The relationship between interactive-imagery instructions and association memory

Jeremy J. Thomas[1] · Kezziah C. Ayuno[1] · Felicitas E. Kluger[1] · Jeremy B. Caplan[1,2]

Abstract

Interactive imagery, one of the most effective strategies for remembering pairs of words, involves asking participants to form mental images during study. We tested the hypothesis that the visual image is, in fact, responsible for its memory benefit. Neither subjectively reported vividness (all experiments) nor objective imagery skill (experiments 1 and 3) could explain the benefit of interactive imagery for cued recall. Aphantasic participants, who self-identified little to no mental imagery, benefited from interactive-imagery instructions as much as controls (experiment 3). Imagery instructions did not improve memory for the constituent order of associations (AB versus BA), even when participants were told how to incorporate order within their images (experiments 1 and 2). Taken together, our results suggest that the visual format of images may not be responsible for the effectiveness of the interactive-imagery instruction and moreover, interactive imagery may not result in qualitatively different associative memories.

Introduction

One of the best-known ways to increase memory for word pairs (e.g., study APPLE-OVEN, when presented APPLE, recall OVEN), is to instruct participants to form a mental image of the two words interacting (Bower, 1970; Bower & Winzenz, 1970; Dunlosky, Hertzog, & Powell-Moman, 2005; Paivio & Yuille, 1969; Paivio & Foth, 1970; Richardson, 1985; 1998). For example, "imagine an APPLE cooked inside an OVEN, in your mind's eye." Participants who receive interactive-imagery instructions perform significantly better at cued recall than participants given no strategy instruction (Richardson, 1985; 1998), and $\sim 20 - 50\%$ higher cued recall accuracy than participants instructed to use rote repetition (Bower & Winzenz, 1970; Bower, 1970). Bower and Winzenz (1970) and Paivio and Foth (1970) found that interactive-imagery instructions could even outperform comparable verbally mediated instructions (e.g., form a sentence with both words) for concrete word pairs, although Dunlosky et al., (2005) found these instructions were comparable.

At face value, interactive-imagery instructions might cause participants to literally construct rich visual representations, directly improving memory in this way (Yates, 1966). However, this hypothesis is hard to test because visual imagery cannot be directly observed. Here we examine the effect of interactive-imagery instructions with two main approaches. First, we test the visually relevant characteristics of the imagery instruction and individual differences characteristics of the participants. Second, we ask whether interactive-imagery changes the formal nature of the representation; specifically, whether or not constituent order (knowledge that it was APPLE–OVEN, not OVEN–APPLE) is coupled with memory for the pairing, itself.

Testing for visual-imagery characteristics of associations formed through interactive imagery One way to interrogate how visual imagery functions is to exploit individual differences. There is large individual variability in the subjective experience of mental imagery (Marks, 1973; Zeman, Dewar, & Della Sala, 2015; Zeman, Milton, Della Sala, Dewar, Frayling, Gaddum, & Winlove, 2020) and objectively scored imagery/visuospatial tasks (Keogh & Pearson, 2018; Sanchez, 2019; Zeman, Della Sala, Torrens, Gountouna, McGonigle, & Logie, 2010). If the visual image itself is fundamental to the benefit of interactive imagery, one would expect that imagery instructions may benefit individuals with vivid or accurate mental imagery more

✉ Jeremy J. Thomas
 jjthomas@ualberta.ca

1 Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

2 Neuroscience and Mental Health Institute, University of Alberta, Edmonton, Alberta T6G 2E1, Canada

than those with poor mental imagery. Alternatively, visual imagery may be epiphenomenal (Pylyshyn, 2002), implying that individual differences in mental imagery should not relate to objective memory performance. Our three experiments test the hypothesis that both mental imagery vividness and skill determine how much an individual benefits from interactive-imagery instructions.

There is considerable support for a central role of imagery in association memory. Instructions to use interactive imagery produces higher cued recall than without imagery instructions, and associations involving words higher in imageability are remembered better (Bower, 1970; Bower & Winzenz, 1970; Paivio, Smythe, & Yuille, 1968; Paivio & Yuille, 1969; Paivio, 1969; Paivio & Foth, 1970). Beyond memory for word pairs, ancient texts claim that forming vivid images can improve memory of various kinds (Foer, 2011; Gesualdo, 1592; Yates, 1966). For example, when using the method of loci, a popular technique for ordered lists, skilled memorizers report forming mental images of to-be-remembered items in various locations (e.g., Maguire, Valentine, Wilding, & Kapur, 2003).

Common advice by skilled memorizers is that vivid imagery is important for the efficacy of mnemonic strategies (e.g., Foer 2011; Konrad 2013; Müller, Konrad, Kohn, Muñoz-López, Czisch, Fernández, & Dresler, 2018). To test this idea, Sanchez (2019) measured individual differences in imagery/visuospatial skill with the Cube Comparisons Task (CCT; a mental rotation task), and the Paper Folding Task (PFT; judging the outcome of multiple folds and hole-punches of a paper) (French, Ekstrom, & Price, 1963), and examined the correlation to memory performance. In Sanchez' (2019) study, aggregate CCT and PFT performance correlated with serial recall performance for participants who were instructed to use the method of loci, but not for participants who were given a control instruction. However, three studies did not find a significant relationship between Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) and successful use of the method of loci (Kliegl, Smith, & Baltes, 1990; Kluger, Oladimeji, Tan, Brown, & Caplan, 2022; McKellar, Marks and Barron, cited as in-preparation by Marks (1972)).

In light of these variable findings, we included the VVIQ (all experiments) and PFT (experiments 1 and 3) to assess subjective quality of imagery and objective imagery ability, respectively. The hypothesis that the construction of a visual image is central to the success of interactive-imagery instructions implies that either or both the VVIQ and PFT should covary with cued recall accuracy. Alternatively, interactive-imagery effects may not depend on vivid or accurate mental images or perhaps do not require any conscious experience of mental imagery at all.

To further test the hypothesis that visual imagery is vital for the benefits of interactive imagery, we tested people with

the phenomenon of aphantasia, extremely low or non-existent self-reported ability to form voluntary mental images. Current interest in aphantasia originated with patient MX (Zeman et al., 2010), who, after undergoing coronary angioplasty, reported a complete inability to form mental images. MX exhibited completely intact performance in imagery/visuospatial related tasks. However, closer examination of behavior and brain activity suggested MX was applying distinct verbal/symbolic strategies to complete tasks typically thought to require mental imagery. Other studies have examined larger populations of self-reported aphantasics who rate significantly low vividness (Zeman et al., 2015), report worse autobiographical memory, and have difficulty recognizing faces (Zeman et al., 2020). Specific to memory, Bainbridge, Pounder, Eardley, and Baker (2021) examined the ability of aphantasics to draw photographs of rooms in a house from memory. Aphantasics were not different from controls in copying a presented image, indicating no deficits to their perceptual ability. Interestingly, aphantasics remembered fewer objects than controls, but for the objects they could remember, they reproduced their spatial arrangement at the same level as controls. These results indicated that aphantasics had specific deficits to object, but not spatial memory. If the visual image is the necessary mechanism by which interactive-imagery instructions increase cued recall accuracy, aphantasics should show no such advantage (experiment 3).

Interactive imagery and the formal properties of associations We could find no formal implementation of imagery in any mathematical model of association memory. However, image-based associations could differ in their qualitative or formal characteristics, which might be meaningful from a mathematical modeling perspective. One hypothesis about the relationship between imagery and the formal characteristics of association-memory emerged while reviewing existing models, as we now elaborate.

Mathematical models make starkly different predictions about memory for the constituent order of associations (AB versus BA) (Kato & Caplan, 2017), a memory task that has only begun to be investigated experimentally. Matrix-based models (Anderson, 1970) and concatenation-based models (Hintzman, 1984; Shiffrin & Steyvers, 1997), which we now refer to as perfect-order models, encode associations with non-commutative operations, and consequently predict that order is remembered perfectly given that the association itself is intact. Convolution-based models (Kelly, Blostein, & Mewhort, 2013; Murdock, 1982; Metcalfe Eich, 1982; Plate, 1995), in contrast, are based on commutative operations that completely discard order (and see Cox & Criss 2017, 2020, and Criss & Shiffrin's 2005 model, which also disregard order). In these models, which we now refer to as order-absent models, information for order, if present, must

be provided by some other term, predicting that the ability to remember the constituent order will be unrelated to remembering the pairing itself. Kato and Caplan (2017) found no evidence for either of these predictions. In their study, word pairs were tested with cued recall, and then, an order recognition task, where participants had to recognize whether a probe was in the correct order (AB), or reversed (BA) (Greene & Tussing, 2001; Kounios, Bachman, Casasanto, Grossman, Smith, & Yang, 2003; Kounios, Smith, Yang, Bachman, & D'Esposito, 2001; Yang, Zhao, Zhu, Mecklinger, Fang, & Han, 2013). Challenging both perfect-order and order-absent models, they found a significant correlation between order recognition and cued recall performance; however, this correlation was significantly smaller than a control correlation (with associative recognition), suggesting associations are not stored with perfect order, nor are they completely order-absent.

If we take imagery at face value, it seems plausible that a visual image could provide an effective means of incorporating order, such as left-to-right within the image, or top-to-bottom. This might be just the thing that participants are missing in their spontaneously adopted strategies. So, in addition to increasing memory accuracy, interactive-imagery instructions might help participants incorporate order, and render the association non-commutative like in a perfect-order model. This was our first hypothesis. The alternative hypothesis is that imagery is simply a good "hook", engaging participants better in the task, but otherwise invoking the same associative mechanism as in conditions without imagery instructions. This hypothesis leads to the prediction that the relationship between order and the association itself will be unchanged with interactive-imagery instructions. We tested these two hypotheses in experiments 1 and 2 with order recognition subsequent to cued recall for all studied pairs in one group, and as a control, associative recognition in another group.

Summary of experiments In all experiments, participants studied lists of eight word-pairs followed by cued recall. First, we obtained a baseline measure of memory with no strategy instructions; then participants were given imagery instructions (all experiments), or a filler instruction (experiment 1). To test the hypothesis that visual images are necessary for memory benefit due to interactive imagery, and that individual differences in imagery ability/vividness should predict memory benefit, vividness was assessed with the VVIQ in all experiments, and imagery skill was assessed with the PFT in experiments 1 and 2. Experiment 3 applied a stronger test of the visual imagery hypothesis by recruiting aphantasics. In experiments 1 and 2, we also tested the hypothesis that imagery could provide a way for participants to incorporate order and generate associations that are more non-commutative (like a matrix model). Cued recall was

followed by either order or associative recognition to test the relationship between constituent order and memory for the pair itself. The prediction is that imagery instructions will increase order recognition, and, moreover, its relationship to cued recall. Finally, we also include supplementary materials with additional analyses.

Experiment 1

Methods

Participants

Participants enrolled in introductory psychology courses at the University of Alberta ($N = 227$) participated for partial course credit. Participants were required to have learned English before the age of six, have normal or corrected-to-normal vision, and be comfortable typing. Participants chose one of 15 testing rooms in order of arrival, blind to condition. One participant was excluded from analyses for not completing the experiment within the allotted 50 min. Procedures in all experiments were approved by a University of Alberta ethical review board.

