THE IMPACT OF DRAWING AND TRACING ON VISUAL LONG-TERM MEMORY

THE IMPACT OF DRAWING AND TRACING ON VISUAL LONG-TERM MEMORY

A Dissertation Presented

by

ANNA C. MCCARTER

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2025

Psychological and Brain Sciences

THE IMPACT OF DRAWING AND TRACING ON VISUAL LONG-TERM MEMORY

A Dissertation Presented

by

ANNA C. MCCARTER

Approved as to style and content by:

Jeffrey Starns, Chair

Kyle Cave, Member

Rebecca Spencer, Member

Torrey Trust, Member

Jeffrey Wammes, Member

Ilia Karatsoreos, Chair
Psychological and Brain Sciences

ACKNOWLEDGMENTS

I am so lucky to have had many wonderful people walk alongside me through this process. First, I would like to thank my dissertation committee members, Jeff Starns, Kyle Cave, Bekki Spencer, Torrey Trust, and Jeff Wammes for their thoughtful advice on this project. Each of you has made contributions that have led to a stronger dissertation. I would like to give a special thanks to Jeff Starns, Kyle Cave, and Bekki Spencer for also advising my comprehensive exam. Those questions helped to form this work.

Next, I would like to thank my mentors from over the years. To begin, I want to thank Dr. Fleetwood, my high school neuroscience teacher, who introduced me to this field and also led me through my first research project. Thank you for making both the neuroscience class and junior research project such positive experiences. I would next like to thank Tim Vickery, my undergraduate and master's advisor, for taking a chance on little freshman me, for believing in my research abilities more than I did, and for trusting me with my own projects. Working in your lab was the most impactful part of my time at UD. Next, I would like to thank Rosie Cowell and Dave Huber for giving me the opportunity to do my PhD here at UMass and for a great first year. Thank you to Denise Pope, Johanna Yunker, and Amanda Hamel for teaching me so much about teaching. Your mentorship and support have led me to my passion as an educator. And lastly, I'd like to thank Jeff Starns for being a fantastic advisor. Your incredible depth of statistical knowledge, flexibility, kind and prompt advice, and relaxed demeanor have helped me grow as a researcher and person.

So many people have brought fun to the last four years. Thank you to all of my lab mates: Jerome, John, Mar, Melisa, Michael, Natasha, Sean, and Tejas. You all have provided lots of research advice but also laughter and encouragement. Thank you to the game night crew: Clara D., Firat, Fran, Kuan-Jung, Melisa, Qing, Sean, Tejas, Trina, Tori, and many others, for bringing so much joy and laughter to my time at UMass. Our Thursday nights together were always an island of calm and happiness in the midst of busy weeks. Thank you to my cohort mates, Aanchal, Jenna, Tori, and Yun, for providing such consistent encouragement along this journey. I am fortunate to have gone through this process with you all. Thank you to my friends outside the program, Bridget, Clara W., Emily, Ethan, Hannah, Jaji, Rebecca, and many others for checking in regularly and reminding me how much more there is to life than grad school.

Lastly, thank you to my family for their endless support, for being incredibly understanding, and for believing in me. Thank you to my grandparents, Uncle Glenn, and my sisters (Paula & Laura) for their regular Zoom calls and motivating messages. I would especially like to remember Granddad for inspiring my work in the field of memory. Most of all I would like to thank my parents, Scott and Donna, and my partner, Eric, for their ceaseless encouragement, love, and support through all that the last four years have brought. I couldn’t have done this without you.

ABSTRACT

THE IMPACT OF DRAWING AND TRACING ON VISUAL LONG-TERM MEMORY

MAY 2025

ANNA C. MCCARTER

B.S., UNIVERSITY OF DELAWARE

M.S., UNIVERSITY OF DELAWARE

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Jeffrey Starns

This dissertation explores the impact of drawing and tracing on subsequent visual long-term memory of object images. Aim 1 tested whether producing a secondary representation (through describing) or gaining additional visual experience (through drawing) was more beneficial for learning visual materials. We found that performance was considerably better for images that participants described and drew compared to items that were studied, with a slight advantage of describing over drawing. This contrasts with studies of verbal content in which drawing leads to significantly better memory than describing.

Aim 2 explored how describing, drawing, and tracing impact memory precision. In this experiment, participants viewed images of objects and either copied the image, traced the image, described the image, or simply studied the image. They then completed a 3-alternative forced choice test with very visually similar alternatives. We found that memory was much better for items that were drawn and described compared to items that were traced or studied, with a benefit of drawing over describing. This shows that, in

comparison to Aim 1, when the goal is very detailed visual memory, drawing is superior to describing.

Aim 3 and Aim 4 of this dissertation explored how drawing and tracing impact the binding of object features in memory. In these experiments, participants viewed images of colored objects in a specific location in a box and either copied the image, traced the image, described the image, or simply studied the image. To test binding, participants were presented with an image of the object and selected the color and location of it. Aim 3 was conducted with younger adults and Aim 4 was conducted with older adults. For young adults, describing and drawing led to the best memory binding. For older adults, describing and studying were optimal.

Overall, this work shows that drawing and describing are useful strategies for learning images compared to simply studying. In addition, it emphasizes that the optimal learning strategy depends on the content, the test format, and the age group.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GENERAL INTRODUCTION

Visual memory is hugely important. It is foundational for work life (e.g. remembering what mug is yours, where you left your laptop, and which colleagues you have met before) and personal life (e.g. remembering what your childhood bedroom looked like, where to put the pieces for your favorite board game, and which baby at daycare is yours). Unfortunately, memory can be quite fallible, especially for older adults and people with disorders such as Alzheimer's Disease. However, there is currently considerable research exploring potential ways to bolster memory (e.g., Brod, 2021). In particular, drawing and tracing are potential means of improving visual memory (e.g., Ainsworth & Scheiter, 2021).

1.1 Movement & Cognition

There is evidence that movements can improve memory (Glenberg et al., 2013). For instance, acting phrases out leads to better memory compared to imagining the phrase (Bäckman & Nilsson, 1985), hearing and seeing the phrase (Bäckman et al., 1986), or seeing someone else act out the phrase (Engelkamp & Zimmer, 1997). Similarly, handwriting leads to better memory than typing (Longcamp et al., 2005; Longcamp et al., 2006; Longcamp et al., 2008), and feeling shapes leads to better memory than just seeing them (Kalenine et al., 2011). Drawing, a focus of this dissertation, has also been found to benefit memory (Van Meter & Garner, 2005; Fernandes et al., 2018; Fan et al., 2023).

One potential reason why movements are beneficial for memory is because they produce a motor-based memory (Engelkamp & Zimmer, 1989). Several fMRI studies provide support for this idea. For instance, Kato and colleagues (1999) had participants learn Chinese characters through repetitive writing and then undergo an fMRI while either covertly or overtly retrieving the characters. Activity was found in the presupplementary motor area, intraparietal sulcus, and left premotor cortex in the covert retrieval condition, indicating that even when no overt writing was needed, retrieving the characters automatically activated motor representations. Longcamp and colleagues (2003) also explored this idea. They had participants undergo an fMRI while first viewing real letters or pseudo-letters and then writing the real letters and pseudo-letters. Activation in the left ventral premotor area was found for viewing real letters but not for viewing pseudo-letters. Activation in this same area was found when participants wrote the letters and pseudo-letters. This suggests that viewing the real letters (which had previously been written) automatically activated the associated motor representations. Roberts and colleagues (2024) explored this idea with drawing. They had participants learn words through either drawing, writing, or describing. Participants then did a recognition test while undergoing an fMRI. They found that the memory boost for drawn items compared to written items was positively correlated with activity in the premotor and supplementary motor areas. This suggests that remembering the drawn items led to the reactivation of motor representations.

Additional evidence for the idea of a motor memory comes from an ingenious behavioral study by Wammes and Fernandes (2017). They had participants learn words through either reading or acting them out. Then participants had a recognition test along with a semantic or motor secondary task. With the semantic secondary task, participants remembered significantly more words that were acted out than words that were read. In contrast, with the motor secondary task, participants remembered an equivalent amount of enacted and read words. This suggests that acting out is beneficial in part because it adds a motor memory component that is robust against semantic interference.

Overall, incorporating movements into learning benefits memory. This is likely due, at least in part, to memory for the motor movements. In this work, the movements we will focus on are drawing and tracing. Some work has been done to explore the benefits of drawing and tracing with verbal and conceptual materials.

1.2 Drawing and Tracing to Learn Verbal and Conceptual Materials

Drawing and tracing are commonly used in the educational domain to improve memory for conceptual and verbal information (Ainsworth et al., 2011; Ainsworth & Scheiter, 2021; Cromley et al., 2020; Fan, 2015). Studies of the impact of drawing and tracing on learning have spanned all ages from 3-year-olds (Hulme et al., 1987) to older adults (Meade et al., 2020a; Meade et al., 2018). In addition, these studies span a variety of educational domains - physics, chemistry, biology, mathematics, vocabulary, letter learning, mechanics, anthropology, earth systems, and anatomy.

1.2.1 Drawing

The benefits of drawing have been explored in many contexts. For instance, several studies have found a benefit of drawing a depiction of a passage compared to simply reading the passage (Balemans et al., 2016; Dean & Kulhavy, 1981; Hall et al., 1997; Leutner et al., 2009; Schmeck et al., 2014; Schwamborn et al., 2010; Schwamborn et al., 2011; Snowman & Cunningham, 1975; Van Meter, 2001; Van Meter et al., 2006). To extend those findings, several studies have found a benefit of drawing a depiction of a passage or video compared to writing an explanation of the passage or video (Alesandrini, 1981; Bobek & Tversky, 2016; Edens & Potter, 2003; Gobert & Clement, 1999; Leopold & Leutner, 2012; Leopold et al., 2013; Scheiter et al., 2017), though Kulhavy et al. (1985) and Ploetzner & Fillisch (2017) found evidence to the contrary. Some groups have also found a benefit of drawing compared to writing for learning vocabulary (Jalava et al., 2023; Meade et al., 2020a; Meade et al., 2020b; Meade et al., 2018; Meade et al., 2019; Roberts & Wammes, 2021; Tran & Fernandes, 2024b; Wammes et al., 2016; Wammes et al., 2017; Wammes et al., 2018a; Wammes et al., 2018b), though Tirre et al. (1979) found a benefit of writing compared to drawing. Additional evidence for the benefit of drawing for conceptual learning has been found with word problems (Van Essen & Hamaker, 1990), a chemistry visualization (Zhang & Linn, 2011), a chemistry animation (Zhang & Linn, 2013), anatomy (Alsaid & Bertrand, 2016; Backhouse et al., 2017), and a mechanical model (Bobek & Tversky, 2016).

1.2.2 Tracing

The benefits of tracing have also been explored in many ways. For instance, several groups have found a benefit of tracing sample math problems compared to simply viewing the problems (Agostinho et al., 2015; Ginns et al., 2016; Hu et al., 2014; Hu et al., 2015). In addition, research has shown a benefit of tracing compared to viewing when learning letters (Bara et al., 2004; Bara et al., 2007; Hulme et al., 1987; Tzeng et al., 1986), words (Ofman & Shaevitz, 1963; Roberts & Coleman, 1958), and diagrams (Macken & Ginns, 2014; Tang et al., 2019). Interestingly, two studies have directly compared drawing and tracing. These studies found that drawing is more effective than tracing for learning about a physics animation (Mason et al., 2013) and word-image pairs (Wammes et al., 2019).

As seen in this section, many studies have been conducted exploring the benefits of drawing and tracing when learning verbal and conceptual materials. Though the results are mixed, the vast majority of studies in this area found a benefit of drawing and tracing.

1.3 Drawing and Tracing to Learn Visual Materials

While there is a wealth of studies exploring the impact of drawing and tracing with conceptual and verbal materials, relatively few studies have investigated the impact of drawing and tracing with visual materials. This section will outline all of the studies that have used visual materials, beginning with letters, letter-like symbols, and abstract figures before moving on to studies that used pictures or line drawings of common objects and places.

1.3.1 Letters

A few studies have explored the impact of drawing and tracing on letter learning. While letters may be considered verbal materials, in this context, the goal was to learn letter shapes, which emphasizes the visual aspect of letters. Thus, studies of letters are included within this visual section. To begin, Hirsch and Niedermeyer (1973) had kindergarteners learn letters through either copying or tracing. Students were then given a letter formation test in which they had to copy letters. They were also given a letter discrimination test in which they had to pick which letter matched a model. Children who learned the letters through copying did significantly better than the tracing group on the letter formation test. This would be predicted by transfer-appropriate processing theory (Morris et al., 1977) since the practice and test are the same task. Interestingly, there was no significant difference between the groups on the letter discrimination test. This suggests that copying is beneficial for memory when the test involves copying the letters, but that on a more simplistic test, copying and tracing are equivalent.

Zemlock and colleagues (2018) also tested the potential benefit of drawing letters. They had preschoolers learn letters and numbers by either copying them or viewing them. Students were given a letter matching test before and after study in which they had to match letters with exemplars. They were also given a 4AFC letter recognition test before and after study in which they were presented with a handwritten letter and had to select which letter was written. There was no significant difference between the groups on the letter matching test. However, the copying group improved significantly more on the recognition test than the viewing group. This suggests that both drawing and viewing are beneficial, but when faced with a more difficult task, drawing leads to better performance.

In alignment with the studies on Roman letters, Experiment 4 of Naka and Naoi (1995) had undergraduate students learn Arabic letters through either viewing or copying. Students were then given free recall and recognition tests. Students who learned the letters through copying performed numerically better on both tests. Similarly, Kim and colleagues (2024) had participants learn English-Korean word pairs by either saying the Korean characters or drawing the Korean characters. They were then visually presented with two sets of Korean characters and needed to pick the correct one. Memory was better for characters learned through drawing than through speaking. Overall, these studies suggest that drawing and tracing are useful strategies to use when learning letter shapes since in all cases, memory following drawing and tracing is equivalent or better than following viewing alone.

1.3.2 Letter-Like Symbols and Abstract Figures

Several groups have explored the impact of drawing and tracing on letter-like symbols and abstract figures. For instance, Hulme (1979) had children learn abstract figures by either tracing them or pointing at them. On a subsequent recognition test, items learned through tracing were remembered significantly better than items learned through pointing. This suggests that tracing is better for memory than pointing. Experiment 4 of Hulme et al. (1987) replicated these findings with abstract letter-like shapes. Additional evidence comes from Askov and Greff (1975) who had children learn letter-like symbols by either copying them or tracing them. Students were then given a cued recall test. Students who copied did significantly better than students who traced. This suggests that copying is even better for memory than tracing.

To further corroborate that result, Lansing (1981) had kindergarteners learn an abstract figure by either drawing, tracing, or viewing. Students were then tested two days later and 30 days later with a cued recall test and a 5AFC recognition test. Students who drew did significantly better than students who traced on both tests at both time points. Students who traced did significantly better than students who viewed on both tests at both time points. This mirrors the results of Hulme (1979), Hulme et al. (1987), and Askov and Greff (1975). Similarly, Experiments 1-3 of Naka & Naoi (1995) involved undergraduate students learning abstract figures by either viewing or copying. Students were then given free recall and recognition tests. Performance was numerically better on both tests for items learned through copying compared to items learned through viewing. Interestingly, three of the studies with abstract figures found null results on a delayed test. First, Experiment 2 of Berman (1939) involved children learning abstract figures by either tracing or viewing. Students were then given a 5AFC recognition test both immediately and at a 24-hour delay. Tracing led to significantly better performance on the immediate test but there was no difference between the groups on the delayed test. Similarly, Gonzalez and colleagues (2011) had university students learn complex lines through either copying or tracing. Students were then given a cued recall test immediately and one week later. On the immediate test, drawings were more accurate for the tracing group than the copying group. On the delayed test, there was no difference in drawing accuracy between the groups. Lastly, Tsutsui and colleagues (2017) had adults learn an abstract figure through either copying or tracing. Participants were then given a free recall test immediately and three days later. On the immediate test, the copying group did better than the tracing group. On the delayed test, there was no difference between the groups. Overall, these studies suggest that drawing and tracing are beneficial, but that the benefits may not last long, though Lansing (1981) did find benefits on a delayed test.

