed power = .431). Thus, there is not enough evidence to say that these two sub-groups performed differently at the square donut scanning task.

For Study 4, accuracy was investigated as the dependent measure. A 2 x 2 x 4 mixed ANCOVA was run using group, degree changed, and trial type ((few imagery dreams aphantasia x some imagery dreams aphantasia) x (near x far) x (three-square x six-square) while entering age and sex as covariates.

There was no main effect of group concerning the estimated marginal means of accuracy between the "few dreams" aphantasia group (M = 42.8%, SE = 1.7%) and the "some dreams" aphantasia group (M = 43.4%, SE = 1.2%; F (1, 90) = 4.29, p = .015, η_p^2 = .039, observed power = .743). Additionally, when accuracy was the dependent measure, group did not interact with any other covariate or variable. This pattern repeated when response times were considered as the dependent measure: no significant main effects or interactions were detected concerning group.

When each trial type was considered separately (i.e., run as an additional analysis as in the Study 4 results section), no significant differences between the "few dreams" and "some dreams" aphantasia groups were uncovered in terms of their overall accuracy, in terms of the slope of their accuracy between categorical and coordinate trials, in terms of their overall response times, or in terms of the slope of their response times between categorical and coordinate trials.

In most cases, the two aphantasia groups divided by dream imagery frequency behaved virtually indistinguishably from one another, with extreme overlap in their 95% confidence intervals for estimated marginal means. Thus, this series of studies was not able to detect a difference between the participants with aphantasia who frequently experienced visual imagery while dreaming and the participants with aphantasia who infrequently experienced visual imagery while dreaming.

VVIQ Categorical Transformation and Analyses

Zeman et al. (2020) proposed a categorization scheme for dividing people into discrete groups of “extreme aphantasia,” “moderate aphantasia,” “typical imagery,” “moderate hyperphantasia,” and “extreme hyperphantasia” based on VVIQ scores. Considering that VVIQ scores are so easily pushed around (as shown in Study 1 and in previous research as discussed in the Introduction), the seemingly arbitrary VVIQ scores chosen for the mid-groups (e.g., the cut-off points for typical imagery), the tendency for VVIQ scores to *not* correlate with “objective” measurements of imagery use during cognitive tasks, and the fact that aphantasia categorization scheme(s) that do not depend on VVIQ scores are already in development (Reeder, 2022), I do not believe that this categorization scheme will be used in the future of aphantasia research. However, I was interested in putting these VVIQ-dependent groups to the test, and so I conducted a few exploratory analyses with them. Table 11 shows how the aphantasia and control groups broke across the five discrete groups defined by Zeman et al. (2020). As can be plainly seen, my sample did not include very many participants with high enough VVIQ scores to be considered to have “moderate hyperphantasia” or “extreme hyperphantasia.” Thus, my sample size (N = 215) was a limiting factor due to the relative rarity of high VVIQ scores. In fact, the cell size for the “moderate aphantasia” group is also smaller than I would like, having only 18 participants. Because of the limited sample size, and because the majority of the people in the aphantasia group fall into the “extreme aphantasia” group, and the majority of the people in the control group fall into the “typical imagery” group, this particular dataset is not ideal for testing the categorization scheme of Zeman et al.

Table 11. *Categorical VVIQ Groups*

Categorization	VVIQ Score	Aphantasia Group	Control Group	Total
Extreme Aphantasia	16	64 (67.4%)	1 (0.8%)	65 (30.2%)
Moderate Aphantasia	17–23	16 (16.8%)	2 (1.7%)	18 (8.4%)
Typical Imagery	24–74	15 (15.8%)	107 (89.2%)	122 (56.7%)
Moderate Hyperphantasia	75–79	0 (0%)	8 (6.7%)	8 (3.7%)
Extreme Hyperphantasia	80	0 (0%)	2 (1.7%)	2 (0.9%)

Note. Total $N = 215$, aphantasia $n = 95$, controls $n = 120$. Categorization scheme suggested by Zeman et al., 2020.

However, it is interesting that the categorization scheme of Zeman et al. places a full 15.8% of my aphantasia participants into the “typical imagery” category. This suggests that perhaps the cut-off point used by Zeman et al. was too conservative, or that the use of VVIQ scores does not fully reflect whatever it is about themselves that people use to determine whether they identify as having aphantasia.

Integration of Response Time and Accuracy Measures as Rate Correct Scores

According to a recent series of studies that investigated different ways of combining speed and accuracy performance measures in cognitive research, such combined measures may be useful, but only if the researcher first tests the component effects (Vandierendonck, 2017, p. 672). That is, accuracy and response time should also be reported, instead of just reporting a combined measure of both accuracy and response time, such as rate correct score. Based on a series of Monte Carlo simulations that tested a variety of integration methods, one linear integration that Vandierendonck recommends is “rate correct score” (RCS) which can be interpreted as “number of correct responses per second of activity” (p. 654). The formula for RCS can be written as seen in Figure 53. It is worth noting that the author points out that “there is no general agreement about the utility of such integrated measures and their efficiency in detecting performance differences” (p. 653). However, he was able to show in his simulated datasets that RCS generally adds value as an extra level of analysis. After having tested the component effects, general recommendations are to apply a linear integration (e.g., rate correct score) rather than a binning procedure, and to keep in mind that integrated measures perform best when both response times and accuracy measures agree with one another (i.e., “point in the same direction”, p. 671).

$$\mathrm{RCS} = \frac{c}{\sum \mathrm{RT}}$$

Figure 53. Rate correct score (RCS) formula from Woltz & Was, 2006. *c* represents the number of correct responses in a particular condition from a particular participant, while the denominator represents the sum total of all response times from that condition for a particular participant.

This score can be easily calculated for a variety of the conditions with this series of studies. Participants in my studies were told, for all five cognitive tasks: “Please try to respond ACCURATELY and QUICKLY.” Thus, they were not specifically told to prioritize speed over accuracy, or vice versa. Because studies 2 and 4 did not have specific hypotheses about whether response time or accuracy would be the dependent measure of most interest, RCS was calculated for those tasks in order to examine whether RCS could assist with interpreting group differences.

While Study 3 did have a specific hypothesis about response times being the more important dependent measure, the RCS for that task was also calculated for completeness.