Groups There were two main experimental groups. The imagery group ($N = 113$) received interactive-imagery instructions halfway through the word lists, and the control group ($N = 114$) received filler instructions halfway through the lists (Fig. 1). Each experimental group was further subdivided into two conditions. Following cued recall, one condition performed order recognition ($N = 57$ and 56 for imagery and control, respectively), and the other condition performed associative recognition ($N = 56$ and 58, respectively). For analyses involving only cued recall, these conditions were collapsed within the imagery group and control group. For all analyses involving recognition tasks, these conditions were separated and named, accordingly, control-order recognition, control-associative recognition, imagery-order recognition, and imagery-associative recognition.

Materials

Stimuli were the 478 nouns from the Toronto Word Pool (Friendly, Franklin, Hoffman, & Rubin, 1982), 4–8 letters and spanning the full ranges of concreteness mean (SD) = 5.32 (1.32), and with frequency = 62.47 (82.45) per million (Kucera & Francis, 1967). Words were assigned to pairs and lists with the computer's random number generator. Study pairs, cued recall, and recognition test probes were presented in uppercase, white, Courier bold font.
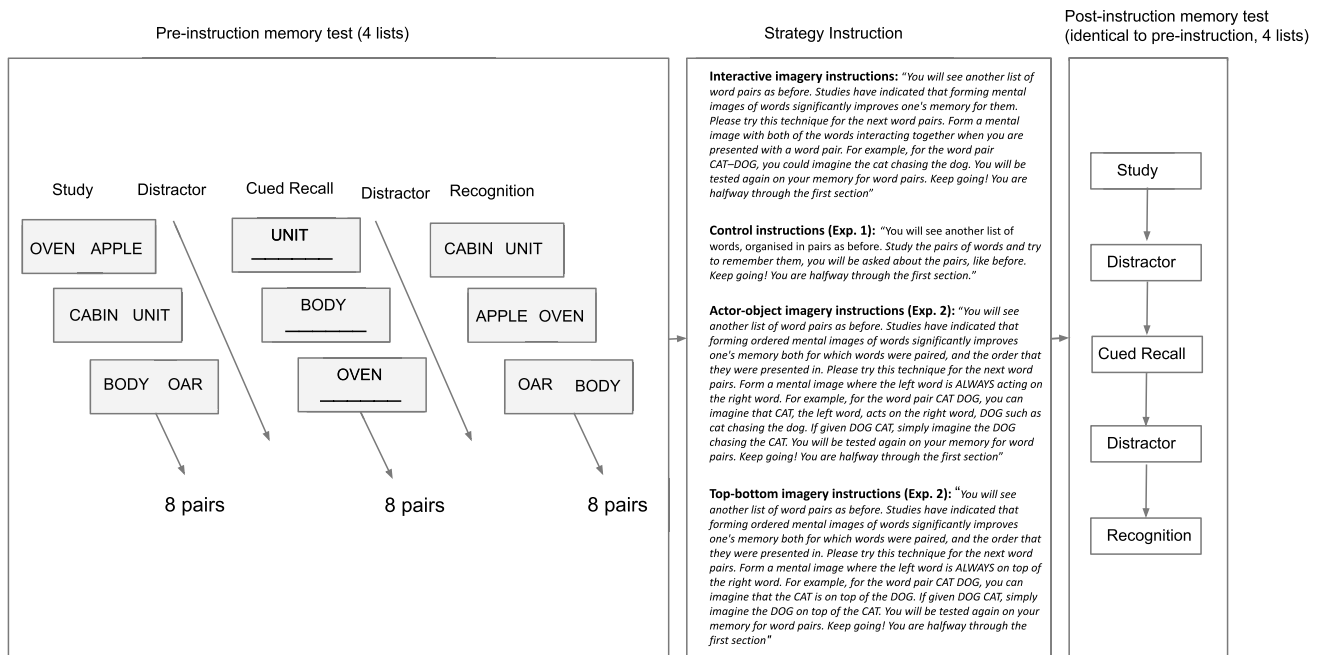
Pre-instruction memory test (4 lists)

Strategy Instruction

Post-instruction memory test
(identical to pre-instruction, 4 lists)

Interactive imagery instructions: *"You will see another list of word pairs as before. Studies have indicated that forming mental images of words significantly improves one's memory for them. Please try this technique for the next word pairs. Form a mental image with both of the words interacting together when you are presented with a word pair. For example, for the word pair CAT–DOG, you could imagine the cat chasing the dog. You will be tested again on your memory for word pairs. Keep going! You are halfway through the first section"*

Control instructions (Exp. 1): *"You will see another list of words, organised in pairs as before. Study the pairs of words and try to remember them, you will be asked about the pairs, like before. Keep going! You are halfway through the first section."*

Actor-object imagery instructions (Exp. 2): *"You will see another list of word pairs as before. Studies have indicated that forming ordered mental images of words significantly improves one's memory both for which words were paired, and the order that they were presented in. Please try this technique for the next word pairs. Form a mental image where the left word is ALWAYS acting on the right word. For example, for the word pair CAT DOG, you can imagine that CAT, the left word, acts on the right word, DOG such as cat chasing the dog. If given DOG CAT, simply imagine the DOG chasing the CAT. You will be tested again on your memory for word pairs. Keep going! You are halfway through the first section"*

Top-bottom imagery instructions (Exp. 2): *"You will see another list of word pairs as before. Studies have indicated that forming ordered mental images of words significantly improves one's memory both for which words were paired, and the order that they were presented in. Please try this technique for the next word pairs. Form a mental image where the left word is ALWAYS on top of the right word. For example, for the word pair CAT DOG, you can imagine that the CAT is on top of the DOG. If given DOG CAT, simply imagine the DOG on top of the CAT. You will be tested again on your memory for word pairs. Keep going! You are halfway through the first section"*

Fig. 1 There were a total of eight lists in experiments 1 and 2. Halfway through the lists, participants either received imagery or control instructions in experiment 1, and either imagery, actor–object or top–bottom instructions in experiment 2. All participants in experiment 3 received imagery instructions. Experiment 3 had a similar design, but without associative or order recognition trials after cued recall, and a total of ten lists

Procedure

The experiment was run in Python, in conjunction with the Python Experiment-Programming Library (Geller, Schleifer, Sederberg, Jacobs, & Kahana, 2007), for the first cohort of participants. Because software updates made lab computers incompatible with PyEPL, we ran the second cohort in a MATLAB port, written with the PsychToolBox experiment programming extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997), and the CogToolBox Library (Fraundorf et al., 2014). Illustrated in Fig. 1, the session included study of word pairs, cued recall, followed by order or associative recognition tests, repeated for eight study sets, with five trials of a mathematical distractor task between study, cued recall and recognition sets. Given that Kato and Caplan (2017) found that initial cued recall tests affected subsequent recognition tests but did not change the coupling of order with association memory, we tested every pair initially with cued recall (as in experiment 1 of Kato & Caplan, 2017) to maximize the data yield (and see page S1). Interactive-imagery instructions or control filler instructions were administered after the fourth list in a pretest (Lists 1–4)/posttest (Lists 5–8) design, allowing us to check for equal baseline performance (pre-instruction), and get a closer estimate of the true effect of imagery instructions above baseline. Participants then completed the VVIQ and the PFT. Halfway through data collection, a section was added after the PFT, where participants were asked to rate how often they used interactive imagery, and then asked to type a free-form response about their strategy use, reported on page S2.

Practice list Participants performed one practice list, excluded from analyses, at the beginning of the session, during which they were walked through the tasks.

Study phase For each list, participants viewed eight pairs in sequence. The two words in a pair were presented side by side, centered on the screen, for 2850 ms, with a 150-ms inter-pair blank.

Distractor Interleaved between study, recall, and recognition, participants were administered a math distractor task. Participants had to solve the sum of three digits, randomly drawn from two to eight within 5000 ms followed by a 200-ms blank inter-trial interval. Participants typed their response, which was displayed on the screen, and upon pressing ENTER, the color of the response digit changed to gray, to show the response registered, and the 200-ms inter-trial interval was initiated after the 5000-ms response interval elapsed.

Cued recall Each studied pair was tested once with cued recall. Direction of cued recall (forward, APPLE–?, or

backward, ?–OVEN) was counterbalanced (Python version: across all lists except the practice; MATLAB version: within each list). The cue word was presented in centrally with a centered response line underneath, regardless of the direction of cued recall. The letters appeared on the line as the participant typed, submitting the word with the ENTER key. The next cued recall trial started 750 ms later. ENTER was only accepted once more than two letters were typed, to reduce participants speeding through. In the Python version, if participants did not press ENTER within 15,000 ms, the trial ended, was scored incorrect, and the next cued recall trial was presented. In the MATLAB version, this time-limit was removed.

Recognition Two probe words were presented side by side centrally, as in the study phase. In order recognition, participants judged if a presented probe was intact (e.g., OVEN APPLE) or reverse (e.g., APPLE OVEN). In associative recognition, participants judged whether a presented probe was intact (e.g., OVEN APPLE) or recombined (e.g., OVEN BUTTON). Key 1 was assigned to intact and key 2 was assigned to reverse or recombined. Other keys were ignored. Recombined probes were only rearranged with other pairs within the current list, and a pair probed with an intact probe was never used to create a recombined probe. Pairs were tested in pseudo-random order. In the Python version, the number of intact and lure (reverse or recombined) probes were counterbalanced over all analyzed lists (excluding practice). In the MATLAB version, trials were counterbalanced over all lists including the practice list.[1] In the Python version, the trial was aborted after 15,000 ms. Rather than score these timed-out trials as incorrect, they were omitted from analyses (two trials in all, both in control-associative participants). To prevent missing data, the 15,000-ms timeout limit was removed in the MATLAB version. The next recognition trial started after a 750-ms blank screen.

Vividness of visual imagery questionnaire Participants completed a computerized version of the Visual Vividness of Imagery Questionnaire (Marks, 1973), which asks participants to imagine four scenes. A description of each scene was displayed on the screen, followed by instructions to imagine four items within the scene and to rate vividness on a scale from one (perfectly vivid imagery), to five (no image at all) using the number keys. To indicate the response

---

[1] Due to programming error, counterbalancing was slightly unbalanced for associative recognition in the MATLAB version. When one recombined trial was randomly assigned to given list, it did not have another recombined pair to exchange words with, and appeared as an intact trial. The occurrence of this error was rare, with 11 participants having one extra intact trial, and one participant having two extra intact trials.

registered, the choice changed to green for 1000 ms, immediately followed by the next item. VVIQ score was the sum of these ratings, ranging from 16 (perfectly vivid imagery) to 80 (no image formed at all).

Paper Folding Task Participants completed a computerized version of the PFT (French et al., 1963), consisting of 20 questions increasing in difficulty. Each question was a series of images that depicted a piece of paper being folded successively and then hole-punched. The question was displayed to the left of a central vertical line, and five possible choices were displayed to the right, selected with the keys 1–5. The chosen option was highlighted in green for 1000 ms, immediately followed by the next question. Mean accuracy and response time were analyzed.

Distribution of VVIQ ratings and PFT ratings Distributions of VVIQ ratings and PFT scores aligned with previous studies (Table 1).