Some groups have used more unique conditions to test the impact of drawing and tracing. For instance, Naka (1998) conducted three experiments exploring what strategies are optimal for children learning abstract figures. In Experiment 1, students either copied or viewed. On a subsequent free recall test, students who copied did better than students who viewed. In Experiment 3, students either traced or copied without marks appearing (drew with the wrong end of the pen). On a subsequent free recall test, students who copied without marks performed better than students who traced. In Experiment 4, students either copied, traced, or viewed. On a subsequent free recall test, students who copied did better than students who traced or viewed. Students who traced and students who viewed performed equivalently. These experiments emphasize the benefit of copying over tracing and viewing.

Peynircioğlu (1989) also used unique conditions to explore the optimal strategies for university students learning abstract figures. Experiment 3 compared copying a figure with drawing the figure based on a description. On a free recall test, items learned through drawing were remembered significantly better than items learned through copying. Experiment 4 compared connecting the dots to form a figure with tracing the figure. On a free recall test, items learned through connecting the dots were remembered significantly better than items learned through tracing. Overall, the studies presented in this section suggest that drawing and tracing are beneficial for learning abstract shapes, with drawing generally superior to tracing.

1.3.3 Pictures and Line Drawings of Common Objects or Scenes

Lastly, some research has explored drawing and tracing with pictures and line drawings of common objects or scenes. For instance, Experiment 2 of Wammes et al. (2018a) involved undergraduate students learning pictures of concrete nouns through either drawing or writing the label of the object. On a subsequent recognition test, the hit rate was significantly higher for objects that were drawn compared to objects that were labeled. This suggests that drawing is a useful strategy for remembering visual materials. In Experiments 1 and 2 of Peynircioğlu (1989), they had university students learn line drawings of objects and scenes. In Experiment 1, students either rated the quality of the drawing or got a description of it and had to draw it. On a subsequent free recall test, items learned through drawing were remembered significantly better than items learned through rating. In Experiment 2, students either rated the quality of the drawing, copied the drawing, or got a description of it and had to draw it. On a subsequent free recall test, items that were drawn and items that were copied were remembered significantly better than items that were rated. In addition, items that were drawn were remembered significantly better than items that were copied. These experiments highlight that drawing is a useful way to learn visual materials.

Experiment 1a of Levin et al. (1975) also involved learning visual stimuli. In this experiment, children learned line drawings of common objects by either verbally labeling them, viewing them, or imagining them. The children were then given a recognition test. Children who labeled did significantly better than children who viewed or imagined. Importantly, in Experiment 1b, they had children learn the same stimuli by drawing them in the air. On the subsequent recognition test, there was no significant difference between students who drew (in 1b) and students who labeled (in 1a). This suggests that drawing is a beneficial way to learn images. To follow up on that finding, Levin et al. (1977) had children learn line drawings of common objects by either drawing them in the air, tracing them with their finger, or with no specific instructions. On a subsequent recognition test, the group that drew did significantly better than the group that traced. There was no significant difference between tracing and not receiving instructions. This experiment suggests that drawing is more beneficial than tracing in terms of subsequent memory.

Overall, these prior studies emphasize that drawing and, in some cases, tracing are beneficial strategies to boost memory for visual materials. However, several questions remain. First, none of these studies have explored how writing a description and drawing compare in terms of remembering visual materials. Second, it is unclear how drawing and tracing compare in terms of the precision of subsequent visual memory. Lastly, it would be interesting to know how drawing and tracing compare in terms of the binding of object features in subsequent visual memory. These three questions are explored in this dissertation.

CHAPTER 2

AIM I: DUAL-CODE THEORY

Many researchers claim that drawing is beneficial in large part because it creates a representation in a new modality so that participants have both verbal and visual representations (Fiorella & Zhang, 2018; Kulhavy et al., 1985; Scheiter et al., 2017; Schmidgall et al., 2019; Van Meter & Garner, 2005; Wammes et al., 2016). However, the prior studies have all explored the benefit of adding a visual representation when learning verbal content. The goal of Aim 1 is to explore whether drawing or describing is more beneficial for learning visual content. If having both verbal and visual representations is optimal, then describing the image will be better than drawing. If drawing has other inherent benefits, then drawing may be better than describing.

2.1 Background

Dual-code theory proposes that having representations in both verbal and visual modalities is beneficial for memory (Paivio, 1991). In the context of learning verbal materials, this means that producing a visual representation is beneficial. One way to create a visual representation is by drawing. Several studies suggest that creating this visual representation is one of the key benefits of drawing (Fiorella & Zhang, 2018; Scheiter et al., 2017; Schmidgall et al., 2019; Wammes et al., 2016). For instance, Schmidgall and colleagues (2019) had participants learn a passage by either drawing, viewing pictures, writing a summary, viewing an animation, imagining, or reading the text alone. At test, they found that drawing and viewing pictures led to better memory than writing or text only. This suggests that having a visual representation is important. In addition, they found that learning through drawing and viewing an animation was better than imagining, suggesting that having an external visual representation is optimal.

Two theories have been proposed that tie together dual code theory and drawing. First, Kulhavy and colleagues (1985) put forth the Conjoint Retention Hypothesis which aimed to explain why map learning is optimized by having both a visuospatial depiction (either provided or drawn) and a written description (either provided or written). This theory states that encoding information in both spatial and verbal modalities strengthens memory because the participant can retrieve information from either modality. In an effort to understand why drawing is beneficial for learning verbal content more broadly, Van Meter and Garner's Generative Theory of Drawing Construction (2005) suggests that the integration of existing verbal information with newly generated visual information is key. Overall, these theories suggest that drawing is beneficial for memory in large part because it creates a visual representation to complement the verbal representation.

This emphasis on the benefit of having both verbal and visual representations leads to the assumption that describing may be more beneficial than drawing when learning visual materials. However, verbal overshadowing suggests that describing can actually be detrimental to visual memory (Schooler & Engstler-Schooler, 1990). In the Schooler study, participants viewed a crime scene video and either wrote a description of it or did an unrelated task. Then, the participants were presented with eight similar faces and needed to pick the person who committed the crime. Interestingly, the group that wrote a description showed worse memory. This would suggest that given the potential negative impact of describing, drawing would be more beneficial when learning visual materials. However, it is important to note that Meissner & Brigham (2001) outline several unsuccessful attempts to replicate the verbal overshadowing findings. In summary, dual code theory and verbal overshadowing provide contradictory predictions about which condition (drawing or describing) will produce optimal memory.

In order to explore the question of whether drawing or describing is more beneficial for learning visual materials, we built on the methods of Meade and colleagues (2018). They had participants learn words by drawing or describing and found that drawing led to significantly better memory than describing. The current study extended Meade et al. (2018) by using visual stimuli instead of verbal stimuli. This study is theoretically important because it parses out the conflicting predictions of dual code theory and verbal overshadowing. In addition, this study is practically relevant since it elucidates what strategy is optimal for remembering visual materials.

2.2 Methods

2.2.1 Participants

We based our sampling plan on recovery simulations that used an unequal-variance signal detection model, and we chose recognition memory differences from Test Half 1 as the key outcome. (Test Half 2 had more trials per condition, so estimation should be more accurate for those analyses. The reasoning behind having two halves of the test will be explained later.) To generate simulated data, we first defined signal-detection parameters for each simulated participant by sampling from distributions that spanned a wide range of plausible values. Thus, the simulations incorporated across-participant variability in memory performance and response strategies (e.g., the overall tendency to say “old” or the overall tendency to use high versus low confidence levels).

We chose levels of variability that seemed plausible in light of previous recognition memory studies. The parameter values for the distribution function can be seen in Table 1.

Table 1

Parameter values for the simulation distribution function for Aim 1.

Parameter	Mean	Standard Deviation
Center	1.7	0.35
Draw Adjustment	0.4	0.3
Describe Adjustment	0	0.3
View Adjustment	-0.4	0.3
Log Sigma	log(1.25)	0.05
C1 Offset	-0.3	0.2
C2 Offset	0	0.2
C3 Offset	0.3	0.2

For each simulated participant, we sampled recognition responses from the signal-detection model using the participant's sampled parameter values and the trial counts for Test Half 1 (15 items per condition). We fit each simulated set of response frequencies using the maximum-likelihood method. The primary goal was to estimate the difference in memory for the studied items, indexed by the signal detection measure d_a, between all pairs of learning conditions. d_a measures the extent to which participants can distinguish targets and lures. It is similar to d' but accounts for the possibility of unequal variance in target and lure memory (Macmillan & Creelman, 2004). We regard a d_a difference of 0.2 as the minimum meaningful value; that is, we are not interested in interpreting d_a differences less than 0.2 as important findings. As such, the key outcomes that we evaluated were 95% uncertainty intervals on the difference in d_a between pairs of learning conditions, and our goal was to have uncertainty intervals that are no wider than 0.4, such that the interval should exclude 0 as a hypothesized effect size if the estimated effect size is greater than 0.2 in absolute value (i.e., if the interval width is less than 0.4, then the limits of the uncertainty interval should be less than 0.2 away from the effect size point estimate in both directions).

We ran many simulations with varying sample sizes, and this process suggested that meeting our desired level of precision for estimating effect sizes was likely to require 50-75 participants. Specifically, a sample size of 50 produced uncertainty intervals at our desired level of precision about half the time, and a sample size of 75 achieved the desired level of precision in over 90% of the simulated experiments. Given this, we created the following stopping rule: run 50 participants, then check the confidence interval widths and continue until all of the interval widths are less than 0.40, up to a maximum of 75 participants.

Participants were undergraduate students from the University of Massachusetts Amherst who participated for course credit. Participants had to be 18 years or older and have normal or corrected-to-normal vision. Fifteen participants were excluded for doing no task or the wrong task on three or more trials (our preregistered exclusion criteria). Four participants were excluded because their hit rate was not at least 0.1 greater than their false alarm rate (our preregistered exclusion criteria). Participants were excluded as data collection proceeded and only usable datasets were counted towards the stopping rule.

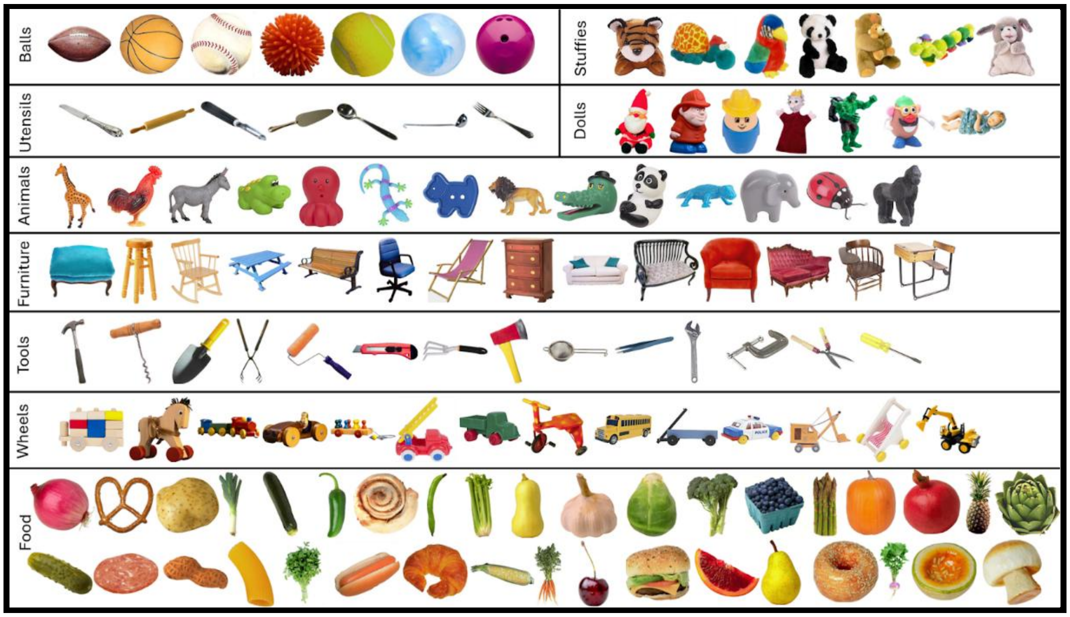
The credible interval widths were still above 0.40 when checked at 53 participants, so we decided to proceed to 75 participants. We ended up with 80 usable participants due to additional sign-ups on the final day. Participants ranged in age from 18 to 23 years with a mean of 19.4 years. The sample was made up of 21 males and 59 females.

2.2.2 Materials

Stimuli were 120 colored images of common objects selected from the Brady et al. (2008) database. All of the images used in this experiment can be seen in Figure 1. For each participant, 85 images were randomly selected to be learned and the other 35 images were used as lures in the recognition test, with 45 of the images learned in Test Half 1 (*Describe*, *Draw*, and *Study*) and the other 40 learned in Test Half 2 (*Describe* and *Draw*). The reasoning behind having two sections of the test will be explained later. In order to make the task more challenging, objects from the same categories were assigned to each item type. Each condition (Test Half 1 *Describe*, Test Half 1 *Draw*, Test Half 1 *Study,* Test Half 2 *Describe,* Test Half 2 *Draw*) and each set of lures (Test Half 1 lures, Test Half 2 lures) was made up of one ball, one stuffed animal, one kitchen utensil, one doll, two animals, two pieces of furniture, two tools, two wheeled toys, and three food items. Since Test Half 1 had 15 items per stimulus set and Test Half 2 had 20 items per stimulus set, each Test Half 2 stimulus set had an additional five food items. Which images from each category were assigned to each item type was randomized across participants.

Figure 1

All of the object images used in Aim 1.



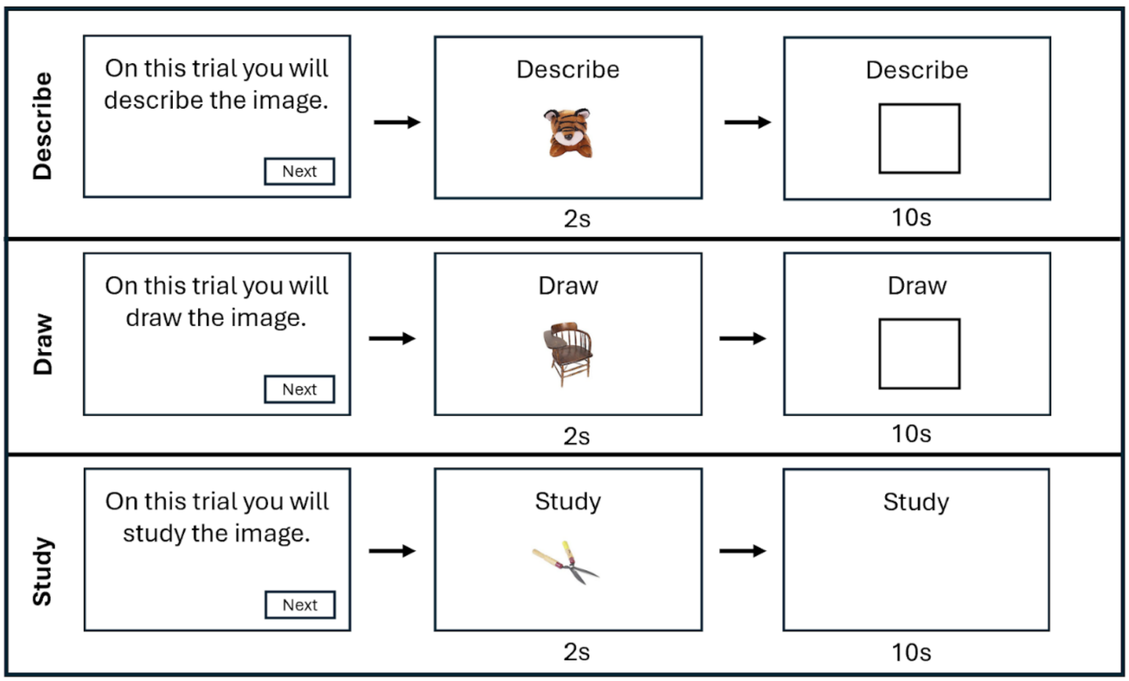
2.2.3 Procedure

Participants completed this experiment on a tablet (Samsung Galaxy Tab S6 Lite) using a stylus. The experiment began with instructions on how to use the stylus and one practice trial of each task (*Describe*, *Draw*, and *Study*). Based on Meade et al. (2018), for the *Describe* task, the participant was instructed to list physical descriptions of the object. They wrote these descriptions on the tablet. For the *Draw* task, the participant was instructed to draw a copy of the image in the space provided on the tablet. For the *Study* task, the participant was instructed to visually rehearse the image. To ensure that participants were not tracing or drawing in the *study* condition, they were instructed to sit with their hands on their lap for the duration of these trials.

Following this instructional phase, the learning phase began. This was a within-participant design with three conditions (*Describe, Draw*, *Study*). Each trial began with instructions on what task they would be doing on that trial. Then, the image was presented for two seconds. The image then disappeared, and the participant had ten seconds to describe, draw, or mentally rehearse the item based on the instructions. They were told to continue their task until the time was up. This repeated for all 85 learned images. See sample trials in Figure 2.

Figure 2

Sample learning phase trials of the Describe, Draw, and Study tasks for Aim 1.

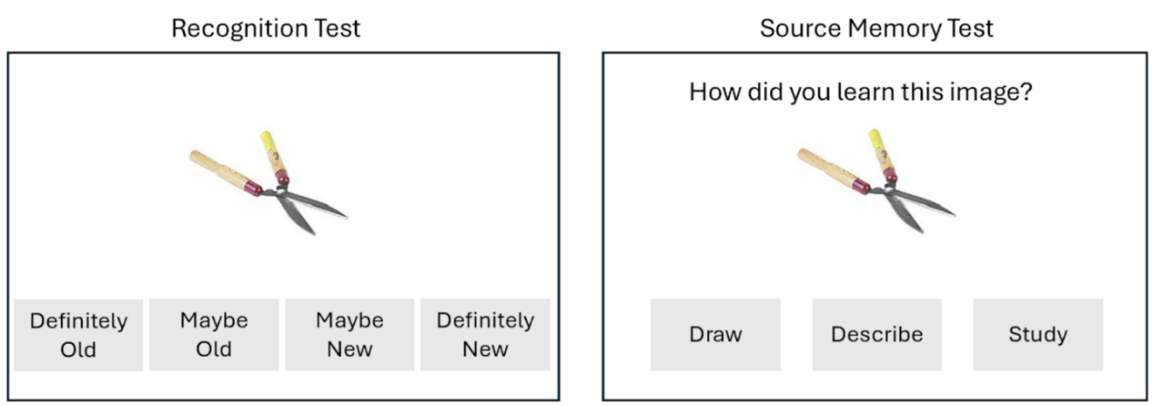


Following the learning phase, a brief filler task occurred. This ensured that participants were not relying on working memory in the recognition test. The filler task involved categorizing 240 scene images as a city, beach, or mountain. This filler task was selected because it involves both a verbal and visual component, ensuring that we were not selectively impairing the *Draw* or *Describe* condition. This took approximately 5 minutes to complete.

Experiment 3 of Meade et al. (2018) used a remember-know-new recognition test in which remember meant that the participant could recall what task they completed for that word, know meant that the participant felt that the word was familiar but could not recall what task they did for it, and new meant that the participant did not remember learning that word. In order to minimize confusion, this experiment separated the recognition judgment and source memory judgment. First, the participant was presented with each image and asked if they learned it or not (old/new). To incorporate their confidence, the four options were “definitely old,” “maybe old,” “maybe new,” and “definitely new.” If the participant indicated that they did remember learning the object (definitely old or maybe old), they were then presented with a source memory judgment asking which condition it was learned in (*Draw*, *Describe*, or *Study*). Both of these judgments were unspeeded decisions. See sample test trials in Figure 3. Across the entire test phase (both halves), participants went through all 120 object images: 85 targets and 35 lures. The images within each test half were presented in a random order.

Figure 3

Sample test trials for Aim 1.



Previous studies have shown that when participants false alarm in a source memory test, they typically attribute the memory to the condition that is perceived as weaker (e.g., Hoffman, 1997). For instance, if their memory is weaker for the describing task, they would be expected to attribute more false alarms to the *Describe* condition. With the *Describe, Draw*, and *Study* conditions, *Study* would likely be the weakest. This does not provide additional information about the comparison between describing and drawing. In order to allow for that comparison, there were two halves to the test. In the first half, the options for the source memory test were *Describe, Draw, and Study* as shown in Figure 3. For the second half, the participants were informed that all of the remaining images were either described, drawn, or not learned and so the options for the source memory test were only *Describe* and *Draw*. This affords us a clearer estimate of whether the *Describe* condition or the *Draw* condition is considered weaker.

At the very end of the experiment, the participants were asked 'Do you typically draw one or more times per week?' and 'Have you had any drawing training?' This allows us to explore whether drawing is more beneficial for people who draw regularly.

2.2.4 Open Practices Statement

This study was preregistered on OSF (https://osf.io/96z3f). In this article, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data, analysis code, and stimuli are available on OSF (https://osf.io/9s5cz/). Data were analyzed using R (R Core Team, 2021).

2.2.5 Model

We used a Bayesian unequal variance signal detection model to estimate d_a and, critically, the difference in d_a between pairs of learning conditions. This model was fit to data from each participant for both Test Half 1 and Test Half 2. The model estimated signal detection parameters for each participant as well as parameters that described the group distributions across participants. The parameters for an individual participant defined the relative mean and standard deviation of the memory evidence distributions for target and lures, which in turn were mapped to d_a values, and the cutoffs for each of the four possible recognition test responses (“Definitely Old”, “Maybe Old”, “Maybe New”, and “Definitely New”). These parameters defined the predictions for each participant by seeing what proportion of the memory distributions fell within each of those response bins, thereby taking into consideration how conservative the participant was (the position of the cutoffs). Instead of having a separate estimate for each of the four conditions, we created a Mu Center and adjustments for each of the three conditions to account for some people being better or worse at the task overall. Performance in each of the conditions is then correlated with that participant’s Mu Center. See Table 2 for the parameters defining the prior distributions.

Table 2

Parameter values for the prior distribution for Aim 1.

Parameter & Description	Mean	Standard Deviation
Mu Center: The center of the memory distribution	Normal distribution with mean of 1 and standard deviation of 2, truncated with a minimum of 0	Normal distribution with a mean of 0.50 and a standard deviation of 0.20, truncated with a minimum of 0.10
Mu Adjust Describe: How much to adjust memory distribution for described items	Normal distribution with mean of 0 and standard deviation of 1	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10
Mu Adjust Draw: How much to adjust memory distribution for drawn items	Normal distribution with mean of 0 and standard deviation of 1	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10
Mu Adjust Study: How much to adjust memory distribution for studied items	Normal distribution with mean of 0 and standard deviation of 1	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10

Table 2 – continued

Parameter & Description	Mean	Standard Deviation
LSig: The standard deviation of the target memory distributions	Normal distribution with mean of log(1.25) and standard deviation of 0.10	Normal distribution with mean of 0.125 and standard deviation of 0.10, truncated with a minimum of 0.05
C: An index of how conservative the participant was	Normal distribution with mean of 0 and standard deviation of 2	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10
D1: The adjustment for the upper cutoff	Normal distribution with mean of 0.75 and standard deviation of 2, truncated with a minimum of 0	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10
D2: The adjustment for the lower cutoff	Normal distribution with mean of -0.75 and standard deviation of 2, truncated with a maximum of 0	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10
Cut Shift: How much to adjust the cutoffs in Test Half 2	Normal distribution with mean of 0 and standard deviation of 0.30	Normal distribution with mean of 0.20 and standard deviation of 0.10, truncated with a minimum of 0.05
Mu Shift: How much to adjust the mean memory distribution for Test Half 2	Normal distribution with mean of 0 and standard deviation of 0.30	Normal distribution with mean of 0.20 and standard deviation of 0.10, truncated with a minimum of 0.05

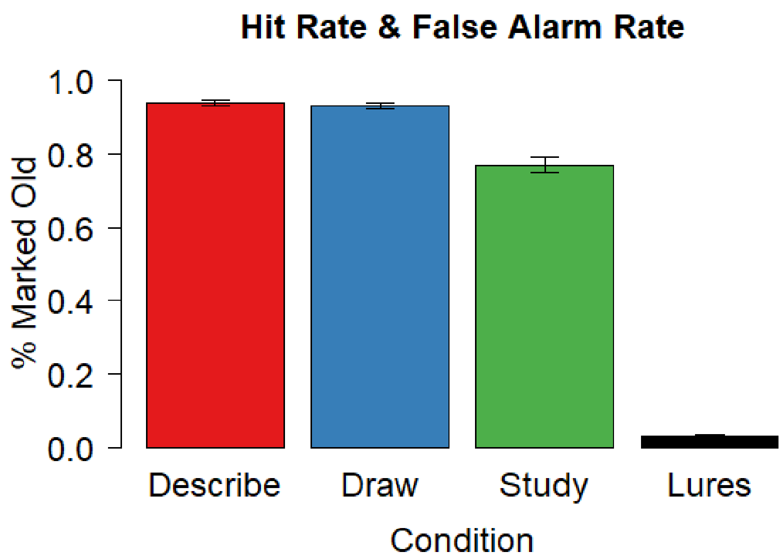
2.3 Results

2.3.1 Recognition

Overall recognition performance was quite high. The hit rate for each condition and the false alarm rate can be seen in Figure 4. The hit rate for the described items was 93.9%, for the drawn items was 93.2%, and for the studied items was 77.0%. The false alarm rate was 2.80%.

Figure 4

The hit rates and false alarm rate for Aim 1.



Note. Error bars represent standard error.

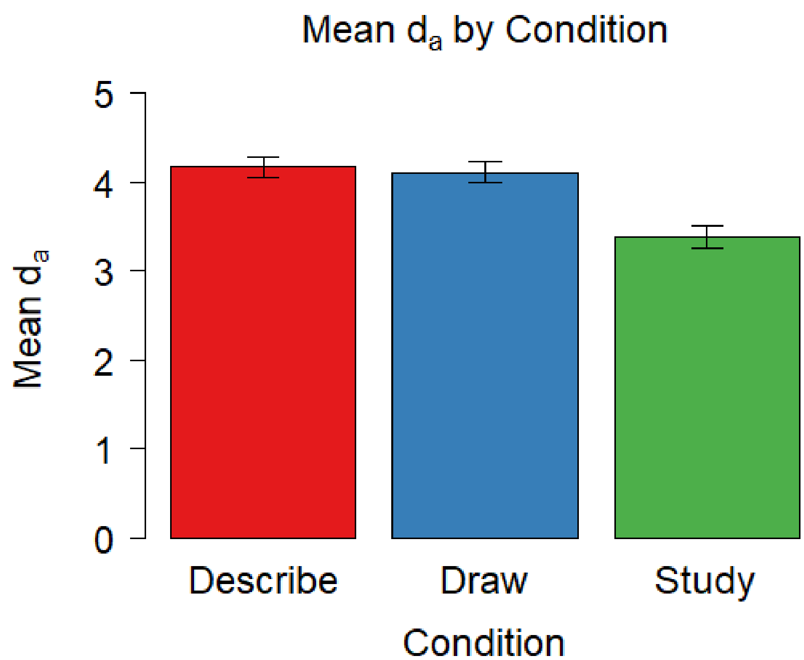
As an initial check, frequentist analyses were conducted. Median d_a values were calculated for each condition for each participant from the Bayesian model posterior distribution. These averages can be seen in Figure 5. Importantly, this was a simplified version of the Bayesian model that did not include participant distributions. This is because in the original Bayesian model, the values for each participant are correlated with each other, thus violating independence. The ANOVA of d_a values across conditions was significant ($F(2, 237) = 13.41$, $p < 0.001$). According to two-tailed[1] paired samples *t*-tests, the d_a for described items was significantly greater than the d_a for studied items ($t(79) = 11.34$, $p < 0.001$, 95% CI [0.66, 0.94]) and the d_a for drawn items was

[1] In the preregistration, we indicated doing one-tailed *t*-tests for the comparisons involving *Study* since we expected that *Study* would always be the worst condition. We chose to do two-tailed *t*-tests instead since *Study* was not always the worst condition and so that we could present 95% confidence intervals.

significantly greater than the d_a for studied items ($t(79) = 9.63$, $p < 0.001$, 95% CI [0.58, 0.88]). The two-tailed paired samples *t*-test comparing the described items and the drawn items showed no significant difference ($t(79) = 1.11$, $p=0.27$, 95% CI [-0.05, 0.18]). This suggests that describing and drawing lead to better memory than studying and that both of these strategies provide similar levels of improvement.

Figure 5

Median d_a value by condition for Aim 1.

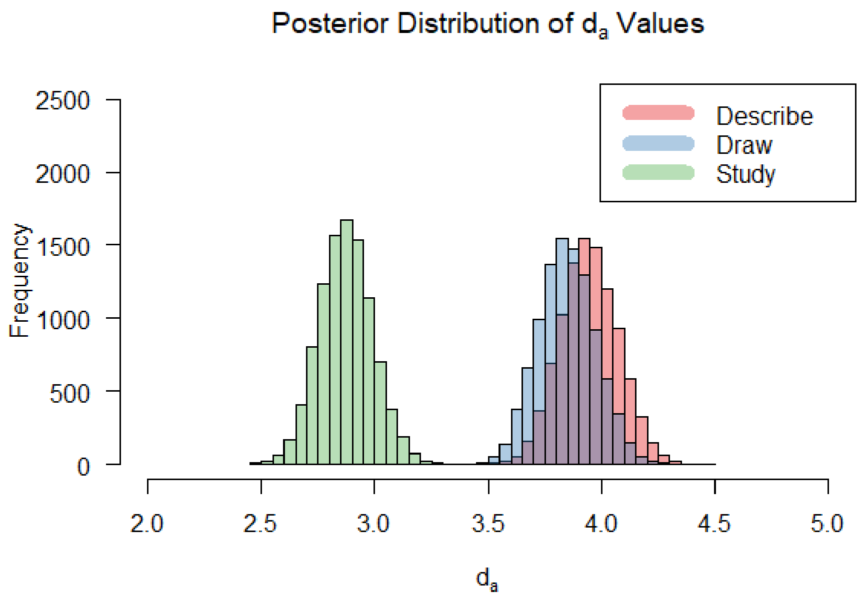


Note. Error bars represent the standard error.