Analyses

To check null effects, we include Bayesian analyses (with uniform priors) run in JASP Team (2021). The Bayes factor is a ratio of evidence, where by convention, when $BF_{10} > 3$, the effect is considered supported, and when $BF_{10} < 0.3$, the effect is considered more consistent with the null. For ANOVAs, $BF_{inclusion}$, which summarizes across all factorial models and quantifies whether each model fits better with the main effect or interaction included versus excluded. We measured order and associative recognition with $d' = z$(hit rate) − z(false alarm rate). Whenever hit or false alarm rate were 0 or 1, one-half an observation was added or subtracted to avoid infinities.

Results and discussion

Cued recall We replicated the interactive-imagery advantage for cued recall. A mixed ANOVA on cued recall accuracy (Fig. 2), with design Group (imagery, control group) × Instruction phase (pre-instruction, post-instruction), returned significant main effects of Instruction phase, $F(1,225) = 110.79$, $MSE = 2.91$, $p < .001$, $\eta_p^2 = 0.33$, $BF_{inclusion} > 1000$, and Group, $F(1,225) = 4.92$, $MSE = 0.41$, $p = .03$, $\eta_p^2 = 0.02$, $BF_{inclusion} > 1000$; however, the interaction was also significant, $F(1,225) = 41.5$, $MSE = 1.09$, $p < .001$, $\eta_p^2 = 0.16$, $BF_{inclusion} > 1000$. Simple effects found no difference between groups pre-instruction ($p = .19$, $BF_{10} = 0.33$), but significantly higher accuracy for the imagery group post-instruction ($p < .001$, $BF_{10} > 1000$). Additionally, for both groups, accuracy significantly increased post-instruction (both $p < .001$, $BF_{10} > 33$). Thus, perhaps due to practice effects, the control group moderately improved as

Table 1 M(SD) (Means and standard deviations) of VVIQ ratings for each group in experiments 1, 2, and 3, and PFT scores in experiment 1 and 3, along with population estimates for VVIQ ratings from McKelvie's (1995), and PFT scores in the control and method of loci group in Sanchez (2019)

Experiment and Group	VVIQ rating	PFT score
Sanchez (2019) method of loci group	N/A	12.52 (2.59)
Sanchez (2019) control group	N/A	11.87 (3.30)
McKelvie (1995) VVIQ population estimate	36.9 (11.07)	N/A
Experiment 1: Imagery-order recognition sub-condition	31.8 (10.86)	13.02 (4.06)
Experiment 1: Imagery-associative recognition sub-condition	32.9 (8.94)	13.70 (3.87)
Experiment 1: Control-order recognition sub-condition	32.5 (8.39)	13.14 (3.75)
Experiment 1: Control-associative recognition sub-condition	32.7 (9.73)	13.83 (4.54)
Experiment 2: Actor-object-order recognition sub-condition	36.2 (12.62)	N/A
Experiment 2: Actor-object-associative recognition sub-condition	36.2 (10.07)	N/A
Experiment 2: Standard-imagery-order recognition sub-condition	36.3 (10.52)	N/A
Experiment 2: Standard-imagery-associative recognition sub-condition	35.2 (8.07)	N/A
Experiment 2: Top-bottom-order recognition sub-condition	36.9 (11.92)	N/A
Experiment 2: Top-bottom-associative recognition sub-condition	35.7 (11.35)	N/A
Experiment 3: Consistent aphantasic group	61.0 (18.06)	12.40 (4.61)
Experiment 3: Consistent non-aphantasic group	38.1 (13.89)	13.15 (4.37)
Experiment 3: Inconsistent responder group	44.7 (15.67)	11.98 (4.50)
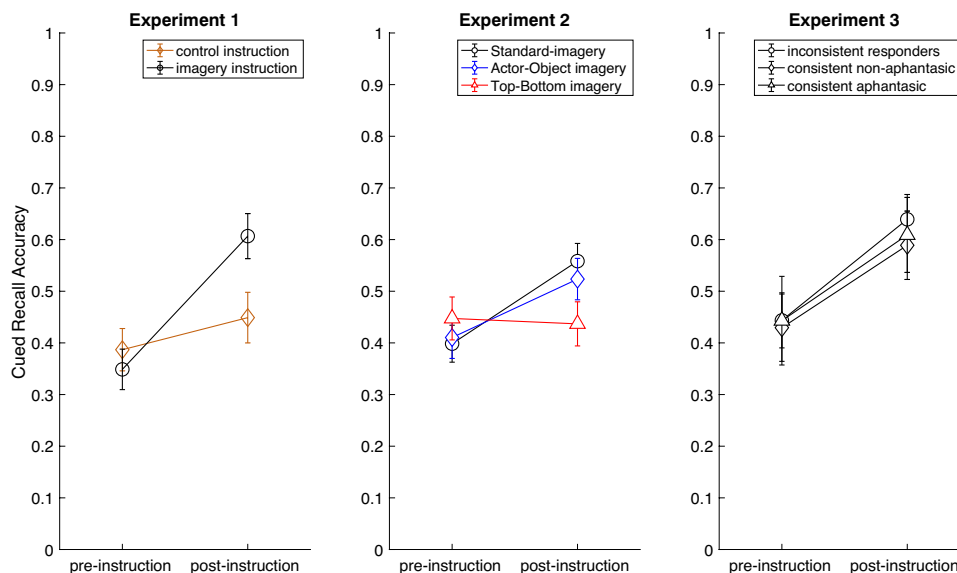
Fig. 2 Pre- and post-instruction cued recall accuracy for all three experiments. (*Left*) In experiment 1, the imagery group received instructions to use interactive imagery halfway through the word lists. The control group was simply instructed to continue with the experiment. (*Middle*) In experiment 2, participants either received standard-imagery, actor-object imagery, or top-bottom imagery instructions. (*Right*) In experiment 3, all participants received imagery instructions. *Error bars* represent 95% confidence intervals based on standard error of the mean

the experiment progressed; however, the imagery group performed significantly better in the post-instruction phase, and exhibited a greater improvement from baseline compared to the control group.[2]

Associative and order recognition A mixed ANOVA on associative recognition d' (Fig. 3), with design Group

---

[2] Expanding on these findings, we also found evidence that imagery instructions were most beneficial for participants with poor baseline performance (page S4).
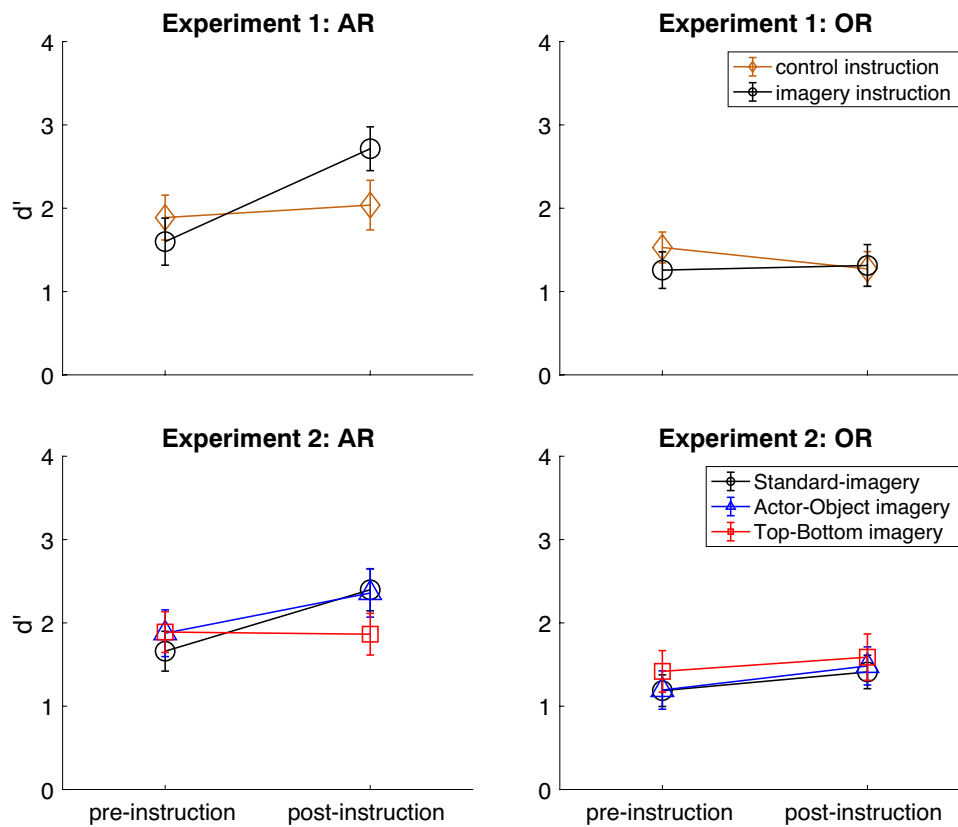
Fig. 3 Pre- and post-instruction order (OR), and associative recognition (AR) performance for experiment 1 and 2. In experiment 1, participants either received standard imagery instructions or control instructions. In experiment 2, participants received either standard-

imagery, top-bottom imagery, and actor-object imagery instructions. *Error bars* represent 95% confidence intervals based on standard error of the mean

(imagery-associative recognition, control-associative recognition) × Instruction phase (pre-instruction, post-instruction) returned a non-significant main effect of Group ($p = .25$, $BF_{inclusion} = 612.89$),[3] a significant main effect of Instruction phase, $F(1,112) = 38.13$, $MSE = 22.79$, $p < .001$, $\eta_p^2 = 0.25$, $BF_{inclusion} > 1000$, and a significant interaction Group × Instruction phase, $F(1,112) = 21.24$, $MSE = 13.29$, $p < .001$, $\eta_p^2 = 0.17$, $BF_{inclusion} > 1000$. Simple effects revealed a non-significant group difference in performance pre-instruction ($p = .14$, $BF_{10} = 0.54$), but the imagery-associative recognition condition performed significantly better post-instruction ($p < .001$, $BF_{10} = 31.12$). Additionally, the imagery-associative recognition condition improved post-instruction ($p < .001$, $BF_{10} > 1000$), but the control-associative recognition condition did not significantly improve ($p =$

.16, $BF_{10} = 0.37$). These analyses indicate that imagery instructions substantially improved associative recognition performance over control instructions.

An ANOVA with the same design, on order recognition d' (Fig. 3) returned non-significant, favored null main effects of both factors (both $p > .2$, $BF_{inclusion} < 0.3$). The interaction Group × Instruction phase nearly reached significance, $F(1,111) = 3.90$, $MSE = 1.61$, $p = .051$, $\eta_p^2 = 0.03$, although the Bayesian analysis favored the null ($BF_{inclusion} = 0.26$). Nonetheless, we cautiously followed up on the interaction with simple effects. The control-order recognition group performed significantly worse post-instruction ($p = .01$, $BF_{10} = 3.07$), while the imagery-order recognition group did not exhibit any significant change ($p = .65$, $BF_{10} = 0.16$). Additionally, the group difference in performance was not significant pre-instruction ($p = .06$, $BF_{10} = 0.98$), or post-instruction ($p = .80$, $BF_{10} = 0.21$). In sum, imagery instructions did not improve order recognition performance, but may have acted against a performance decrease observed in the control-order recognition group.

---

[3] A non-significant effect can have strong evidence in a Bayesian analysis because JASP's implementation of Bayesian model selection refuses to consider models including interactions without the main terms. Thus, if there is strong evidence for the interaction, it will also return strong evidence for the main terms included in interactions .

The relationship among mental imagery skill, vividness, and the effectiveness of interactive-imagery instructions Next, we asked if any individual difference measure would explain individual differences in memory performance (Tables S1–S3). Correlations between VVIQ ratings and cued recall accuracy were all non-significant and either were, or were nearly, supported null effects (all $p > .09$, $BF_{10} < 0.45$), and likewise for order recognition (all $p > .15$, $BF_{10} < 0.46$). VVIQ ratings significantly correlated with post-instruction associative recognition performance in the imagery-associative recognition condition, $r(54) = -.44$, $p < .001$, $BF_{10} = 44.10$, but this correlation was not significant post-instruction for control-associative recognition group, $r(56) = -.04$, $p = .78$, $BF_{10} = 0.17$; and these correlations differed significantly (Fisher's test, $p = .024$). Thus, individual differences in mental imagery vividness explained differences in associative recognition performance under interactive-imagery conditions,[4] but could not explain the interactive imagery advantage for cued recall.

PFT accuracy exhibited significant, positive correlations with nearly all memory tasks, and not only with memory performance in the imagery group (Tables S1–S3). Although the tables show some exceptions, our results, particularly the presence of pre-instruction correlations, suggest that PFT accuracy does not specifically relate to interactive imagery, and may have either reflected a general factor such as motivation, task engagement or a distinct cognitive process such as working memory.

PFT response time was not significantly related to the memory measures apart from a significant positive correlation with post-instruction cued recall accuracy, $r(111) = .27$, $p = .004$, $BF_{10} = 7.49$, and post-instruction associative recognition performance, $r(54) = .32$, $p = .017$, $BF_{10} = 2.74$, both in the imagery group. If *longer* PFT response times indicate worse performance, these correlations would be counter-intuitive. A simpler interpretation is that longer PFT latencies are a consequence of greater general effort or engagement (a successful speed–accuracy trade-off) rather than mental imagery skill. Thus, the pattern argues against the idea that mental imagery accuracy or skill is required for the memory benefit.[5]

The relationship of order recognition to cued recall Figure S11 plots log-odds transformed cued recall accuracy versus both order recognition and associative recognition d', for both imagery and control groups. Pre-instruction, the associative recognition–cued recall correlations (imagery: $r(56) = .86$, $p < .001$, control: $r(56) = .83$, $p < .001$), were larger than the order recognition–cued-recall correlations (imagery: $r(55) = .43$, $p < .001$, control: $r(54) = .46$, $p < .001$). The difference in correlations was significant for both groups pre-instruction (Fisher's tests, imagery: $p < .001$, control: $p < .001$). This pattern persisted post-instruction; associative recognition-cued recall correlations (imagery: $r(54) = .70$, $p < .001$, control: $r(56) = .81$, $p < .001$) were also larger than order recognition-cued recall correlations (imagery: $r(55) = .31$, $p = .020$, control: $r(54) = .37$, $p = .005$; Fisher's test, imagery: $p < .001$, control: $p = .005$). Thus, consistent with Kato and Caplan (2017), order recognition exhibited a smaller correlation to cued recall accuracy than associative recognition.[6]

Importantly, Fisher's tests between the control and imagery group OR-CR correlations were not significant pre- ($p = .85$) and post-instruction ($p = .70$), and AR-CR correlations pre- ($p = .57$) and post-instruction ($p = .15$), suggesting that imagery instructions did not affect the dependence of order or associative recognition on cued recall. This result does not support the hypothesis that imagery instructions help participants incorporate order. Instead, we have evidence for the alternative hypothesis that imagery does not change the formal characteristics of the association.[7]

Summary of experiment 1 Interactive-imagery instructions increased cued recall accuracy and associative recognition d' above baseline, and compared to the control group. Imagery instructions did not improve order recognition, or change its relationship to cued recall. Both imagery vividness and skill did not predict the effectiveness of imagery instructions.

Experiment 2

The results of experiment 1 raised an additional question. Although interactive imagery failed to improve order recognition, if participants were given a specific way to incorporate order into their image, could that improve order recognition?

---

[4] The significant correlation between VVIQ ratings and post-imagery instruction associative recognition d' was not replicated in experiment 2, thus, we do not consider this a robust finding and do not discuss it in the general discussion.

[5] We also found no support for the idea that significant pre-instruction PFT correlations were due to high PFT scorers spontaneously adopting imagery before being instructed to do so (page S4).

[6] When interpreting these results, one might consider the effect of testing pairs with cued recall before order recognition. Indeed, this was a major point addressed by Kato and Caplan (2017), who, in their second experiment, withheld half the pairs from cued recall testing, and in their third experiment, moved cued recall to the end of the session. In both cases, they found that the order-cued recall relationship persisted, which we also found when analyzing testing effects in our own dataset, reported on page S1.

[7] The *within-subject* analysis of the OR–CR relationship for experiment 1 and 2 are reported on page S5 and S13.

We addressed this question by modifying the interactive-imagery instruction in two ways (see Fig. 1 for instructions). First, physically enacting verbal stimuli (e.g., hit the NAIL) improves benefits memory (enactment effects; cf. Allen, Waterman, Yang, & Jaroslawska, 2022; Engelkamp, 1991; Engelkamp, 1995; Sivashankar & Fernandes, 2021), even when imagined (Allen et al., 2022; Yang et al., 2021). We hypothesized that imagining an actor–object relationship might not only exploit this benefit but also incorporate order into the image. Second, whereas the left–right axis is generally symmetric, gravity can break the symmetry; for example, a MOUSE on top of an ELEPHANT conjures a different meaning than the ELEPHANT on the MOUSE. We thus added two imagery instructions, where images were to comprise actor–object or top–bottom relationships, respectively.

Experiment 2 was pre-registered. All pre-registered analyses are reported. For analyses of the *within-subject* relationship of order/associative recognition to cued recall of pairs, see page S13.

Methods

Participants

Participants ($N = 433$) were recruited through Prolific (https://www.prolific.co), and compensated £6.50 for a 50-min session. Participants were required to have English as their first language, be fluent in English, and have a Prolific approval rating above 70%. Our initial pre-registered exclusion criteria included failure to pass two attention checks, and/or exceeding a specified floor or ceiling threshold for recognition performance. Instead, we excluded participants who demonstrated clear evidence of disengagement, rather than exclude participants may have responded earnestly but performed extremely poorly or well: 13 were excluded because they re-wrote the presented probe in cued recall, suggesting they did not understand the task; three were excluded because they did not respond to any cued recall trial; seven were excluded because they responded to < 10% of recognition trials (Table 2).

Table 2 Experiment 2: Included and excluded participants for each group and sub-condition. A total of 23 participants were excluded

Group/Condition	Included	Excluded
Standard-imagery/Associative recognition	73	2
Standard-imagery/Order recognition	91	4
Actor-Object imagery/Associative recognition	72	6
Actor-Object imagery/Order recognition	68	3
Top-Bottom imagery/Associative recognition	76	4
Top-Bottom imagery/Order recognition	53	4

Groups Three main experimental groups were each divided into two sub-conditions: i) standard-imagery/associative recognition, ii) standard-imagery/order recognition, iii) actor-object/associative recognition, iv) actor-object/order recognition, v) top-bottom/associative recognition, vi) top-bottom/order recognition. Groups/sub-conditions were assigned with a random number generator function.

Materials and procedures

Materials and procedures were identical to experiment 1; however, with the following differences: (1) Experiment 2 was conducted online, with recruitment from https://www.prolific.co, hosted on Pavlovia.org. Groups were assigned with a random number generator. (2) The Paper Folding Task was omitted to save session time. (3) After the mid-session strategy instruction, participants were asked "Please explain back to us, in your own words, what we have asked you to do on the previous screen". Short-answer responses were rated by two coders (KA and JT) blinded to group to quantify comprehension of instructions (corresponding on page S9).[8] (4) After completing the VVIQ, participants rated, on a five-point scale, their frequency of incorporating mental imagery, interactivity, and order during study (page S7). (5) Participants answered a reversed-sense aphantasia question (see experiment 3 methods). Five aphantasic participants are presented as case studies in supplementary materials on page S13. (6) Two engagement checks were included; participants were presented a short message, "NOTE: Remember the number: X", in the top-right corner of the screen, highlighted in blue, and against a grey foreground, once during the mid-session strategy instruction, and again, immediately after the VVIQ. Participants were asked to recall the number shortly after; however, two participants indicated their monitor cut off this number from the screen, thus, we applied different criteria, stated above. (7) Distractor trials were held for a fixed 1000-ms period after the response was entered, regardless of response time. Additionally, there was a 5000-ms maximum time-limit, and a blank 200-ms inter-trial interval. (8) Recognition trials were counterbalanced over all trials, including the practice list. However, there were two programming errors with associative recognition; i) a single recombined trial assigned to a list appeared as an intact trial, because could not exchange items with another pair. ii) random shuffling of recombined probes sometimes resulted in the original pairing. $N = 198$ participants had more intact probes than recombined probes, and of these participants, there was an average of nine extra intact trials. However, baseline associative recognition d'

---

[8] Note that in these analyses, we did not perform the chi-squared tests proposed in the pre-registration.

was comparable to experiment 1 (Fig. 3), suggesting mean associative recognition performance was not sensitive to this design difference. (9) Recognition trials initially had a 15,000-ms time-limit. For d' calculations, rather than omit these trials from analyses outright, a correction was applied for each timed-out trial; if an intact trial was timed-out, 0.5 of an observation was added to hits and to misses. Likewise, if a recombined/reversed trial timed-out, 0.5 of an observation was added to false alarms and to correct rejections. In this way, timed-out trials pushed the overall d' to 0, where $d' = 0$ represents no memory, as if the participant were guessing. Thus, with this correction, we assume that when a trial times-out, a participant has no knowledge, and would have guessed if given the opportunity. A total of 23 trials timed-out and were corrected in this manner. To remove the need for this estimation and obtain a response from each participant to each trial, time-limits were removed for recognition trials halfway through data-collection.

Distribution of VVIQ ratings VVIQ rating distributions were comparable to experiment 1 (Table 1).

Results and discussion

Cued recall A mixed ANOVA on cued recall accuracy (Fig. 2) with design Group (standard-imagery, actor-object, top-bottom) × Instruction phase (pre-instruction, post-instruction) returned a significant main effect of Instruction phase, $F(1,430) = 71.13$, $MSE = 1.64$, $p < .001$, $\eta_p^2 = 0.14$, $BF_{\text{inclusion}} > 1000$. The main effect of Group was not significant, $F(2,430) = 1.15$, $MSE = 0.10$, $p = .32$, $\eta_p^2 = 0.005$, $BF_{\text{inclusion}} > 1000$, but had strong evidence in the Bayesian analysis[3]. However, the Group × Instruction phase interaction was significant, $F(2,430) = 24.74$, $MSE = 0.57$, $p < .001$, $\eta_p^2 = 0.10$, $BF_{\text{inclusion}} > 1000$. Simple effects returned a supported null effect of Group pre-instruction ($p = .19$, $BF_{10} = 0.13$), but significant effect post-instruction ($p < .001$, $BF_{10} = 379.6$). Follow-up t tests on the post-instruction Group difference indicated a non-significant, supported null difference between the standard-imagery and actor-object imagery, $p = .19$, $BF_{10} = 0.29$. Additionally, cued recall accuracy was significantly lower in the top-bottom imagery compared to the standard-imagery ($p < .001$, $BF_{10} > 1000$), and actor-object ($p = .004$, $BF_{10} = 7.27$) imagery groups. Simple effects also returned a significant effect of Instruction phase for the actor-object, and standard-imagery group (both $p < .001$, $BF_{10} > 1000$), both of which increased in performance post-instruction, but a supported null difference for the top-bottom imagery group ($p = .60$, $BF_{10} = 0.11$). In sum, the actor-object imagery instructions matched the robust benefits of standard interactive-imagery instructions for memory, but top-bottom instructions were ineffective.