Subsequently, a fully Bayesian analysis was conducted. A posterior distribution of across-participant average d_a values by condition is shown in Figure 6. The 95% credible interval for the *Describe-Study* comparison is 0.88 to 1.26, for the *Draw-Study* comparison is 0.78 to 1.17, and for the *Describe-Draw* comparison is -0.07 to 0.26. The histograms and credible intervals show a considerable advantage of the *Describe* and *Draw* conditions over the *Study* condition. As seen in the histograms, there is substantial overlap between the *Describe* and *Draw* conditions, and the credible interval includes zero, so there is not a large difference between *Describe* and *Draw*, though *Describe* does seem to have a slight advantage.

Figure 6

Posterior distribution of d_a values by condition for Aim 1.

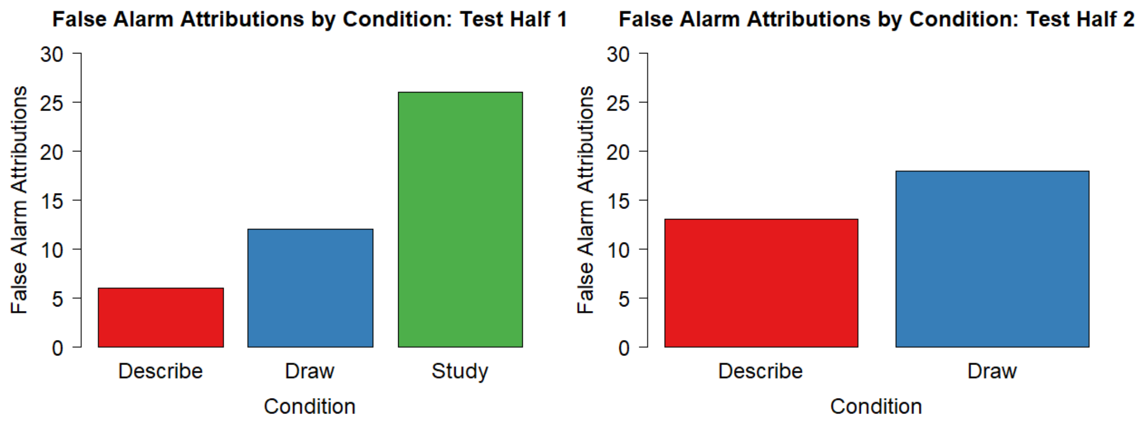


2.3.2 False Alarm Attributions

Overall, there were very few false alarms in this experiment. About half of the participants (n = 42) had no false alarms. Given this, our original plan of analyzing the percentage of false alarms attributed to each condition by participant would not be valid. So instead, we summed the number of false alarms attributed to each condition across all participants (Figure 7).

Figure 7

Total false alarm attributions by condition across all participants for Test Half 1 (left) and Test Half 2 (right) for Aim 1.



As an initial test, we conducted a multinomial test which found significant differences between the attributions for Test Half 1 and chance levels (p = 0.001). Subsequent binomial tests found that attributions to the *Describe* condition were significantly lower than chance (p = 0.006), attributions to the *Draw* condition were not significantly different from chance (p = 0.43), and attributions to the *Study* condition were significantly higher than chance (p < 0.001). In addition to these frequentist analyses, we used a Bayesian Dirichlet model to get a point estimate and confidence interval for each condition. For *Describe*, the estimated percentage of false alarm attributions is 14.4% with a 95% credible interval from 6.4% to 26.3%. For *Draw*, the estimated percentage of false alarm attributions is 27.3% with a 95% credible interval from 16.0% to 41.1%. It is worth noting that there is considerable overlap between the 95% credible intervals for *Describe* and *Draw*. For *Study*, the estimated percentage of false alarm attributions is 57.5% with a 95% credible interval from 43.2% to 71.1%. Overall, this shows that for Test Half 1, more false alarms are attributed to *Study* with not much of a difference between *Describe* and *Draw*.

We conducted a binomial test comparing the attributions for Test Half 2 and chance level which found no significant difference ($p = 0.47$). In addition to this frequentist analysis, we used a Bayesian beta-distribution model to get a point estimate and credible interval for each condition. For *Describe*, the estimated percentage of false alarm attributions is 42.2% with a 95% credible interval from 26.4% to 59.4%. For *Draw*, the estimated percentage of false alarm attributions is 57.7% with a 95% credible interval from 40.6% to 73.6%.[2] Since these credible intervals include 50%, there does not seem to be a significant difference between the number of misattributions to *Describe* and to *Draw* in Test Half 2.

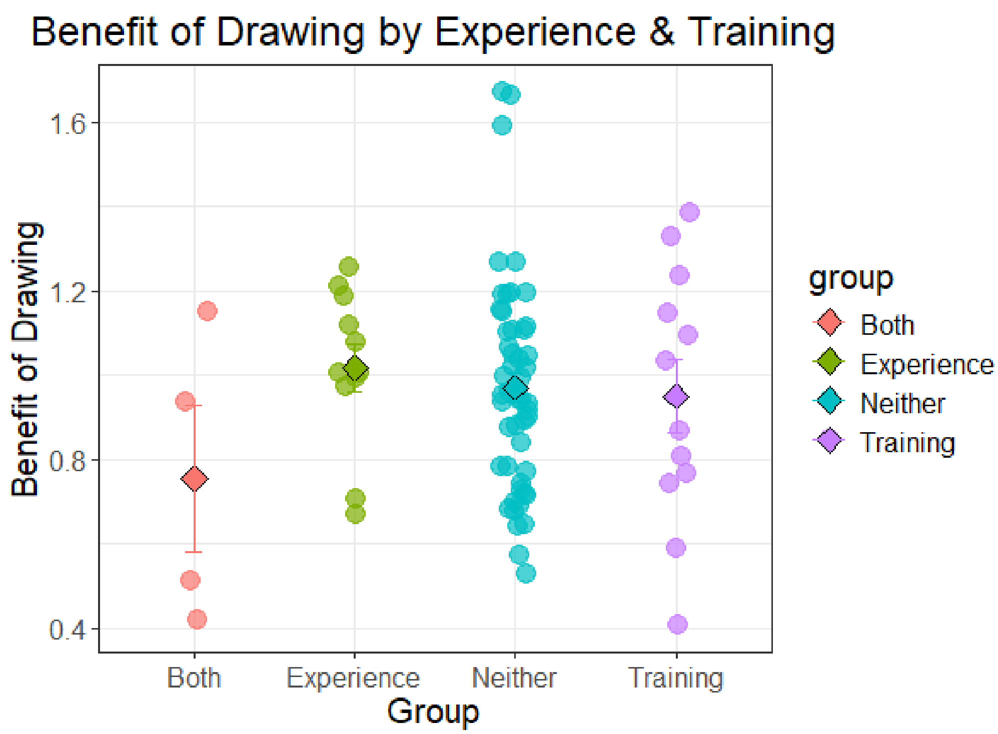
2.3.3 Training & Experience

At the end of the experiment, we asked participants whether they had any drawing training and whether they draw on a regular basis. Most participants ($n = 53$) had neither training nor experience, 12 had training, 11 had experience, and 4 had both training and experience. The measure of interest was the benefit of drawing (median d_a draw - median d_a study). These data are plotted in Figure 8. To achieve more equal samples, the participants with training and/or experience were merged into one group ($n = 27$). Unexpectedly, there was no significant difference in drawing benefit between those who had training and/or drawing experience and those who did not ($t(48.28) = 0.35$, $p = 0.73$).

[2] Note that, given the two-option task, the *Draw* point estimate is complementary to the *Describe* point estimate, within rounding error. Moreover, the lower bound of the *Describe* interval is the complement of the upper bound of the *Draw* interval, and vice versa.

Figure 8

The benefit of drawing (median d_a draw - median d_a study) based on the participants' drawing experience and training for Aim 1.



Note. Error bars represent the standard error.

2.4 Discussion

These results show that both drawing and describing are beneficial for memory in comparison to studying, with a slight benefit of describing over drawing. These results suggest that verbal overshadowing was not occurring here and that instead, multi-modal representations were beneficial for memory.

It is interesting how these results contrast with those of Meade and colleagues (2018), which this study mirrored. They found that memory was significantly better for the words that were learned through drawing than the words that were learned through describing. Interestingly, we found very little difference between performance for drawn images and described images, with a slight advantage of describing over drawing. This suggests that the optimal strategy depends on the type of content that is being learned.

We anticipated that having drawing experience and/or training would be beneficial since these advantages are likely to promote more accurate drawings, which tend to be associated with better memory (Schwamborn et al., 2010). However, we surprisingly found that drawing is not more beneficial for people who have drawing experience and/or training compared to those who do not. This result aligns with another study that found no correlation between drawing history and the benefit of drawing (Wammes et al., 2016). This lack of a benefit may be due to the brief duration that participants were given to draw. Ten seconds is enough time for a quick sketch but not a detailed drawing. If people with drawing experience and/or training were attempting to create a thorough drawing, they would have run out of time.

There are some limitations to this study. First, memory performance was quite good, with very few false alarms. This makes it more challenging to differentiate between the conditions. To account for that, Aim 2 uses a difficult three-alternative forced choice test and a two-day delay between learning and test. Second, participants could correctly respond in the recognition test based on remembering only a brief description of the item. For instance, they could remember seeing a red ball and then respond correctly at test since there were not highly similar lures. With the three-alternative forced choice test in Aim 2, participants will need detailed visual memory of the item in order to respond correctly. Thus, Aim 2 will expand on this study by testing the precision of the participants' visual memory.

CHAPTER 3

AIM II: PRECISION

Aim 2 sought to explore whether the benefit of describing found in Aim 1 would persist if the test relied on memory for precise visual details. The difficult 3AFC test used in this experiment could not be accurately completed based on a brief description alone. Given this, we expected drawing to outperform describing in this study. In addition, tracing was added as another condition in this experiment since tracing involves a lot of attention paid to the details of the shape outline. Thus, the goal of Aim 2 was to explore how describing, drawing, tracing, and studying images compare in terms of the precision of the resulting visual memories.

3.1 Background

On the surface, drawing and tracing appear to be quite similar tasks. However, there are some key differences that may lead to better memory precision for objects that are traced in comparison to objects that are drawn. To start, tracing involves continuous feedback. As the participant traces over the object shape, they can see where their pen strokes deviate from the model. This may lead to more detailed information about the exact shape of the object. In contrast, drawing lacks this continuous feedback since a drawing is spatially separate from the model. Another key difference between drawing and tracing is the eye and hand movements involved. Gowen and Miall (2006) found that eye-hand movements are more tightly coupled during tracing than drawing, likely due to the continuous monitoring of the pen strokes against the model. The eyes and hands remain near the model during tracing, but while drawing, they may go back and forth between the model and participant-produced drawing. Since overt visual attention tends to follow where the eyes are focused, tracing may lead to more attention to the model than drawing. In addition, where the hands are located is also prioritized in terms of attention (Abrams et al., 2008; Reed et al., 2006). In tracing, the hands are located at the model, whereas in drawing, the hands are located at the participant-produced drawing which may not be as accurate. This increased attention to the model during tracing could lead to more precise memories of the shape of the object in comparison to drawing. Studying the object without performing a secondary task could be comparable to tracing if the participant closely follows the shape with their eyes, but without explicit instruction to do so, studying likely involves more cursory studying of the object. In addition, the hands are not near the model, which might further decrease the attention paid to it.

The question of how motor engagement impacts subsequent precision has been explored in a few studies. Fan and colleagues (2020) had participants repeatedly draw real-world objects while getting an fMRI. Across this training, neural evidence showed enhanced discriminability between the neural representations of these objects. Similarly, Baratz and colleagues (2023) had participants trace abstract shapes. They found enhanced neural discrimination for shapes learned through tracing in comparison to shapes learned through watching a video of the shape being traced or viewing the shape image alone. These studies provide evidence that motor engagement leads to more precise neural representations of the objects, but neither of these studies measured if these neural changes led to behavioral changes.

Fan and colleagues (2018) looked into behavioral changes in object discrimination following drawing. They had participants do a pre-training discrimination test, draw objects, and then do a post-training discrimination test. The discrimination test involved seeing a morph of two objects and assigning it to the object it more closely resembles. They found that the slope parameter changed significantly more for objects that were drawn compared to objects that were not drawn. This suggests that drawing leads to enhanced discrimination, which may reflect more precise memory for the objects. Importantly, none of the studies thus far have compared drawing and tracing.

Gonzalez and colleagues (2011) addressed this idea by comparing memory precision following drawing or tracing abstract lines. On an immediate test, objects that were traced showed marginally better shape accuracy than objects that were drawn. On the delayed test one week later, there was no significant difference between objects that were traced and objects that were drawn in terms of shape accuracy. Several methodological issues may account for these non-significant results. On the delayed test, participants performed equivalently to how they did on the very first test block (which followed only four learning trials). This suggests that the participants may have never sufficiently learned the items. In addition, there were only 8 participants in each condition. This lack of power could also have contributed to the null results.

The study proposed here is similar to Gonzalez' study in that we compare drawing and tracing in terms of shape precision. However, the study presented here uses real-world objects instead of abstract lines which likely leads to better learning, uses a more straightforward 3-alternative forced choice (3AFC) test to provide a clearer index of participant performance, and uses a within-participant design and more participants to ensure sufficient power to detect group differences. This study helps elucidate which strategy (describing, drawing, or tracing) is optimal for producing precise visual memories.

3.2 Methods

3.2.1 Participants

We based our sampling plan on recovery simulations. To generate simulated data, we first defined values for each simulated participant by sampling from distributions. Thus, the simulations incorporate across-participant variability in memory performance. We chose levels of variability that seemed plausible in light of previous memory studies. The parameter values for the distribution function can be seen in Table 3.

Table 3

Parameter values for the simulation distribution function for Aim 2.

Parameter	Value (in log odds)
Center Point Mean	1
Center Point Standard Deviation	0.85
Mean of Draw Offset	0.4
Mean of Trace Offset	0.6
Mean of Study Offset	-0.6
Mean of Describe Offset	0.2
Standard Deviation of Offsets	0.4
Log Sigma Mean	log(1.25)
Log Sigma Standard Deviation	0.2

We fit each simulated data set using a hierarchical Bayesian model. The primary goal was to estimate the difference in memory for the studied items, indexed by the log odds correct on the 3AFC task, between all pairs of learning conditions. We regard a log odds difference of 0.2 as the minimum meaningful value; that is, we are not interested in interpreting log odds differences less than 0.2 as important findings. As such, the key outcomes that we evaluated were 95% credible intervals on the difference in log odds between pairs of learning conditions. Our goal is to have credible intervals that are no wider than 0.4, such that the interval will exclude 0 as a hypothesized effect size if the estimated effect size is greater than 0.2 in absolute value (i.e., the limits of the uncertainty interval should be less than 0.2 away from the effect size point estimate in both directions).

We ran many simulations with varying sample sizes and found approximately 80-100 participants allows for credible intervals less than 0.4 in width. That is, a sample size of 80 produced credible intervals at our desired level of precision about 1/3 of the time, and a sample size of 100 achieved the desired level of precision in nearly 100% of the simulated experiments. Given this, we implemented the following stopping rule: run at least 80 participants, then check the average credible interval width for the six comparisons (*Draw-Describe*, *Draw-Trace*, *Draw-Study*, *Describe-Trace*, *Describe-Study, Trace-Study*) and continue until the average is 0.4 or less up to a maximum of 100 participants.

Participants were undergraduate students from the University of Massachusetts at Amherst recruited through the SONA system. They were compensated with course credit. Participants were 18 years or older and had normal or corrected-to-normal vision. Eleven participants were excluded for not completing the study, two participants were excluded because their data failed to save, one participant was excluded for having an average accuracy below 0.43 (our preregistered cutoff), and one participant was excluded for having five or more trials where they did the incorrect task (our preregistered cutoff). Participants were excluded as data collection proceeded and only usable datasets were counted towards the stopping rule.

We first checked the credible intervals at 81 usable data sets. All of the credible intervals were over 0.4, so we scheduled another week of participants. Following this week, we had a total of 111 usable data sets. Of these 111, ages ranged from 18 to 25 years (avg. = 19.1 years). The sample was made up of 31 males, 79 females, and 1 nonbinary individual.

3.2.2 Stimuli

Object images were selected from the Konkle et al. (2010) database. 80 object categories were used and within each object category, three images were selected. For instance, three button images were selected. For each participant, one image from each category was randomly selected to be the target image. The other images from that category served as lures in the 3AFC recognition test. All images were made grayscale so that recognition relied on object shape, not on object color.

3.2.3 Procedure

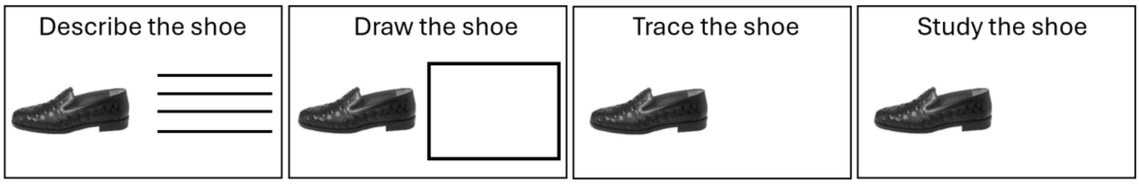
Participants completed this experiment on a tablet (Samsung Galaxy Tab S6 Lite) using a stylus. The experiment began with instructions on how to use the stylus and one practice trial of each task (*Describe*, *Draw*, *Trace*, and *Study*). Participants were informed that they would have a memory test, but they did not know what the test would involve.

Following this instructional phase, the learning phase began. This was a within-participant design with four conditions (*Describe*, *Draw*, *Trace*, *Study*). Twenty target objects were randomly assigned to each condition for each participant. Each trial began with instructions on what task they would be doing on that trial. Then, participants were presented with an image and had 20 seconds to do the associated task. They were instructed to continue doing the task until the time was up. At the top of the screen, there was an instruction that included what they should do on that trial and what the name of the object was. The instructions for the *Describe* and *Draw* trials were identical to Aim 1. For the *Trace* trials, participants were instructed to trace over the object image. For the

Study trials, participants were instructed to visually study the object. See sample trials in Figure 9.

Figure 9

Sample learning phase trials of the Describe, Draw, Trace, and Study tasks for Aim 2.



Following the learning phase, participants had a two-day break, and then they returned to the lab for the 3AFC recognition test. In the test phase, the participant had to select the image that they learned out of three visually similar objects from the same category. All 80 learned images were tested in a random order. This was an unspeeded decision. See an example 3AFC recognition test trial in Figure 10.

Figure 10

Example trial of the 3AFC recognition test for Aim 2.

3.2.4 Open Practices Statement

This study was preregistered on OSF (https://osf.io/vpgzm). In this article, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data, analysis code, and stimuli are available on OSF (https://osf.io/grbmu). Data were analyzed using R (R Core Team, 2021).

3.2.5 Model

We used a Bayesian hierarchical model to estimate the difference in memory for the learned items between all pairs of learning conditions, indexed by the log odds correct on the 3AFC task. This model was fit for each participant and took into consideration overall performance across participants through across-participant distributions for each of the individual-level parameters. The model took in the correct and incorrect counts for each condition for each participant and a prior distribution. The model output estimated log odds of a correct response for each condition for each participant. Instead of having a separate estimate for each of the four conditions, we created a Log Odds Center and adjustments for each of the four conditions to account for some people being better or worse at the task overall. Performance in each of the conditions is then correlated with that participant's Log Odds Center. See Table 4 for the parameters defining the prior across-participant distribution.

Table 4

Parameter values for prior across-participant distribution for Aim 2.

Parameter	Mean	Standard Deviation
Log Odds Center	Normal distribution with mean of 1 and standard deviation of 1.50	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10 and maximum of 1
Log Odds Adjustment Write	Normal distribution with mean of 0 and standard deviation of 0.50	Normal distribution with mean of 0.30 and standard deviation of 0.20, truncated with a minimum of 0.10 and maximum of 1
Log Odds Adjustment Draw	Normal distribution with mean of 0 and standard deviation of 0.50	Normal distribution with mean of 0.30 and standard deviation of 0.20, truncated with a minimum of 0.10 and maximum of 1

Table 4 – continued

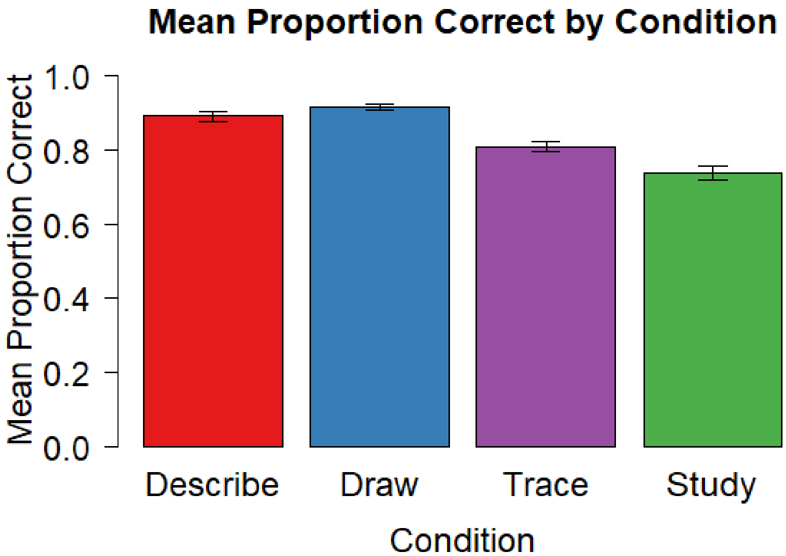
Parameter	Mean	Standard Deviation
Log Odds Adjustment Trace	Normal distribution with mean of 0 and standard deviation of 0.50	Normal distribution with mean of 0.30 and standard deviation of 0.20, truncated with a minimum of 0.10 and maximum of 1
Log Odds Adjustment Study	Normal distribution with mean of 0 and standard deviation of 0.5	Normal distribution with mean of 0.30 and standard deviation of 0.20, truncated with a minimum of 0.10 and maximum of 1

3.3 Results

As an initial check, frequentist analyses were conducted. 3AFC performance was calculated as the proportion of correct responses for each condition. As can be seen in Figure 11, items learned in the *Draw* condition were remembered best, followed by the *Describe* condition, then the *Trace* condition, with *Study* being the worst. The proportion correct for described items was 89.1%, for drawn items was 91.6%, for traced items was 80.8%, and for studied items was 73.8%.

Figure 11

Mean proportion correct by condition for Aim 2.



Note. Error bars represent the standard error.

An ANOVA found significant differences across conditions ($F(3,440) = 37.12$, $p < 0.001$). The results of subsequent two-tailed *t*-tests can be found in Table 5. All of the comparisons are significant. These results suggest that drawing is significantly better than the other three strategies, describing is significantly better than tracing or studying, and tracing is significantly better than studying.

Table 5

The results of the t-tests for Aim 2.

Comparison	*t*-test Results
Draw-Describe	$t(110) = 2.36$, $p = 0.02$, 95% CI [0.004, 0.047]
Draw-Trace	$t(110) = 8.05$, $p < 0.001$, 95% CI [0.08, 0.13]
Draw-Study	$t(110) = 10.31$, $p < 0.001$, 95% CI [0.14, 0.21]
Trace-Study	$t(110) = 4.20$, $p < 0.001$, 95% CI [0.04, 0.10]
Describe-Trace	$t(110) = 6.03$, $p < 0.001$, 95% CI [0.06, 0.11]
Describe-Study	$t(110) = 9.81$, $p < 0.001$, 95% CI [0.12, 0.18]

Subsequently, the Bayesian hierarchical model was used to analyze the log odds correct for each condition. The 95% credible intervals comparing the log odds correct for each condition can be seen in Table 6. It is worth noting that none of the comparisons (including drawing vs. describing) include zero, replicating the frequentist analysis findings.

Table 6

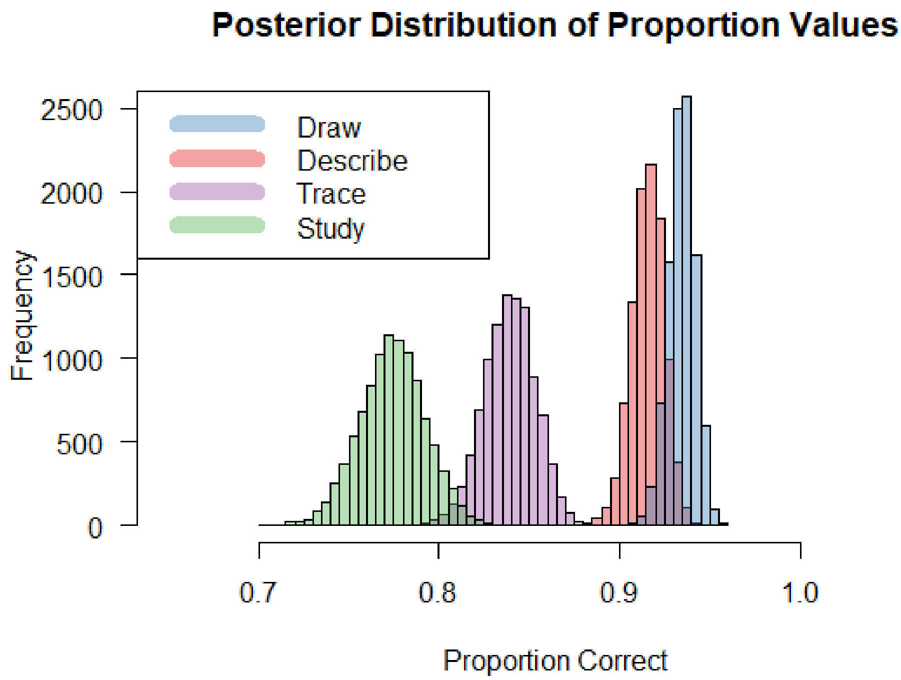
The 95% credible intervals of the log odds correct for each condition comparison for Aim 2.

Comparison	Log Odds 95% Credible Interval
Draw-Describe	0.013-0.527
Draw-Trace	0.773-1.249
Draw-Study	1.19-1.68
Trace-Study	0.209-0.626
Describe-Trace	0.517-0.972
Describe-Study	0.926-1.39

The log odds were then converted back into proportions to aid with the interpretation of the results. The posterior distribution of these across-participant average proportions is visible in Figure 12. In line with the frequentist analyses, we see that drawing is the best strategy, followed by describing, then tracing, and then studying.

Figure 12

Posterior distribution of the proportion values for each condition for Aim 2.



3.4 Discussion

We hypothesized that drawing and tracing would outperform describing in this experiment given the reliance on precise visual memory to accurately complete the 3AFC test. In alignment with this hypothesis, there is evidence from both the frequentist and Bayesian analyses that drawing is better than describing in this experiment, though the difference is small. The Draw-Describe credible interval includes a lot of values below 0.20 which was our preregistered minimum cutoff for a meaningful difference and the frequentist confidence interval on proportion correct indicates that differences less than 0.01 cannot be ruled out. This small benefit of drawing over describing is surprising since a brief description is unlikely to sufficiently differentiate visually similar objects. Additionally, it was surprising to see that tracing led to worse memory compared to both drawing and describing. A potential explanation for this is a lack of active engagement in the task. Both drawing and describing are generative tasks which have been shown to lead to better memory. Tracing adds a motor component but does not generate a new representation. However, it is important to note that tracing is superior to simply studying the image.

Overall, this study suggests that drawing, describing, and tracing are effective strategies for learning precise visual materials. Out of the three strategies, drawing is the most effective in this case with describing very close behind. In conjunction with Aim 1, this provides compelling evidence that the optimal learning strategy depends on the test type[3]. If learning the gist of the item is sufficient (i.e., Aim 1), describing is optimal. In contrast, if learning precise details of the item is important (i.e., Aim 2), drawing is optimal.

---

[3] The 95% credible interval for the interaction between the *Describe-Draw* effect in Aim 1 and in Aim 2 is 0.042 – 0.917. This suggests that there is an interaction, though it may be quite small.

CHAPTER 4

AIM III: FEATURE BINDING WITH YOUNG ADULTS

Aim 3 explores how drawing, tracing, describing, and studying objects compare in terms of the binding of object features in the resulting visual memories. In this experiment, the binding of object identity with object color and object location was explored with young adult participants.

4.1 Background

Feature binding is a critical aspect of visual memory. It is not enough to have a jumbled collection of objects, colors, and locations in memory. For instance, if a colleague asks if they can borrow a red pen from your office, it is not useful to remember that you have one writing instrument on the desk and one writing instrument on the table, one of which is red and one of which is blue. Instead, it is imperative to know which features go with which objects. For instance, you have a red pen on your desk and a blue highlighter on the table.

Interestingly, the question of how drawing, tracing, and describing may impact feature binding has not been previously investigated. Prior experiments have only studied memory for the overall object, without considering how object features are unified in memory. Additionally, of the studies that have investigated drawing and tracing with visual content, all of them used black-and-white stimuli and none included a spatial component. Thus, it is not possible to explore feature binding from the results of these prior studies. Since binding is an integral aspect of our day-to-day visual memory, learning which strategy (drawing, tracing, or describing) is optimal for learning feature binding is important.

4.2 Methods

4.2.1 Participants

We based our sampling plan on recovery simulations. At the time of planning this study, we had data from the older adult version of this study. Though these populations may differ with respect to the benefits of these strategies, it served as a good starting point for our simulations. To generate simulated data, we first defined values for each simulated participant for the color task by sampling from distributions based on the Aim 4 older adult data. The simulations incorporate across-participant variability in memory performance. We chose levels of variability that were close to the values estimated from the Aim 4 older adult data. The parameter values for the distribution function can be seen in Table 7.

Table 7

Parameter values for the simulation distribution function for Aim 3.

Parameter	Value (in log odds)
Center Mu	0.05
Draw Effect Mu	-0.05
Describe Effect Mu	0.35
Trace Effect Mu	-0.45
View Effect Mu	0.15
Center Standard Deviation	0.75
Effect Standard Deviation	0.25

We fit each simulated data set using a hierarchical Bayesian model. The primary goal was to estimate the difference in memory for the learned items, indexed by the log odds correct on the color task, between all pairs of learning conditions. We regard a log odds difference of 0.2 as the minimum meaningful value (i.e., the log odds of responding to the color question correctly in one condition is 0.2 higher or lower than the log odds in another condition). That is, we are not interested in interpreting differences of less than

0.2 as important findings. As such, the key outcomes that we evaluated were 95% credible intervals on the difference in log odds between pairs of learning conditions. Our goal was to have credible intervals no wider than 0.4, such that the interval will exclude 0 as a hypothesized effect size if the estimated effect size is greater than 0.2 in absolute value (based on the logic that half of the interval lies between the mean and 0).