Associative and order recognition Broadly speaking, the results for associative recognition paralleled those for cued recall; standard and actor-object imagery instructions were effective to improve performance and top-bottom instructions were ineffective. A mixed ANOVA on associative recognition d' (Fig. 3), with design Group [3] × Instruction phase [2] returned significant main effects of Instruction phase, $F(1,195) = 21.38$, $MSE = 15.34$, $p < .001$, $\eta_p^2 = 0.10$, $BF_{\text{inclusion}} > 1000$, and significant Group × Instruction phase interaction, $F(2,195) = 7.56$, $MSE = 5.43$, $p < .001$, $\eta_p^2 = 0.07$, $BF_{\text{inclusion}} = 22.13$. Simple effects indicated that associative recognition performance increased post-instruction in both the actor-object group ($p = .003$, $BF_{10} = 9.65$) and standard-imagery group ($p < .001$, $BF_{10} > 1000$), while the top-bottom group had a supported null difference between instruction phases ($p = .86$, $BF_{10} = 0.13$). Simple effects with the factor Group returned a supported null difference pre-instruction ($p = .34$, $BF_{10} = 0.16$), but a significant difference post-instruction ($p = .005$, $BF_{10} = 5.82$). Follow-up t-tests on the post-instruction group difference indicate that actor-object and standard-imagery had a supported null difference ($p = .84$, $BF_{10} = 0.21$), but both groups performed significantly better than the top-bottom group ($p = .017$, $BF_{10} = 3.75$ and $p = .003$, $BF_{10} = 9.86$, respectively).

Results for order recognition diverged from the other tasks. A mixed ANOVA on order recognition d' (Fig. 3), with design Group [3] × Instruction phase [2] returned a significant main effect of Instruction phase, $F(1,232) = 12.89$, $MSE = 6.02$, $p < .001$, $\eta_p^2 = 0.053$, $BF_{\text{inclusion}} = 37.83$, indicating that order recognition d' improved in all three groups post-instruction. A significant improvement in order recognition somewhat diverged from null effects observed in experiment 1; however, the effect in all three groups was small in magnitude (d' post-minus-pre $\approx +0.25$, Fig. 3), and post-instruction performance was in the range of values from experiment 1, suggesting the effect on order recognition was small in comparison to associative recognition. Importantly, both the main effect and interaction involving Group were supported null (both $p > .32$, $BF_{\text{inclusion}} < 0.3$), indicating that emphasizing order in the imagery instructions did not improve order recognition more than standard interactive-imagery instructions.

The relationship between mental imagery vividness and the effectiveness of interactive-imagery instructions VVIQ ratings had a supported null relationship to cued recall in three groups and instruction phases (all $p > .15$, $BF_{10} < 0.3$), replicating and extending findings from experiment 1 and 2. A single exception was found in the top-bottom imagery group pre-instruction, $r(54) = -.18$, $p = .03$, $BF_{10} = 1.09$, although a Bayesian correlation returned inconclusive evidence for this relationship (Tables S4–S6). Correlations between

Table 3 Experiment 2: Correlations between log-odds cued recall accuracy and both order and associative recognition collapsed across participants, and separated into groups

	Pre-instruction		Post-instruction	
	r	p	r	p
All participants/Associative recognition	.67	< .001	.66	< .001
All participants/Order recognition	.54	< .001	.47	< .001
All participants Fisher's test (Order versus Associative)	$z = 1.99, p = .047$		$z = 2.93, p = .003$	
Standard-imagery/Associative recognition	.64	< .001	.72	< .001
Standard-imagery/Order recognition	.58	< .001	.32	.0017
Standard-imagery Fisher's test (Order versus Associative)	$z = 0.59, p = .55$		$z = 3.62, p = .0003$	
Actor-Object/Associative recognition	.70	< .001	.66	< .001
Actor-Object/Order recognition	.44	< .001	.50	< .001
Actor-Object Fisher's test (Order versus Associative)	$z = 2.18, p = .030$		$z = 1.34, p = .18$	
Top-Bottom/Associative recognition	.67	< .001	.61	< .001
Top-Bottom/Order recognition	.64	< .001	.68	< .001
Top-Bottom Fisher's test (Order versus Associative)	$z = 0.29, p = .77$		$z = 0.64, p = .53$	

VVIQ ratings and both order recognition, and associative recognition were non-significant, supported null effects (all $p > .36$, $BF_{10} < 0.31$). The failure to replicate the correlation between VVIQ and associative recognition in experiment 1 suggests that this finding is not particularly robust and will not be discussed further. Thus, vividness ratings in the VVIQ could not explain the advantage of standard-imagery instructions, nor memory performance under any imagery instruction variant.

The relationship of order recognition to cued-recall Due to low trial counts for recombined trials (see Methods), the associative recognition measures are noisy and should be interpreted with caution. However, with maximal power by collapsing across groups (Fig. S15, Table 3), the OR-CR correlation was significantly lower than the AR-CR correlation, both pre- and post-instruction ($p = .047$, $p = .0034$, respectively, Fisher's tests), replicating experiment 1 and Kato and Caplan (2017). Next, we asked if, for any instruction, the OR-CR correlation changed from pre- to post-instruction. These comparisons were non-significant for top-bottom ($p = .71$, Fisher's test) and actor-object group ($p = .63$), but there was a significant decrease post-instruction for the standard-imagery group ($p = .034$). This pre- versus post-instruction difference in the standard-imagery group was largely driven by a single outlier (Fig. S15) who performed extremely poorly in cued recall, but extremely well in order recognition. When removed, the comparison was non-significant ($p = .14$).

Summary of experiment 2 Standard interactive-imagery and actor-object imagery instructions boosted cued recall and associative recognition above baseline, and compared to the top-bottom imagery instructions. Surprisingly, both imagery instructions that emphasized order had a negligible effect on order recognition, and did not affect its relationship to cued recall. Replicating experiment 1, imagery vividness did not predict the effectiveness of imagery instructions.

Experiment 3

Experiment 1 suggested the large benefit to cued recall of interactive imagery has little to do with subjective detail or objective visual imagery skill. In experiment 3, we recruited aphantasics, who self-report an inability to form visual imagery, and non-aphantasics, to do cued recall, VVIQ and PFT as in experiment 1. If the presence of visual images is required for interactive imagery, then aphantasics should show substantially less benefit from imagery instructions than non-aphantasics.

Methods

Participants

Just as in experiment 1, participants ($N = 122$) were enrolled in an introductory psychology class at the University of Alberta, and recruitment had the same basic restrictions. Participants who had enrolled in experiment 1 were not permitted to participate in this study. Four participants were excluded from analyses because they accessed the online link and completed the experiment twice; both sessions were excluded. One participant was excluded for providing no cued recall or math distractor responses.

Recruitment Before the experimental session, potential aphantasics and non-aphantasics were identified via online

mass-testing questionnaires administered to the University of Alberta introductory psychology students at the beginning of the Fall 2020 ($N = 2357$) and Winter 2021 ($N = 1975$) semesters. Along with many other items that were part of different studies, questionnaire participants responded yes/no to "Are you able to form mental images (i.e., pictures) in your mind's eye?".

Recruitment for experiment 3 was conducted after the Winter 2021 questionnaire was administered, and was restricted to participants who responded to this question in *either* the Fall or Winter questionnaire. We note here that filling out a mass questionnaire did not guarantee that a student signed-up for our experiment. Participants could only sign up if they had answered the aphantasia question in the mass testing. A different project code was visible to those who answered yes and no, respectively, to roughly equate recruitment rates. However, we further classified the 122 who participated with the additional in-session, reversed-sense aphantasia question.

Aphantasia classification We classified aphantasia in these 122 participants based on three different criteria, which we call "consistent", "moderate" and "extreme" aphantasics, respectively.

The first criterion was based on consistent response to the yes/no aphantasia question. Participants who consistently indicated being unable to form mental images in mass-testing and in-session, were classified as "consistent aphantasic" ($N = 25$). Those who consistently indicated the opposite were "consistent non-aphantasic" ($N = 34$). Those who were inconsistent in their responses to this question formed a third "inconsistent-responder" group ($N = 64$). Because inconsistent responders changed their answers across testing sessions, we were hesitant to classify them as either aphantasic or non-aphantasic, as they might have been unsure of their status. Additionally, because the recruitment question was embedded within a much longer questionnaire this raised the possibility that individuals would not respond conscientiously to each questionnaire item. This provided more reason for classifying aphantasia based on multiple responses.

To be more selective, we also applied more conservative second and third criteria from Zeman et al., (2020). Of the "consistent aphantasics," participants rating 73–79 (maximum 80) VVIQ in-session were considered "moderate" aphantasics ($N = 7$), while ratings of 80/80 were considered "extreme" aphantasics ($N = 3$). VVIQ criterion aphantasic participants are reported as case studies (Table 4).

A strength of our procedure was that our experimental session was separated by days or weeks from the Winter mass-testing questionnaire. The in-session reversed-sense aphantasia question and VVIQ were at the end of the session. We thought this should make the constructs of aphantasia

and even visual imagery less front-of-mind for participants than in previous aphantasia studies.

Mass questionnaire aphantasia prevalence rates Next, we applied our three aphantasia classification criteria to mass questionnaire data to provide an estimate of the prevalence of aphantasia in our student population. Note that the following numbers are based *solely* on mass questionnaire data and not on the sub-sample tested with memory tasks in experiment 3.

We identified 772 participants who answered the aphantasia question in both the Fall and Winter mass testing sessions. Of these participants, 30 indicated being unable to form mental images in both sessions (3.9%). This approached Faw's (2009) previously estimated rate of 2–3%.

Our conservative aphantasia classification criteria based on VVIQ cutoffs were identical to Zeman et al., (2020), who observed the rate of moderate aphantasia ($73 - 79/80$) and extreme aphantasia ($80/80$) to be 2.6% and 0.7% in their mass-testing questionnaire. First, of the $N = 2000$ who completed the VVIQ in the Fall 2020 mass-testing, 23 (0.9%) and 9 (0.4%) met these VVIQ cutoffs, respectively. Next, of the 1975 participants who responded to the VVIQ in Winter 2021 mass testing questionnaire, 43 (2.2%) and 26 (1.3%) participants met the moderate and extreme VVIQ cutoffs, respectively. In sum, the prevalence rates that were derived from the Fall 2020 questionnaire were considerably lower than previous observations, while the rates that were derived from the Winter 2021 questionnaire were closer to Zeman et al., (2020). The extreme cutoff appears far more highly selected than prior aphantasic samples.