We ran many simulations with varying sample sizes and found that running approximately 80-100 participants consistently allowed for credible intervals less than 0.4 in width. That is, a sample size of 80 produced credible intervals at our desired level of precision about 90% of the time, and a sample size of 100 achieved the desired level of precision in nearly all of the simulated experiments. Given this, we implemented the following stopping rule: run at least 80 participants, then check the average credible interval width for the six comparisons (*Draw-Describe*, *Draw-Trace*, *Draw-Study*, *Describe-Trace*, *Describe-Study*, *Trace-Study*) and continue until the average is 0.40 or less up to a maximum of 100 participants.

Participants were undergraduate students from the University of Massachusetts at Amherst recruited through the SONA system. They were compensated with course credit. Participants were 18 years or older and had normal or corrected-to-normal vision. Ten participants were excluded for the following reasons: five participants had five or more trials where they did the incorrect task (our preregistered cutoff), one participant did not finish in the allotted time, the tablet froze for two participants, the data failed to save for one participant, and the tablet was not calibrated for one participant. Participants were excluded as data collection proceeded and only usable datasets were counted towards the stopping rule.

Data collection moved swiftly, and we surpassed 80 participants and 100 participants in the same week, so data collection stopped after that week regardless of credible interval widths. We ended up with a total of 109 usable data sets. Of these 109, ages ranged from 18 to 24 years (avg. = 19.2 years). The sample was made up of 33 males and 76 females.

4.2.2 Stimuli

Seventy-five object images from the Brady et al. (2013) database were used. The selected objects did not have a typical color associated with them. For instance, a banana or school bus were not used because they are typically yellow. Sixty objects were learned (15 in each condition) while the remaining 15 served as lures in the recognition test. For each participant, the color and location of the learned objects were randomly determined at the beginning of the experiment. The color options were red, orange, yellow, green, blue, or purple.

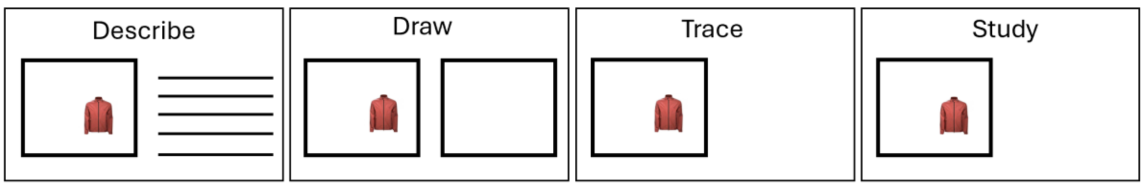
4.2.3 Procedure

Participants completed this experiment on a tablet (Samsung Galaxy Tab S6 Lite) using a stylus. The experiment began with instructions on how to use the stylus and one practice trial of each task (*Draw*, *Trace*, *Describe*, and *Study*) as well as a practice test trial.

Following this instructional phase, the learning phase began. This was a within-participant design with four conditions (*Draw*, *Trace*, *Describe*, *Study*). Fifteen images were randomly assigned to each condition for each participant. On each trial, the participant was presented with the image of a colored object at a certain location in a box. At the top of the screen, there was an instruction that included what they should do on that trial. See sample trials in Figure 13. The images were the same size in all four conditions and participants were given 20 seconds to complete the task in all four conditions. They were instructed to continue the task until the time was up. To ensure that participants were not tracing or drawing in the *Study* condition, they were instructed to sit with their hands in their lap for the duration of these trials.

Figure 13

Sample learning phase trials of the Describe, Draw, Trace, and Study tasks for Aim 3.

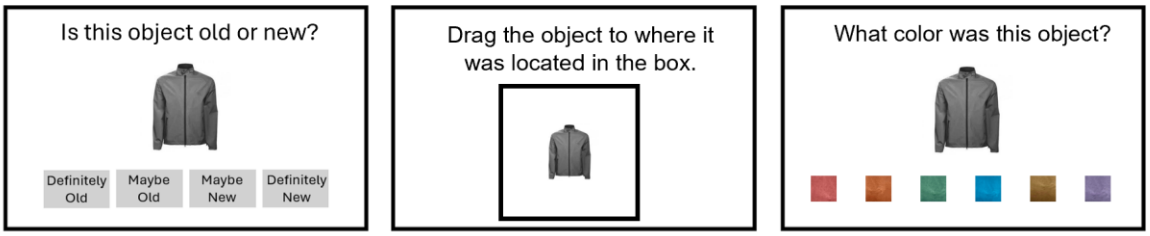


Following the learning phase, a brief filler task occurred. This ensured that participants were not relying on working memory in the recognition test. The filler task involved categorizing 120 scene images as a city, beach, or mountain. This filler task was selected because it involves both a verbal and visual component, ensuring that we were not selectively impairing the *Draw* or *Describe* condition. This took approximately 5 minutes to complete.

Lastly, a memory test was conducted. Participants first had a recognition test in which a grayscale object was presented, and they responded with an old/new judgment. Participants were instructed to focus solely on the object identity in this judgment, ignoring if the object they studied was a different color. In order to incorporate confidence judgments, participants indicated whether the object was “definitely old,” “maybe old,” “maybe new,” or “definitely new.” For all of the objects that were learned (regardless of whether the participant indicated them as old or new), there were two additional questions. First, the participant was asked to drag the object to where it was located in the box when it appeared in the study list. Then, the participant was instructed to click on the color that the object appeared in at study. Sample trials of the three test questions are shown in Figure 14.

Figure 14

Sample memory test trials for Aim 3.



4.2.4 Open Practices Statement

This study was preregistered on OSF (https://osf.io/cn4qf). In this article, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data, analysis code, and stimuli are available on OSF (https://osf.io/zh5m2/). Data were analyzed using R (R Core Team, 2021).

4.2.5 Models

For the recognition data, we used a Bayesian unequal variance signal detection model to estimate d_a and, critically, the difference in d_a between pairs of learning conditions. This model was fit to data from each participant. The model estimated signal detection parameters for each participant as well as parameters that described the group distributions across participants. The parameters for an individual participant defined the relative mean and standard deviation of memory evidence distributions for target and lures, which in turn were mapped to d_a values, and the cutoffs for each of the four possible recognition test responses ("Definitely Old", "Maybe Old", "Maybe New", and "Definitely New"). These parameters defined the predictions for each participant by determining what proportion of the memory distributions fell within each of those response bins, thereby taking into consideration how conservative the participant was (the position of the cutoffs). Instead of having a separate estimate for each of the four conditions, we created a Mu Center and adjustments for each of the four conditions to account for some people being better or worse at the task overall. Performance in each of the conditions is then correlated with that participant's Mu Center. See Table 8 for the parameters defining the prior distribution.

Table 8

Parameter values for the prior distribution for the recognition model in Aim 3.

Parameter & Description	Mean	Standard Deviation
Mu Center: The center of the memory distribution	Normal distribution with mean of 1 and standard deviation of 2, truncated with a minimum of 0	Normal distribution with a mean of 0.50 and a standard deviation of 0.20, truncated with a minimum of 0.10
Mu Adjust Describe: How much to adjust memory distribution for described items	Normal distribution with mean of 0 and standard deviation of 1	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10
Mu Adjust Draw: How much to adjust memory distribution for drawn items	Normal distribution with mean of 0 and standard deviation of 1	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10
Mu Adjust Trace: How much to adjust memory distribution for traced items	Normal distribution with mean of 0 and standard deviation of 1	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10
Mu Adjust Study: How much to adjust memory distribution for studied items	Normal distribution with mean of 0 and standard deviation of 1	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10
LSig: The standard deviation of the target memory distributions	Normal distribution with mean of log(1.25) and standard deviation of 0.10	Normal distribution with mean of 0.125 and standard deviation of 0.10, truncated with a minimum of 0.05

Table 8 – continued

Parameter & Description	Mean	Standard Deviation
C: An index of how conservative the participant was	Normal distribution with mean of 0 and standard deviation of 2	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10
D1: The adjustment for the upper cutoff	Normal distribution with mean of 0.75 and standard deviation of 2, truncated with a minimum of 0	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10
D2: The adjustment for the lower cutoff	Normal distribution with mean of -0.75 and standard deviation of 2, truncated with a maximum of 0	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10

A secondary Bayesian hierarchical model was used to analyze the binding judgments. This model was run separately on the color judgment data and location judgment data. This model estimated the difference in memory for the learned items between all pairs of learning conditions, indexed by the log odds correct on each task. The model estimated parameters for each participant as well as parameters that described the group distributions across participants. The model took in the correct and incorrect counts for each condition for each participant and a prior distribution. The model output estimated log odds of a correct response for each condition for each participant. Instead of having a separate estimate for each of the four conditions, we created a Log Odds Center and adjustments for each of the four conditions to account for some people being better or worse at the task overall. Performance in each of the conditions is then correlated with that participant's Log Odds Center. See Table 9 for the parameters defining the prior distribution.

Table 9

Parameter values for the prior distribution for the color/location model in Aim 3.

Parameter & Description	Mean	Standard Deviation
Log Odds Center: The center of the memory distribution	Normal distribution with mean of 1 and standard deviation of 1.50	Normal distribution with mean of 0.50 and standard deviation of 0.20, truncated with a minimum of 0.10 and maximum of 1
Log Odds Adjustment Write: How much to adjust memory distribution for described items	Normal distribution with mean of 0 and standard deviation of 0.50	Normal distribution with mean of 0.30 and standard deviation of 0.20, truncated with a minimum of 0.10 and maximum of 1
Log Odds Adjustment Draw: How much to adjust memory distribution for drawn items	Normal distribution with mean of 0 and standard deviation of 0.50	Normal distribution with mean of 0.30 and standard deviation of 0.20, truncated with a minimum of 0.10 and maximum of 1
Log Odds Adjustment Trace: How much to adjust memory distribution for traced items	Normal distribution with mean of 0 and standard deviation of 0.50	Normal distribution with mean of 0.30 and standard deviation of 0.20, truncated with a minimum of 0.10 and maximum of 1
Log Odds Adjustment Study: How much to adjust memory distribution for studied items	Normal distribution with mean of 0 and standard deviation of 0.5	Normal distribution with mean of 0.30 and standard deviation of 0.20, truncated with a minimum of 0.10 and maximum of 1

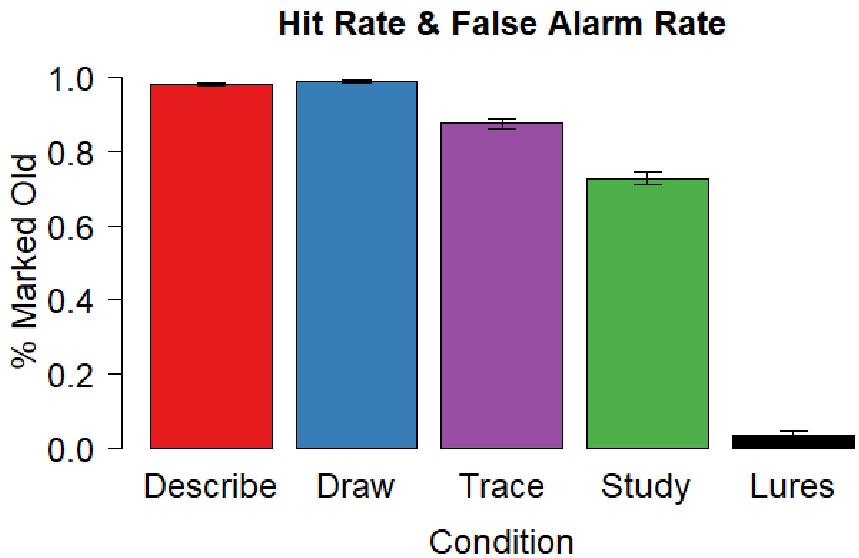
4.3 Results

4.3.1 Recognition

Overall recognition performance was quite high. The hit rate for each condition and the false alarm rate can be seen in Figure 15. The hit rate for the described items was 98.1%, for the drawn items was 98.9%, for the traced items was 87.6%, and for the studied items was 72.7%. The false alarm rate was 3.49%.

Figure 15

The hit rates and false alarm rate for Aim 3.

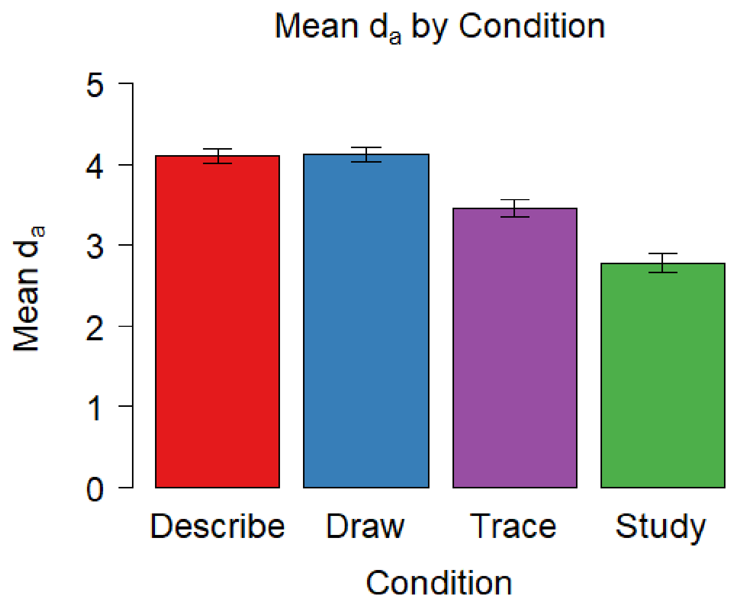


Note. Error bars represent the standard error.

As an initial check, frequentist analyses were conducted. Median d_a values were calculated for each condition for each participant from the Bayesian model posterior distribution. Importantly, this was a simplified version of the Bayesian model that did not include across-participant distributions. This is because in the full Bayesian model, the values for each participant are correlated with each other, thus violating independence. The average d_a values can be seen in Figure 16. Items learned in the *Draw* condition are remembered best followed very closely by items in the *Describe* condition. Tracing is worse than drawing and describing and studying is the worst strategy.

Figure 16

Median d_a value by condition for recognition in Aim 3.



Note. Error bars represent the standard error.

The ANOVA of d_a values across conditions was significant ($F(3,432) = 39.49$, $p < 0.001$). Results of the subsequent two-tailed paired samples *t*-tests are shown in Table 10. All of the comparisons are significant except for *Draw* vs. *Describe*. This suggests that drawing and describing are significantly better than tracing and studying and that tracing is significantly better than studying.

Table 10

The t-test results for recognition for Aim 3.

Comparison	*t*-test Results
Draw-Describe	$t(108) = 0.57$, $p = 0.57$, 95% CI [-0.06, 0.11]
Draw-Trace	$t(108) = 10.71$, $p < 0.001$, 95% CI [0.55, 0.79]
Draw-Study	$t(108) = 20.76$, $p < 0.001$, 95% CI [1.22, 1.48]
Trace-Study	$t(108) = 8.23$, $p < 0.001$, 95% CI [0.52, 0.85]

Table 10 – continued

Comparison	*t*-test Results
Describe-Trace	$t(108) = 9.15$, $p < 0.001$, 95% CI [0.50, 0.78]
Describe-Study	$t(108) = 18.25$, $p < 0.001$, 95% CI [1.18, 1.47]

Subsequently, the full Bayesian hierarchical model was used to analyze the d_a for each condition. A posterior distribution of the across-participant average d_a values by condition is shown in Figure 17. As seen in the histograms, there is considerable overlap between the memory distributions for *Draw* and *Describe*. Both *Draw* and *Describe* produce much stronger memories than *Trace* or *Study*, with *Trace* being better than *Study*. The 95% credible intervals comparing the d_a for each condition can be seen in Table 11. In line with the frequentist analyses, the only credible interval that includes zero is the *Draw-Describe* comparison. As is evident from both the posterior distribution and credible intervals, performance for the drawn and described items was nearly identical. In addition, performance for drawn and described items was well superior to performance for the traced and studied items. Performance for traced items was superior to performance for studied items.