Materials and procedures

Materials and procedures were identical to experiment 1 except: (1) This experiment was conducted completely online, on Pavlovia.org. The experiment was created using the PsychoPy Builder interface (Peirce et al., 2019) and translated to a PsychoJS experiment (Bridges, Pitiot, MacAskill, & Pierce, 2020). As in experiment 1, recruitment was conducted through the University of Alberta psychology research participation pool, but participants completed the experiment on their personal devices. (2) All participants were instructed to use interactive imagery half-way through the session (no control group) (3) Recognition tasks were omitted; pairs were only tested with cued recall. (4) To use the additional testing time freed up from the recognition tasks, participants studied 10 lists (cf. eight in experiment 1). (5) The PFT was re-added to the design, and administered after the VVIQ just like in experiment 1. (6) After the PFT, participants answered a

Table 4 Experiment 3: Change in cued recall accuracy, strategy self-report, VVIQ rating, PFT accuracy and response times for "consistent aphantasics" who scored higher than 73 on the VVIQ. Bold-face entries indicate responses from extreme aphantasic participants who rated 80/80 on the VVIQ

Participant	Change in Accuracy	Strategy self-report	VVIQ (out of 80)	PFT accuracy (out of 20)	PFT response time (seconds)
1	+22.5%	"I chose to use visual imagery or the memory of thinking about it since I have trouble imagining things in my mind."	77	16	8.84
2	+10%	"I did attempt to do as asked for some of the pairs but I also tried to use short phrases to remember alongside the imagery."	76	10	27.77
3	−2.5%	"I tried to remember any word combinations that stood out based on if they made sense together or not or if the words presented were relevant to me."	78	7	6.02
4	+10%	**"I cannot really picture things so I just said the words out loud and tried to create jokes that included both words as they came up."**	80	15	19.65
5	−25%	"Initially I was saying associations out loud and that worked well, then with the imagery it was hard because I have a hard time envisioning things quickly and a lot of the images would have multiple aspects so I would get confused on what I was meaning to associate."	76	12	17.38
6	+10%	**"I tried to find a connection between the two words so I can remember them better."**	80	16	19.25
7	+22.5%	**"I said the words aloud as they appeared in pairs and didn't do the visualization thing."**	80	19	24.61
8	+25%	"In the beginning I was trying to memorize them just by saying them but when you told me to memorize them by thinking of an image with them I would think of a scenario where the two words would go together for example ice cream and mistake would be dropping ice cream."	76	4	14.20
9	+15%	"I can't picture anything in my mind so I couldn't do that, I just kept repeating the words as many times before they disappeared."	77	12	13.55
10	+32.5%	"I attempted to use visual imagery but I can't get a visual imagine in my mind so I just thought of short scenarios of the two words merged together."	78	10	13.69

single free-form question about their strategy-use question. (7) Cued recall direction (forward versus backward) was counterbalanced over all trials, including the practice list. (8) After the strategy-use question (i.e., at the end of the session) a reversed-sense version of the aphantasia recruitment question was administered: "Are you unable to form mental images (i.e., pictures) in your mind's eye?". (9) Distractor trials were identical to experiment 2, except that immediately after the response was entered, the screen was held for 2000-ms fixed period (versus the 1000-ms fixed period in experiment 2).

VVIQ test–retest reliability We analyzed test–retest reliability of the VVIQ between mass questionnaires and the in-session administration, reported on page S19.

Analysis of gender and interactive-imagery effects We obtained data on self-reported gender for participants in experiment 3. These are reported on page S18.

Free-form strategy self-report After the PFT, participants were asked to "describe how you studied the word pairs, whether or not that included the use of visual imagery as instructed, in a short one or two sentence response." These responses were rated by two coders, blinded to condition, for two measures of interest. Firstly, rated either 1) response includes imagery, 2) response explicitly excludes imagery, 3) response leaves open the possibility of imagery but was not explicit. Second, rated for whether it referred to interactivity or connection between words (yes/no). Analyses incorporating these ratings are reported on page S16.
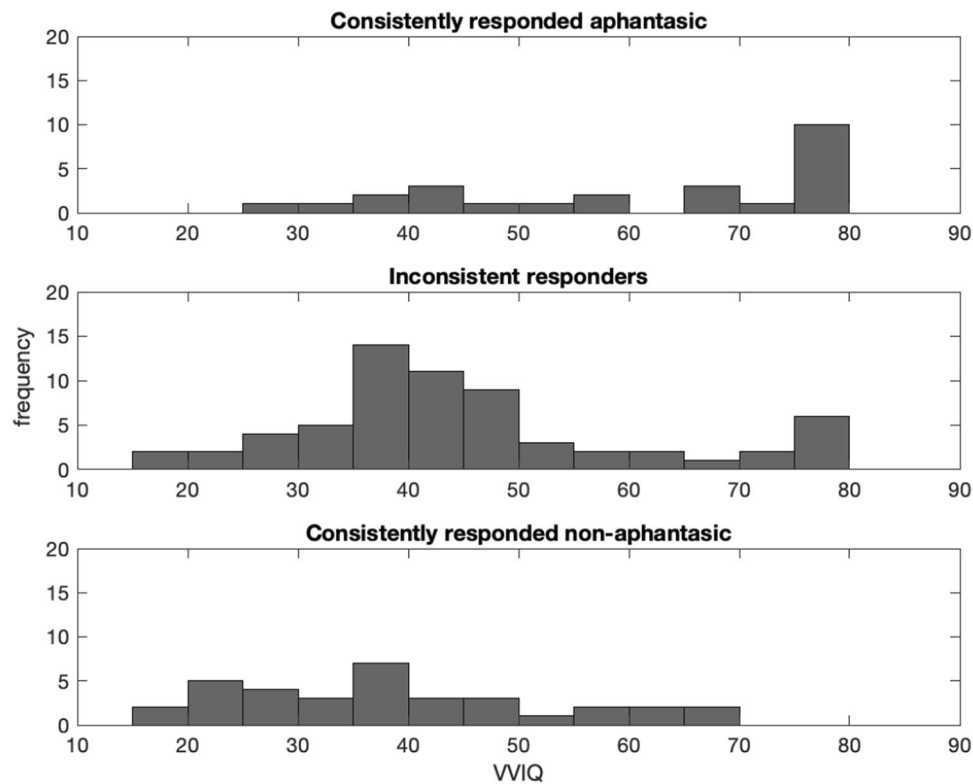
Fig. 4 Experiment 3: Distributions of VVIQ responses for experimental group from experiment 3. Note, lower scores indicate higher vividness.

Results and discussion

Of 122 participants, 25 were consistent aphantasics, 34 were consistent non-aphantasic and 63 were inconsistent responders.

Self-reported vividness Supporting the validity of our yes/no aphantasia self-identification question, consistent aphantasic responders scored significantly higher (lower vividness) than the non-aphantasic group ($p < .001$, Mann–Whitney U test[9]) and the inconsistent responder group ($p < .001$) on the VVIQ, where higher scores indicate lower vividness. The difference between inconsistent responders and consistent non-aphantasic responders nearly reached significance ($p = .07$). Additionally, the average VVIQ rating for consistent aphantasic responders was well above values in experiments 1 and 2 (Table 1). Visual inspection reveals a number of characteristics of the VVIQ responses. First, the inconsistent responders contained participants who exhibited both extremely high and extremely low vividness. Second, a sizeable number of consistent aphantasics nonetheless reported

moderate amounts of vividness in the VVIQ, with ratings within the middle of the VVIQ distribution for consistent non-aphantasics. We do not think that participants are simultaneously reporting an inability to form images (aphantasia question) while reporting vivid mental images (VVIQ). Instead, consistent aphantasics who rated high vividness might have either responded carelessly, or interpreted vividness in terms of the amount of detail within a non-visual representation.

Cued recall A mixed ANOVA on cued recall accuracy (Fig. 2), with design Group (consistent aphantasic, inconsistent responders, consistent non-aphantasics) × Instruction phase (pre-instruction, post-instruction), returned a significant main effect of Instruction phase, $F(1,119) = 91.02$, $MSE = 1.59$, $p < .001$, $n_p^2 = 0.43$, $BF_{inclusion} > 1000$. However, Group, and Group × Instruction phase, were supported null effects (all $p > .5$, $BF_{inclusion} < 0.3$), indicating that aphantasia status did not influence the benefit of interactive-imagery instructions. Additionally, the cued recall accuracy achieved after the imagery instruction in each group was comparable to the imagery group from experiment 1 ($\approx 60\%$), suggesting that the imagery manipulation was successful, and all three groups from experiment 3 would presumably have scored higher than a control group, had it been included.

---

[9] We tested group differences with non-parametric tests due to the skewed vividness rating distribution in the consistent aphantasia group (Fig. 4).

Paper-folding task A one-way ANOVA on PFT accuracy with Group[3] returned non-significant, supported null effect ($p = .52$, $BF_{inclusion} < 0.3$), and likewise for PFT response time ($p = .83$, $BF_{inclusion} < 0.3$). Thus, aphantasic participants did not exhibit worse visuospatial skill, measured objectively, and achieved comparable scores to participants in other experiments (Table 1). These results suggest that the PFT may be added to a class of visuospatial tasks for which aphantasics are fully competent (Zeman et al., 2020), such as mental rotation (Shepard & Metzler, 1973), and the Brooks' matrix spatial task (Brooks, 1967), which we revisit in the general discussion.

The relationship among mental imagery skill, vividness, and the effectiveness of interactive-imagery instructions First, including all participants, VVIQ ratings had a supported null correlation with cued recall accuracy (both $p > .39$, $BF_{10} < 0.30$), and both PFT accuracy and response times had a positive correlation to cued recall accuracy in both instruction phases (Table S9), replicating experiment 1, and with broader coverage of the range of VVIQ values.

Next, we asked whether variability within each group of participants might show different effects. With correlations computed separately for consistent aphantasics, consistent non-aphantasics and inconsistent responders, VVIQ ratings again had a supported null relationship to cued recall accuracy in both instruction phases and all groups ($p > .29$, $BF_{10} < 0.36$), except for inconsistent responders in the pre-instruction phase, $r(61) = -.27$, $p = .03$, $BF_{10} = 1.42$, although the Bayesian correlation was inconclusive. Importantly, VVIQ ratings did not determine the effectiveness of the interactive imagery within the group of consistent aphantasics.

PFT accuracy positively correlated with cued recall accuracy for all three groups and in both instruction phases, and PFT response time had significant positive correlations with cued recall accuracy in both the pre- and post-instruction phases. Thus, skill on this visuospatial task did not predict the effectiveness of interactive imagery even within the consistent aphantasic group.

More conservative criteria for aphantasia Next, we applied increasingly conservative criteria for classification of aphantasics, as described in the Methods. Given the low numbers, these should be interpreted as multiple case studies. Our goal was to check if applying more strict classification criteria would show hints of increased group differences, even while reducing statistical power.

Inconsistent with this, three one-way ANOVAs, with factor Group (VVIQ criterion consistent aphantasics, non-VVIQ criterion consistent aphantasics, inconsistent responders, consistent non-aphantasics) on PFT accuracy, PFT response time and Change in Accuracy returned favored null effects of Group (all $p > .57$, $BF_{inclusion} < 0.3$). Five of the ten VVIQ criterion participants reported, unprompted, difficulty forming visual images. Eight exhibited at least a 10% increase in cued recall following the imagery instruction, with four increasing by 22.5% or more.

Eight participants explicitly reported the use of alternative strategies. It was unclear if participant 1 was referring to mental imagery or not, but described some difficulty with imagining and resorting to "memory of thinking about it". Two participants (7 and 9) reported rote repetition, known to be a poor associative strategy (Bower & Winzenz, 1970), yet still increased substantially (+ 22.5% and + 15%). Two participants did not benefit from the imagery instruction; participant 3 exhibited a small negative change (− 2.5%), participant 5 exhibited a substantial reduction (− 25%) in performance and, interestingly, was the only VVIQ criterion aphantasic who reported trying to implement imagery instructions, suggesting that strict adherence to the imagery instructions may not be beneficial to aphantasics.

Our extreme aphantasics, participants 4, 6, and 7, are of particular interest. Each reported no vividness, were perfectly consistent across multiple administrations of the aphantasia question, and described using non-imagery strategies, consistent with their complete lack of mental imagery. All three benefited from the imagery instruction (+ 10%,+ 10%, and + 22.5%).

In sum, the reduction in sample size was not offset by any hint of an emerging deficit of aphantasics to respond to interactive-imagery instructions, converging with our other evidence against the centrality of visual imagery for interactive-imagery instructions.

General discussion

We replicated the positive effect of interactive-imagery instructions on cued recall (Bower & Winzenz, 1970; Bower, 1970; Paivio, 1969; Paivio & Yuille, 1969; Paivio & Foth, 1970; Richardson, 1985; 1998) compared to control instructions (experiment 1), compared to the no-instruction baseline (all experiments), and compared to the "top-bottom" variant of standard interactive-imagery instructions (experiment 2). Correlations between characteristics of a participant's visual imagery (individual differences in visuospatial skill and vividness) and the effectiveness of interactive imagery produced supported null effects.[10] Furthermore, aphantasics showed no trace of impairment despite their self-diagnosed inability to form visual imagery (experiment 3). Thus, we

---

[10] Additionally, correlations between post-minus-pre instruction memory performance and our visual imagery measures produced supported null effects (page S19)

found no support for the hypothesis that visual images are necessary for interactive-imagery benefits, raising the possibility of alternative explanations.

Curiously, order recognition was not improved by interactive imagery (experiment 1), nor even instructions incorporating order into the image (experiment 2). Whatever additional detail/information is afforded by interactive-imagery instructions evidently does not provide order. Moreover, the relationship between order recognition and cued recall was not influenced by instruction. These results argue against the hypothesis that imagery strategies result in formally different association memories that contain more order. Instead, our results were more consistent with the alternative hypothesis that imagery produces associations that are qualitatively the same as non-imagery conditions.

Subjective vividness does not explain imagery-instruction benefits to cued recall In all three experiments, subjective vividness of mental imagery (VVIQ rating) did not explain the effectiveness of interactive imagery for cued recall. This was reinforced in experiment 3, where aphantasics (high VVIQ) benefited from interactive-imagery instructions as much as others (Fig. 2). All VVIQ-criterion aphantasics that benefited post-instruction reported either solely using non-imagery strategies or a combination of imagery and non-imagery strategies, but evidently with no consequence for their benefit from interactive-imagery instructions. Even three participants who reported exactly no vividness benefited from imagery instructions while reporting using imagery-free strategies. This seems consistent with the observation that congenitally blind participants can effectively apply the method of loci, which is typically described as heavily dependent upon visual imagery (de Beni & Cornoldi, 1985), and with null correlations of the VVIQ with this strategy (Kliegl et al., 1990; Kluger et al., 2022).

Although the VVIQ has been widely used to assess subjective imagery vividness (Marks, 1973), and is a primary way to classify aphantasia (Zeman et al., 2015), there have been specific critiques about its content validity that may be important to consider (McKelvie, 1995; Pylyshyn, 2002). McKelvie (1995) suggested the VVIQ may not capture important dimensions of imagery experience, such as the distinction between imagery vividness and generation. Future studies should focus on qualities of visual imagery experience that the VVIQ may not adequately capture, like imagery generation.

Objective imagery skill does not relate to interactive imagery PFT accuracy did not predict the effectiveness of the interactive-imagery instructions, but covaried with performance even before strategy instructions were given (experiments 1 and 2). Although this does not rule out the PFT as a measure of other memory processes like working

memory or visuospatial ability, it weakens the argument that imagery skill determines success with interactive-imagery instructions.

Interestingly, there was a supported null difference between PFT performance in aphantasics and non-aphantasics in experiment 3, which may place the PFT in a class of visuospatial tasks that aphantasics perform without any clear deficits (Zeman et al., 2010). Both Zeman et al., (2010) and Bainbridge et al., (2021) suggested that aphantasics use symbolic/verbal strategies for visuospatial tasks. Thus, the cognitive processes required for this task may not necessarily depend on visual images, which suggests a dissociation between conscious mental imagery experience and the cognitive processes engaged when solving complex visuospatial problems. Furthermore, because the PFT could not explain the benefits of interactive imagery, its intact status in aphantasics cannot explain why aphantasics showed virtually no reduced benefit from these instructions.

Validity of aphantasia-status classified by self-report Our three criteria for classifying aphantasia in experiment 3 (multiple consistent responses to the aphantasia recruitment question, and two VVIQ cutoffs), produced prevalence rates that approached the estimates in previous studies (see methods), suggesting that methods of classifying aphantasia in experiment 3 aligned well with previous aphantasia studies. Despite this, there are broader critiques of classifying aphantasia by self-report. For example, de Vito and Bartolomeo (2016) suggested aphantasics may underestimate a latent ability to form mental images. Perceived absence of mental imagery experience may then be due to poor/altered metacognition rather than fundamental differences in cognitive representations. However, even if aphantasia is due to an inaccurate sense of one's own imagery ability, our findings still show that this kind of imagery self-efficacy is immaterial to memory-success following interactive-imagery instructions, again problematic for the hypothesis that interactive imagery acts through the formed image, itself.

Interactive-imagery effects without visual imagery Our findings challenge the notion that visual imagery, in any literal sense, is essential for the benefit to cued recall of interactive-imagery instructions. In other words, the subjective experience of mental imagery is experienced by those who are able, but is not required for later memory benefits. This resonates with Pylyshyn's (2002) argument that the experience of mental imagery may be epiphenomenal, and not necessarily causal.

A similar story is emerging from recent research on word concreteness/imageability effects. High-imageability words are recalled better low-imageability words (Paivio, 1969). Hockley (1994) found better associative recognition for higher concreteness word pairs. Paivio and colleagues

explained concreteness as providing participants the greater availability to construct visual image mediators for concrete/imageable than abstract/low-imageable words, confirmed by findings of more frequent self-reported use of imagery strategies during the study of high imageability word pairs (Paivio et al., 1968; Paivio and Yuille, 1969). Thus, the historical understanding of the concreteness/imageability effects is functionally linked to visual imagery-related strategies like interactive imagery.

However, behavioral and neuroimaging findings have challenged the idea that concreteness effects can be explained via visual imagery. Westbury et al., (2013) and Westbury, Cribben, and Cummine (2016) showed that concreteness effects on lexical decision could be explained by non-imagery factors like size/density of a word's context and its emotional associations (see Fiebach and Friederici (2004), and see Cox, Hemmer, Aue, and Criss (2018) who found semantic diversity, alongside concreteness, to be a strong predictor of memory performance). In neuroimaging studies, one can look for memory-related activity in brain regions that are involved in mental imagery, such as posterior visual-processing regions and right-lateralized activity. However, Caplan and Madan (2016) found no brain activity reminiscent of visual imagery explaining word-imageability effects on cued recall (see also Klaver et al., 2005). Rather, higher imageability was associated with more hippocampal activity (somewhat left-dominant), which in turn, apparently increased memory. Similarly, Duncan, Tompary, and Davachi (2014) found that functional connectivity between hippocampus and ventral tegmental area during interactive-imagery instructions predicted retrieval success, regions that are not specialized for imagery.

An alternative explanation of interactive-imagery effects Vincente and Wang (1998) emphasized the idea that expert-memory effects depend on participants engaging with stimuli in a manner that is relevant to their expert domain. Extrapolating to non-expert domains, perhaps interactive-imagery acts primarily by inspiring participants to engage with word pairs in a manner that leads to this kind of meaningful or deep processing.

But what is the nature of this deeper processing, and how does it improve memory? Some hints may be gleaned from experiment 2. Standard-imagery and actor-object imagery both resulted in benefits to memory. Given the high similarity between the examples given for both instructions, both instructions may have engaged the same mechanisms, perhaps revealing some role of motor imagery (Allen et al., 2022; Yang et al., 2021) in interactive-imagery effects. In contrast, top-bottom instructions which ask participants to imagine a spatially organized image including both words, and do not explicitly refer to the words interacting, did not change cued recall or associative recognition from baseline.

Top-bottom imagery may be difficult to implement, especially for certain word pairs. For example, it is easier to conceptualize a spatially organized image of APPLE DRAGON, compared to ASPECT LEVEL (both of which were possible pairings in our study); however, this challenge would also exist with standard and actor-object strategies (concreteness effects; cf. Hockley 1994; Paivio 1969). Alternatively, top-bottom instructions may miss a key component— explicit instructions to conceptualize an interactive, functional relationship between the items. Top-bottom imagery may resemble explicitly non-interactive "separation-imagery" instructions, where participants are asked to form mental images of each word in isolation, which does not improve association-memory (Bower, 1970; Dempster & Rohwer, 1974; Hockley & Cristi, 1996).

In contrast, by leading participants to think about an interactive relationship between words, effective associative strategies like interactive imagery may facilitate encoding of additional item features that are pair-unique. To illustrate how this may occur, consider an associative recognition task for the pairs APPLE TEACHER and TABLE OVEN. An image (or non-visual analogue) of a TEACHER with an APPLE (intact, here) may generate a stereotypical image of a crisp, red apple on a teacher's desk, whereas an image of an OVEN with an APPLE (recombined, here) might bring to mind baked apples. The more a participant focuses on how the words might interact, the more detailed and pair-specific the stored representations might be (see the modelling work of Caplan, Chakravarty, & Dittmann, in press, Cox & Criss 2017, 2020, and Benjamin, 2010). For example, Cox and Criss (2020) showed how similarity can cause the representations of two items to become correlated, by drawing attention to their common features. One intriguing possibility is that interactive imagery amplifies this very same effect by drawing the participant's attention to shared features.

Supporting encoding of more detailed item representations, item recognition improves alongside associative memory performance, when comparing interactive imagery to rote repetition (Dempster & Rohwer, 1974; Hockley & Cristi, 1996).[11] Such a mechanism could conceivably occur without visual imagery. This is consistent with findings that verbally mediated strategies for association-memory (e.g., form a sentence including both words) are nearly as effective (Dunlosky et al., 2005; Hockley & Cristi, 1996).

---

[11] Both Hockley and Cristi (1996) and Dempster and Rohwer (1974) also found that separation imagery improved item recognition, suggesting that interactivity is not *required* to encode more detailed item representations. However, the additional item features granted by non-interactive strategies would likely not be pair-specific, which may explain the lack of effects on associative memory.

Interactive-imagery instructions do not change model-relevant characteristics of the association Largely replicating and extending the boundary conditions of Kato and Caplan (2017), order recognition significantly correlated with cued recall accuracy, but significantly weaker than the correlation between associative recognition and cued recall (Figs. S11, S15, and S16). Despite large effects on association-memory, imagery instructions did not modulate these findings (Figs. S11, S15, and S16). Whatever additional detail/information is afforded by imagery instructions does not improve memory for order. An interesting possibility here is that order and associative information are somehow represented differently in memory, explaining why manipulations of association-memory do not affect memory for order. Cox and Criss (2020) suggested order could be represented by item features distinct from associative features. In any case, our findings indicate that challenges to perfect-order models, which predict a perfect relationship between order recognition and cued recall, and order-absent models, which predict no relationship, are not particular to uninstructed participants, but generalize to several instructed strategies. This increases the need for models that can accommodate moderate-level order within associations.