Figure 17

Posterior distribution of recognition d_a values by condition for Aim 3.

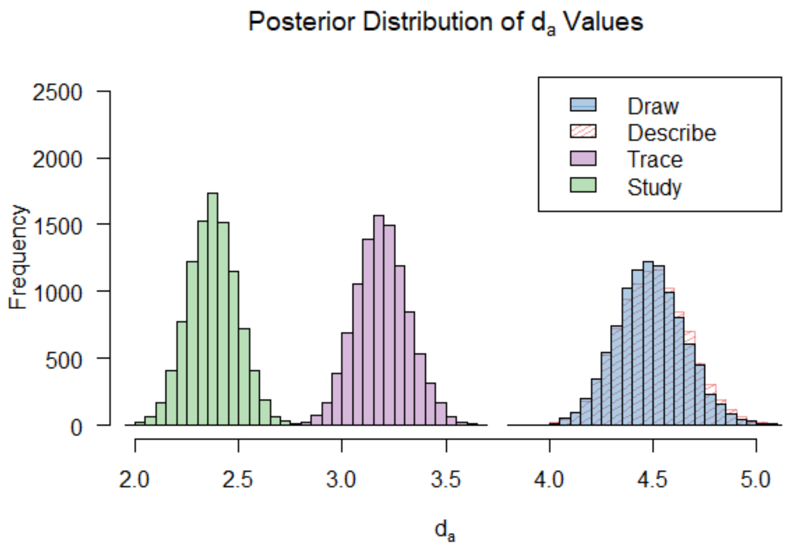


Table 11

The 95% credible intervals of the d_a for each condition comparison for Aim 3.

Comparison	95% Credible Interval
Draw-Describe	-0.34 to 0.31
Draw-Trace	1.03 to 1.58
Draw-Study	1.85 to 2.38
Trace-Study	0.62 to 1.01
Describe-Trace	1.02 to 1.61
Describe-Study	1.85 to 2.42

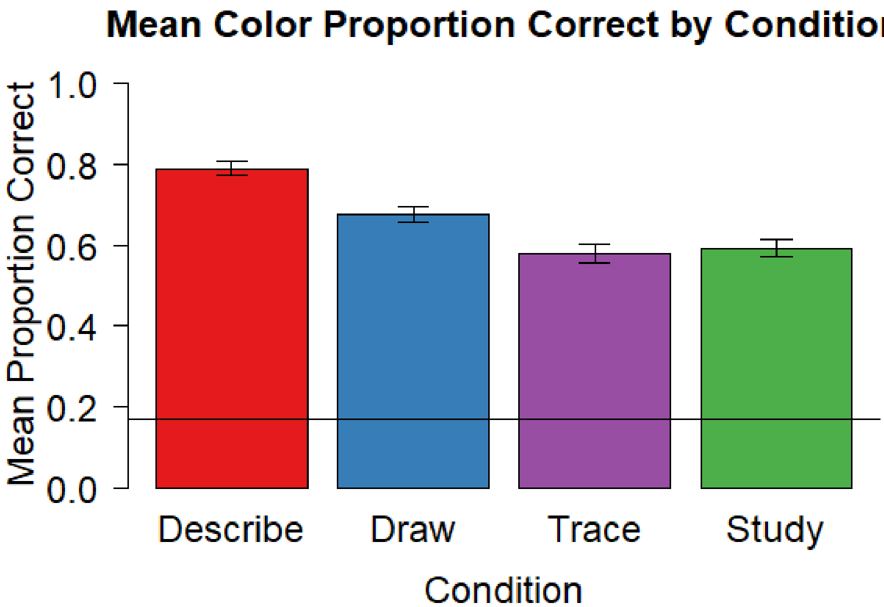
4.3.2 Color Judgment

As an initial check of the differences between conditions for the color judgment, frequentist analyses were conducted. The average proportion correct on the color judgment was calculated for each condition. As seen in Figure 18, color memory was best for items learned in the *Describe* condition followed by *Draw*, then *Study*, and then

Trace. The mean color proportion correct for described items was 78.9%, for drawn items was 67.5%, for traced items was 58.0%, and for studied items was 59.2%. Chance-level performance on this task is 16.7%.

Figure 18

The mean proportion correct by condition for the color judgment in Aim 3.



Note. Error bars represent the standard error. The black line indicates chance-level performance.

The ANOVA of proportion values across conditions was significant ($F(3,432) = 22.95$, $p < 0.001$). Results of the subsequent two-tailed paired samples *t*-tests are shown in Table 12. All of the comparisons are significant except for *Trace* vs. *Study*. This suggests that describing is significantly better than drawing, tracing, and studying and that drawing is significantly better than tracing and studying.

Table 12

The t-test results for the color judgment for Aim 3.

Comparison	t-test Results
Draw-Describe	$t(108) = -6.66$, $p < 0.001$, 95% CI [-0.15, -0.08]
Draw-Trace	$t(108) = 5.13$, $p < 0.001$, 95% CI [0.06, 0.13]
Draw-Study	$t(108) = 4.33$, $p < 0.001$, 95% CI [0.04, 0.12]
Trace-Study	$t(108) = -0.60$, $p = 0.55$, 95% CI [-0.05, 0.03]
Describe-Trace	$t(108) = 11.36$, $p < 0.001$, 95% CI [0.17, 0.25]
Describe-Study	$t(108) = 10.26$, $p < 0.001$, 95% CI [0.16, 0.24]

Subsequently, the Bayesian hierarchical model was used to analyze the log odds correct for each condition. The 95% credible intervals comparing the log odds correct for each condition can be seen in Table 13. In line with the frequentist results, the *Trace-Study* comparison is the only credible interval that includes zero.

Table 13

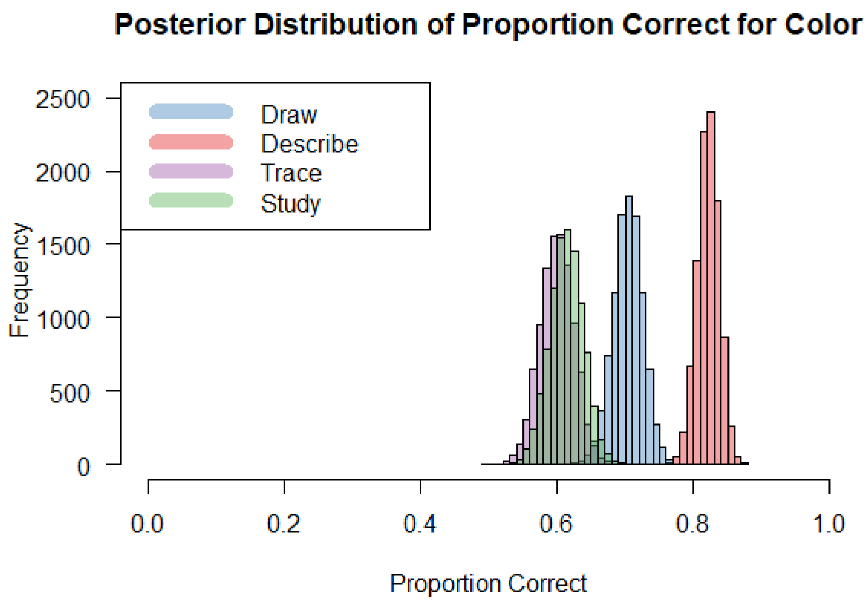
The 95% credible intervals of the log odds for the color judgment for each condition comparison for Aim 3.

Comparison	Log Odds 95% Credible Interval
Draw-Describe	-0.86 to -0.47
Draw-Trace	0.28 to 0.64
Draw-Study	0.21 to 0.59
Trace-Study	-0.25 to 0.13
Describe-Trace	0.94 to 1.31
Describe-Study	0.87 to 1.26

The log odds were then converted back into proportions to aid with the interpretation of the results. The posterior distribution of these across-participant average proportions is visible in Figure 19. In line with the frequentist results, describing is the best strategy followed by drawing, and then studying with tracing very close behind.

Figure 19

Posterior distribution of the proportion values for the color judgment for each condition in Aim 3.

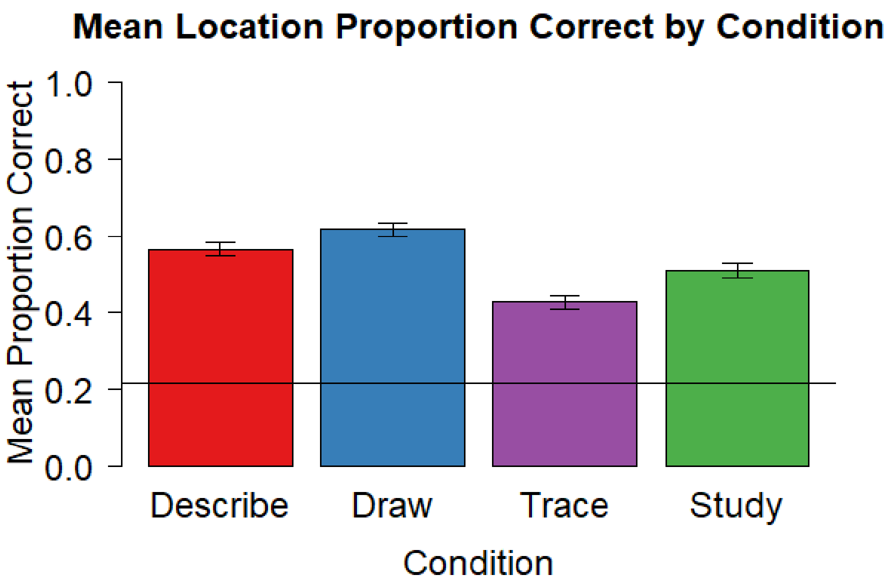


4.3.3 Location Judgment

As an initial check of the differences between conditions for the location judgment, frequentist analyses were conducted. A location placement was considered correct if the x location was within 87 pixels (25% of the range of the possible x locations) of the original x location and the y location was within 87 pixels (25% of the range of the possible y locations) of the original y location. The average proportion correct on the location judgment was calculated for each condition. These averages can be seen in Figure 20. Drawing led to the best location memory, followed by describing, then studying, and then tracing. The mean location proportion correct for described items was 56.5%, for drawn items was 61.7%, for traced items was 42.7%, and for studied items was 50.8%. Chance level on this task is 21.4%.

Figure 20

The mean proportion correct by condition for the location judgment in Aim 3.



Note. Error bars represent the standard error. The black line indicates chance-level performance.

The ANOVA of proportion values across conditions was significant ($F(3,432) = 19.78$, $p < 0.001$). Results of the subsequent two-tailed paired samples *t*-tests are shown in Table 14. All of the comparisons are significant. This suggests that drawing is significantly better than describing, tracing, and studying; describing is significantly better than studying and tracing; and studying is significantly better than tracing.

Table 14

The t-test results for the location judgment for Aim 3.

Comparison	*t*-test Results
Draw-Describe	$t(108) = 3.00$, $p = 0.003$, 95% CI [0.02, 0.09]
Draw-Trace	$t(108) = 9.47$, $p < 0.001$, 95% CI [0.15, 0.23]
Draw-Study	$t(108) = 5.26$, $p < 0.001$, 95% CI [0.07, 0.15]

Table 14 – continued

Comparison	*t*-test Results
Trace-Study	$t(108) = -3.82$, $p < 0.001$, 95% CI [-0.12, -0.04]
Describe-Trace	$t(108) = 7.25$, $p < 0.001$, 95% CI [0.10, 0.18]
Describe-Study	$t(108) = 2.78$, $p = 0.006$, 95% CI [0.02, 0.10]

Location judgment accuracy was slightly better for items learned in the middle (55.5%) than for items learned at the edges (51.6%). For this analysis, the middle was considered within 87 pixels from the center in both the x and y directions. All other locations were considered the edges.

Subsequently, the Bayesian hierarchical model was used to analyze the log odds correct for each condition. The 95% credible intervals comparing the log odds correct for each condition can be seen in Table 15. In line with the frequentist analyses, none of the intervals include zero suggesting a meaningful difference between each of the conditions.

Table 15

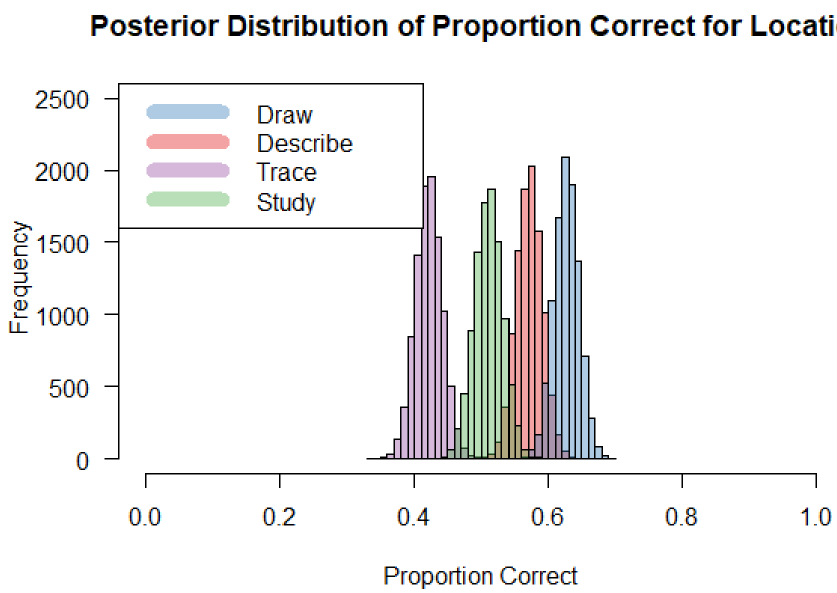
The 95% credible intervals of the log odds for the location judgment for each condition comparison for Aim 3.

Comparison	Log Odds 95% Credible Interval
Draw-Describe	0.07 to 0.40
Draw-Trace	0.66 to 1.01
Draw-Study	0.30 to 0.65
Trace-Study	-0.54 to -0.17
Describe-Trace	0.43 to 0.77
Describe-Study	0.07 to 0.42

The log odds were then converted back into proportions to aid with the interpretation of the results. The posterior distribution of these across-participant average proportions is visible in Figure 21. In line with the frequentist results, drawing is the best strategy followed by describing, then studying, and then tracing.

Figure 21

Posterior distribution of the proportion values for the location judgment for each condition in Aim 3.



4.3.4 Color & Location Dependency

There is evidence suggesting that an object's featural information is often bound to its location (Nissen, 2016). So, if participants get the location of an object correct, they are more likely to get the color of the object correct as well. To test if that holds for our data, we analyzed the proportion of color judgments that were correct given that the location judgment was correct or incorrect. Out of the trials where the participant got the object's location correct, 72.3% (2488/3441) of the time they also got the object's color correct. Out of the trials where the participant got the object's location incorrect, 58.6%

(1798/3068) of the time they got the object's color correct. This suggests that participants were more likely to get the object's color correct if they got the object's location correct.