Conclusions

Interactive-imagery instructions improve associative memory without requiring vividness, visual-imagery skill, nor even the subjective sense that one can create visual imagery. The instruction may instead lead participants to conceptualize elaborate, interactive relationships, leading to storage of more distinctive features. Finally, whatever additional detail aids associative memory does not provide order.

References

Allen, R.J., Waterman, A.H., Yang, T., & Jaroslawska, A.J. (2022). Working memory in action: Remembering and following instructions. In R.H. Logie, Z. Wen, S.E. Gathercole, N. Cowan, & R.W. Engle (Eds.) *Memory in science for society: There is nothing as practical as a good theory*: Oxford University Press.

Anderson, J.A. (1970). Two models for memory organization using interacting traces. *Mathematical Biosciences*, *8*, 137–160.

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.

Benjamin, A.S. (2010). Representational explanations of process dissociations in recognition: The dryad theory of aging and memory judgments. *Psychological Review*, *117*(4), 1055–1079.

Bower, G.H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, *9*, 529–533.

Bower, G.H., & Winzenz, D. (1970). Comparison of associative learning strategies. *Psychonomic Science*, *20*, 119–120.

Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Bridges, D., Pitiot, A., MacAskill, M., & Pierce, J. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*.

Brooks, L.R. (1967). The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology*, *19*, 139–159.

Caplan, J.B., Chakravarty, S., & Dittmann, N.L. (in press). Associative recognition without hippocampal associations. *Psychological Review*.

Caplan, J.B., & Madan, C.R. (2016). Word-imageability enhances association-memory by increasing hippocampal engagement. *Journal of Cognitive Neuroscience*, *28*(10), 1522–1538.

Cox, G.E., & Criss, A.H. (2017). Parallel interactive retrieval of item and associative information from event memory. *Cognitive Psychology*, *97*(5), 31–61.

Cox, G.E., & Criss, A.H. (2020). Similarity leads to correlated processing: A dynamic model of encoding and recognition of episodic associations. *Psychological Review*, *127*(5), 792–828.

Cox, G.E., Hemmer, P., Aue, W.R., & Criss, A.H. (2018). Information and processes underlying semantic and episodic memory across tasks, items, and individuals. *Journal of Experimental Psychology: General*, *147*(4), 545–590.

Criss, A. H., & Shiffrin, R. M. (2005). List discrimination in associative recognition and implications for representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1199–1212.

de Beni, R., & Cornoldi, C. (1985). The effects of imaginal mnemonics on congenitally total blind and on normal subjects. In D. F. Marks, & D. Russell (Eds.) *Imagery*, (Vol. 1 pp. 56–59). Dunedin, N.Z.: Human Performance Associates.

de Vito, S., & Bartolomeo, P. (2016). Refusing to imagine? On the possibility of psychogenic aphantasia. A commentary on Zeman et al. (2015). *Cortex*, 334–335.

Dempster, F.N., & Rohwer, W.D. (1974). Component analysis of the elaborative encoding effect in paired-associate learning. *Journal of Experimental Psychology*, *103*(3), 400–408.

Duncan, K., Tompary, A., & Davachi, L. (2014). Associative encoding and retrieval are predicted by functional connectivity in distinct hippocampal area CA1 pathways. *Journal of Neuroscience*, *34*(34), 11188–11198.

Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology*, *41*(2), 389–400.

Engelkamp, J. (1991). Imagery and enactment in paired-associate learning. In R. H. Logie, & D.M (Eds.) *Mental images in human cognition* (pp. 119–128). Amsterdam: North Holland Press.

Engelkamp, J. (1995). Visual imagery and enactment of actions in memory. *British Journal of Psychology*, *86*, 227–240.

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

Fiebach, C.J., & Friederici, A.D. (2004). Processing concrete words: fMRI evidence against a specific right-hemisphere involvement. *Neuropsychologia*, *42*(1), 62–70.

Foer, J. (2011) *Moonwalking with Einstein: The art and science of remembering everything*. New York, NY: Penguin Press.

Fraundorf, S.H., Diaz, M., Finley, J., Lewis, M.L., Tooley, K.M., Isaacs, AM., & Brehm, L (2014). CogToolbox for MATLAB[computer software]. CogToolbox for MATLAB [computer software]. Retrieved from. http://www.scottfraundorf.com/cogtoolbox.html

French, J.W., Ekstrom, R.B., & Price, L.A. (1963) *Kit of reference tests for cognitive factor*. Princeton, NJ: Educational Testing Service.

Friendly, M., Franklin, P.E., Hoffman, D., & Rubin, D.C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods Instruments, & Computers*, *14*, 375–399.

Geller, A.S., Schleifer, I.K., Sederberg, P.B., Jacobs, J., & Kahana, M.J. (2007). PyEPL: a cross-platform experiment-programming library. *Behavior Research Methods*, *39*(4), 950–958.

Gesualdo, F. (1592). Plutosofia. Padua.

Greene, R.L., & Tussing, A.A. (2001). Similarity and associative recognition. *Journal of Memory and Language*, *45*, 573–584.

Hautus, M.J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods Instruments, & Computers*, *27*(1), 46–51.

Hintzman, D.L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods Instruments, & Computers*, *16*(2), 96–101.

Hockley, W.E. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition*, *22*(6), 713–722.

Hockley, W.E., & Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition*, *24*, 202–216.

JASP Team (2021). JASP(Version 0.15) [Computer Software]. JASP(Version 0.15) [computer software]. Retrieved from https://jasp-stats.org/

Kato, K., & Caplan, J.B. (2017). Order of items within associations. *Journal of Memory and Language*, *97*, 81–102.

Kelly, M.A., Blostein, D., & Mewhort, D.J.K. (2013). Encoding structure in holographic reduced representations. *Canadian Journal of Experimental Psychology*, *67*(2), 79–93.

Keogh, R., & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, *7*, 53–80.

Klaver, P., Fell, J., Dietl, T., Schür, S., Schaller, C., Elger, C.E., & Fernández, G. (2005). Word imageability affects the hippocampus in recognition memory. *Hippocampus*, *15*, 704–712.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in psychtoolbox-3? *Perception*, *36*(14), 1–16.

Kliegl, R., Smith, J., & Baltes, P.B. (1990). On the locus and process of magnification of age differences during mnemonic training. *Developmental Psychology*, *26*, 894–904.

Kluger, F. E., Oladimeji, D. M., Tan, Y., Brown, N. R., & Caplan, J. B. (2022). Mnemonic scaffolds vary in effectiveness for serial recall. *Memory*.

Konrad, B.N. (2013) *Superhirn - Gedächtnistraining mit einem Weltmeister*. Goldegg Verlag: Vienna.

Kounios, J., Bachman, P., Casasanto, D., Grossman, M., Smith, W., & Yang, R.W. (2003). Novel concepts mediate word retrieval from human episodic associative memory: evidence from event-related potentials. *Neuroscience Letters*, *345*, 157–160.

Kounios, J., Smith, R.W., Yang, W., Bachman, P., & D'Esposito, M. (2001). Cognitive association formation in human memory revealed by spatiotemporal brain imaging. *Neuron*, *29*, 297–306.

Kucera, H., & Francis, W. (1967) *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press.

Maguire, E.A., Valentine, E.R., Wilding, J.M., & Kapur, N. (2003). Routes to remembering: the brains behind superior memory. *NatureNeuroscience*, *6*(1), 90–95.

Marks, D.F. (1972). Individual differences in the vividness of visual imagery and their effect on function. In P.W. Sheehan (Ed.) *The function and nature of imagery* (pp. 83–106). New York: Academic Press.

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

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.

Metcalfe Eich, J. (1982). A composite holographic associative recall model. *Psychological Review*, *89*(6), 627–661.

Müller, N.C.J., Konrad, B.N., Kohn, N., Muñoz-López, M., Czisch, M., Fernández, G., & Dresler, M. (2018). Hippocampal–caudate nucleus interactions support exceptional memory performance. *Brain Structure and Function*, *223*(3), 1379–1389.

Murdock, B.B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609–626.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *76*(3), 241–263.

Paivio, A., & Foth, D. (1970). Imaginal and verbal mediators and noun concreteness in paired-associate learning: The elusive interaction. *Journal of Verbal Learning and Verbal Behavior*, *9*, 384–390.

Paivio, A., Smythe, P.C., & Yuille, J.C. (1968). Imagery versus meaningfulness of nouns in paired associate learning. *Canadian Journal of Psychology*, *22*, 427–441.

Paivio, A., & Yuille, J.C. (1969). Changes in associative strategies and paired-associate learning over trials as a function of word imagery and type of learning set. *Journal of Experimental Psychology*, *79*(3), 458–463.

Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., & Lindeløv, JK (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Pelli, D.G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Plate, T.A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks*, *6*(3), 623–641.

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

Richardson, J.T.E. (1985). Converging operations and reported mediators in the investigation of mental imagery. *British Journal of Psychology*, *75*, 205–214.

Richardson, J.T.E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, *5*(4), 597–614.

Sanchez, C. (2019). The utility of visuospatial mnemonics is dependent on visuospatial aptitudes. *Applied Cognitive Psychology*, *33*, 519–529.

Shepard, R.N., & Metzler, J. (1973). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

Shiffrin, R.M., & Steyvers, M. (1997). A model for recognition memory: REM— Retrieving Effectively From Memory. *Psychonomic Bulletin & Review*, *4*, 145–166.

Sivashankar, Y., & Fernandes, M.A. (2021). Enhancing memory using enactment: does meaning matter in action production? *Memory*, *30*, 147–160.

Vincente, K.J., & Wang, J.H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review*, *105*(1), 33–57.

Westbury, C. F., Cribben, I., & Cummine, J. (2016). Imaging imageability: Behavioral effects and neural correlates of its interaction

with affect and context. *Frontiers in Human Neuroscience, 10*, 346.

Westbury, C.F., Shaoul, C., Hollis, G., Smithson, L., Briesemeister, B.B., Hofmann, M.J., & Jacobs, A.M. (2013). Now you see it, now you don't: on emotion, context, and the algorithmic prediction of human imageability judgments. *Frontiers in Psychology, 4*(991), 1–13.

Yang, J., Zhao, P., Zhu, Z., Mecklinger, A., Fang, Z., & Han, L. (2013). Memory asymmetry of forward and backward associations in recognition tasks. *Journal of Experimental Psychology: Learning Memory, and Cognition, 39*(1), 253–269.

Yang, T., Allen, R. J., Waterman, A. H., Zhang, S., Su, X., & Chan, R. C. K. (2021). Comparing motor imagery and verbal rehearsal strategies in children's ability to follow spoken instructions. *Journal of Experimental Child Psychology, 203,* 10533.

Yates, F.A. (1966) *The art of memory*. Chicago: University of Chicago Press.

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

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

Zeman, A., Milton, F., Della Sala, S., Dewar, M., Frayling, T., Gaddum, J., & Winlove, C. (2020). Phantasia - the psychological significance of lifelong visual imagery vividness extremes. *Cortex, 130*, 426–440.