We were also curious to see if color and location judgments would be more accurate for items that were correctly recognized (hit) than items that were missed. For hit items, the color judgment accuracy was 70.2% (4079/5806) and the location judgment accuracy was 55.8% (3239/5806). For missed items, the color judgment accuracy was 29.7% (207/696) and the location judgment accuracy was 28.9% (201/696). So, location and color accuracy were considerably better for correctly remembered items (hit) than missed items.

4.3.5 Analysis of Descriptions

To gain a better understanding of why performance was so good for the items in the *Describe* condition, the participants' descriptions were analyzed in terms of whether they noted the color and location of the object. Location was mentioned in the participant's description on 23.7% (385/1627) of *Describe* trials. Color was mentioned in the participant's description on 96.9% (1576/1627) of *Describe* trials. Despite being instructed that color and location are the crucial factors, only 9 participants wrote about both the color and location on every *Describe* trial. Overall, these analyses show that participants were much more likely to write about the object's color than its location. In addition to this, we looked into whether the participants' description quality influenced their performance on the color and location test. For the trials in which the participant got the location correct, 26.6% (244/919) had written a description of the location of that object. For the trials in which the participant got the location incorrect, 19.9% (141/708) had written a description of the location of that object. For the trials in which the participant got the color correct, 97.0% (1244/1283) had written a description of the color of that object. For the trials in which the participant got the color incorrect, 96.5% (332/344) had written a description of the color of that object. These results show that writing about the color and/or location was associated with a modest increase in accuracy on the test judgments for that object.

4.3.6 Analysis of Drawings

We were interested in seeing if participants drawing the object in the incorrect location during the learning phase would negatively skew their memory for the object location, leading to more incorrect test location judgments. To determine the center of the participant's drawing, we found the maximum x value and minimum x value of the drawing (within the confines of the box) and then calculated the middle of those. The same strategy was used to find the center for the y-axis. Overall, participants were fairly accurate with the locations of their drawings. The drawing was placed in the correct location in the box on 86.8% (1402/1616) of trials. (The criterion for correctness was the same as the above location analysis – within 87 pixels of the original x and y coordinates.) So, the object was drawn in an incorrect location on 13.2% (214/1616) of trials. Of the trials in which the drawing was in an incorrect location, in the test, the participant placed the object at the incorrect drawn location on 28.0% (60/214) of trials. This suggests that participants' incorrect location judgments were typically not due to being influenced by drawing it in an incorrect location. Participants rarely drew in the wrong location and when they did, their test judgment normally was not influenced by this.

4.4 Discussion

In terms of object recognition, these results replicate those of Aim 1 and Aim 2 showing that drawing and describing are quite similar in the benefits that they provide in comparison to studying and tracing.

For the color judgments, describing was the best strategy followed by drawing. Tracing and studying led to similar performance. This suggests that writing out the name of the color aided memory for the object's color. The distinction between drawing and tracing is somewhat unexpected but likely is due to the higher level of engagement needed for the drawing task.

For the location judgments, drawing was the best strategy followed by describing. Interestingly, tracing was worse than studying. The high performance for the drawn items makes sense because the participant was instructed to draw the object in the same location in the box as it was in the original box. Thus, they needed to pay a lot of attention to the object's location in order to determine where to draw. For describing, participants could write a verbal description of where the object was located (e.g., upper left corner). The deficit of tracing compared to studying was unexpected, but upon further consideration does make sense. When tracing, the participant is focusing almost exclusively on the object and its details, not the broader context. When studying, the participant is surveying the entire screen, including taking in the location information.

Overall, this study suggests that in order to properly bind object features in memory, drawing and describing are the most effective strategies. Interestingly, for feature binding, studying is as effective or more effective than tracing.

CHAPTER 5

AIM IV: FEATURE BINDING WITH OLDER ADULTS

The goal of Aim 4 was to see whether the results of Aim 3 hold true for older adults. So, in this experiment, the binding of object identity with object color and object location was explored with older adult participants.

5.1 Background

There is evidence that memory binding is impaired in older adults (e.g., Chalfonte & Johnson, 1996). Based on the work with younger adults in Aim 3, drawing and describing are potential learning strategies to implement to ameliorate binding-related issues. However, older adults differ from younger adults in many ways that could impact how beneficial these strategies are. For instance, some older adults have motoric challenges such as hand tremors that might impact how well these strategies work for them. In addition, prior research has shown that how beneficial drawing is depends on the participant’s age. One study found that older adults show less of a benefit of drawing compared to younger adults (Tran & Fernandes, 2024a). In contrast, a different study found the opposite effect - more of a benefit of drawing for older adults compared to younger adults (Meade et al., 2018). As such, it is important to directly test whether these strategies are beneficial for older adult populations in this context.

5.2 Methods

5.2.1 Participants

Participants were members of the Amherst Massachusetts community who volunteered to participate. Participants were compensated $15 per hour. Participants were 65 years or older, had normal or corrected-to-normal vision, and were not diagnosed with dementia. No participants were excluded based on our preregistered exclusion criteria (doing the wrong task on five or more trials; having a hit rate that is not at least 0.10 greater than their false alarm rate). We had funds for 39 participants. Of these 39, ages ranged from 65 to 87 years (avg. = 73.3 years). The sample was made up of 10 males and 29 females.

5.2.2 Stimuli

The stimuli were identical to those used in Aim 3.

5.2.3 Procedure

At the very end of the experiment, participants completed the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005). This took approximately 10 minutes to complete. Otherwise, the procedure was identical to that of Aim 3.

5.2.4 Open Practices Statement

This study was preregistered on OSF (https://osf.io/at4dx). In this article, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data, analysis code, and stimuli are available on OSF (https://osf.io/q9jhy/). Data were analyzed using R (R Core Team, 2021).

5.2.5 Models

The models used for this experiment are identical to those used for Aim 3.

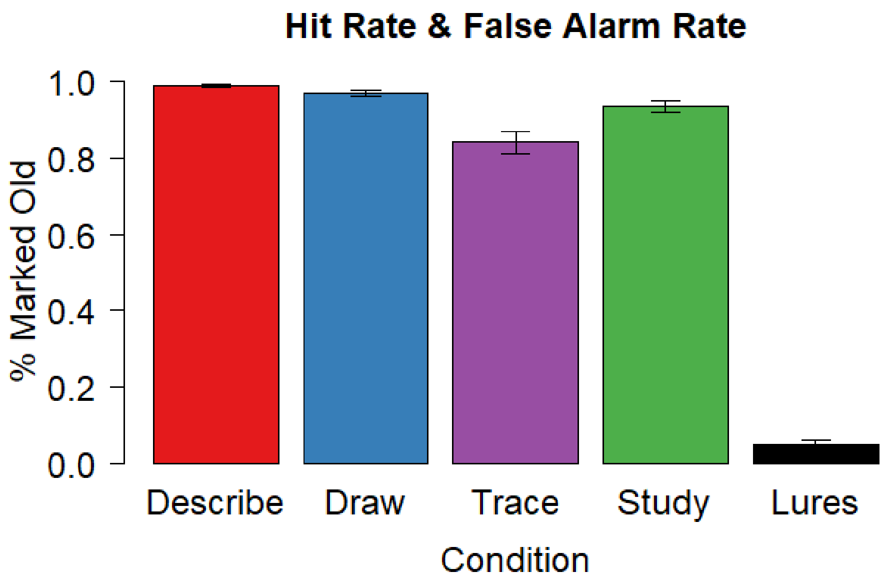
5.3 Results

5.3.1 Recognition

Overall recognition performance was quite high. The hit rate for each condition and the false alarm rate can be seen in Figure 22. The hit rate for the described items was 99.0%, for the drawn items was 97.1%, for the traced items was 84.1%, and for the studied items was 93.5%. The false alarm rate was 4.79%.

Figure 22

The hit rates and false alarm rate for Aim 4.

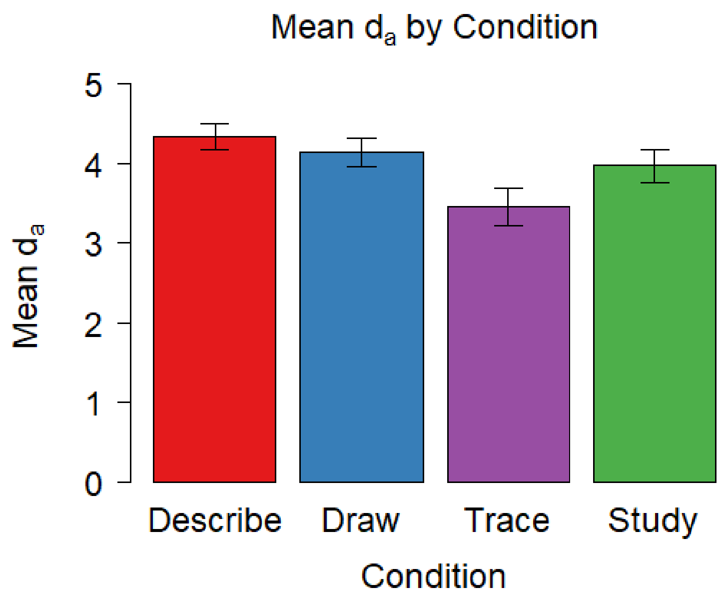


Note. Error bars represent the standard error.

As an initial check, frequentist analyses were conducted. Median d_a values were calculated for each condition for each participant from the Bayesian model posterior distribution. Importantly, this was a simplified version of the Bayesian model that did not include cross-participant adjustments. This is because in the original Bayesian model, the values for each participant are correlated with each other, thus violating independence. The average d_a values can be seen in Figure 23. The items that were described were learned best, followed by drawing, then studying, and then tracing.

Figure 23

Median d_a values by condition for recognition in Aim 4.



Note. Error bars represent the standard error.

The ANOVA of d_a values across conditions was significant ($F(3,152) = 3.68$, p = 0.014). Results of the subsequent two-tailed paired samples *t*-tests are shown in Table 16. All of the comparisons are significant except for *Draw* vs. *Study*. So, describing is significantly better than drawing, tracing, and studying; drawing is significantly better than tracing; and studying is significantly better than tracing.

Table 16

The t-test results for recognition for Aim 4.

Comparison	*t*-test Results
Draw-Describe	$t(38) = -2.82$, $p = 0.008$, 95% CI [-0.34, -0.06]
Draw-Trace	$t(38) = 6.10$, $p < 0.001$, 95% CI [0.46, 0.91]
Draw-Study	$t(38) = 1.63$, $p = 0.112$, 95% CI [-0.04, 0.37]
Trace-Study	$t(38) = -4.36$, $p < 0.001$, 95% CI [-0.76, -0.28]

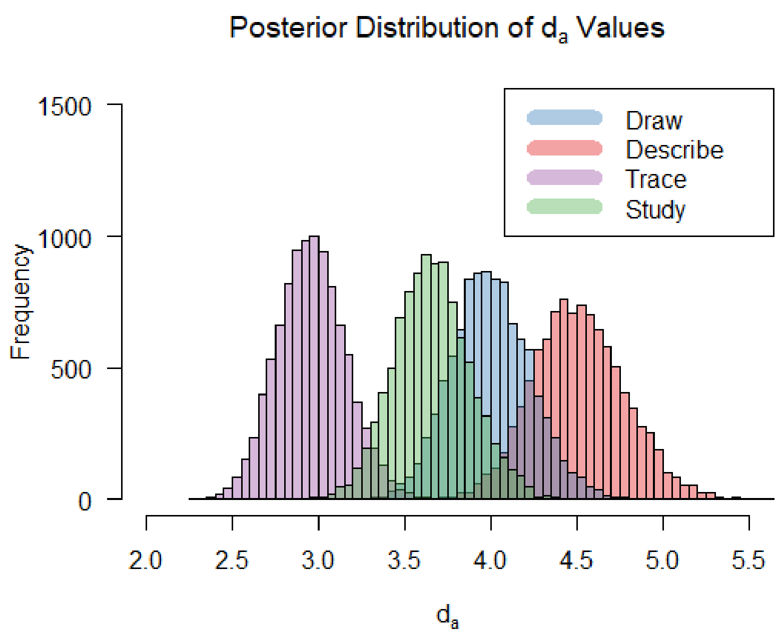
Table 16 – continued

Comparison	*t*-test Results
Describe-Trace	$t(38) = 8.14$, $p < 0.001$, 95% CI [0.66, 1.10]
Describe-Study	$t(38) = 3.75$, $p < 0.001$, 95% CI [0.17, 0.56]

Subsequently, the Bayesian hierarchical model was used to analyze the d_a for each condition. A posterior distribution of across-participant average d_a values by condition is shown in Figure 24. In alignment with the frequentist analysis, describing was the best strategy followed by drawing, then studying, and then tracing.

Figure 24

Posterior distribution of recognition d_a values by condition for Aim 4.



The 95% credible intervals comparing the d_a for each condition can be seen in Table 17. In alignment with the frequentist analysis, the only credible interval that includes zero is the *Draw-Study* comparison. As is evident from both the posterior distribution and credible intervals, performance for the described items was best followed by the drawn items, and then studied items. The traced items were remembered worst.

Table 17

The 95% credible intervals of the d_a for each condition comparison for Aim 4.

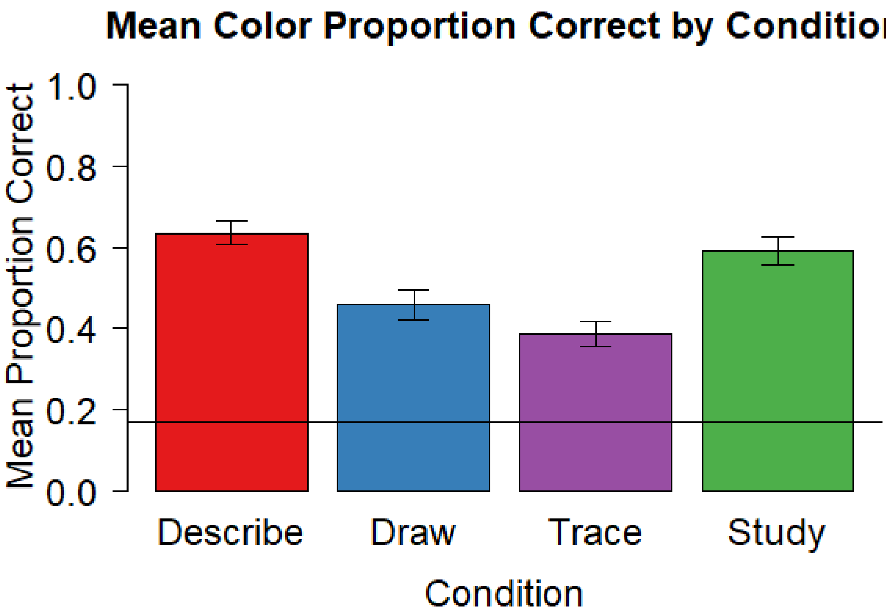
Comparison	95% Credible Interval
Draw-Describe	-0.96 to -0.08
Draw-Trace	0.70 to 1.41
Draw-Study	-0.03 to 0.71
Trace-Study	-1.04 to -0.36
Describe-Trace	1.14 to 2.00
Describe-Study	0.42 to 1.30

5.3.2 Color Judgment

As an initial check of the differences between conditions for the color judgment task, frequentist analyses were conducted. The average proportion correct on the color judgment was calculated for each condition. These averages can be seen in Figure 25. Interestingly, drawing and tracing are both worse than studying on this measure. Additionally, describing is barely superior to studying. The mean color proportion correct for described items was 63.4%, for drawn items was 45.8%, for traced items was 38.7%, and for studied items was 59.0%. Chance-level performance on this task is 16.7%.

Figure 25

The mean proportion correct by condition for the color judgment in Aim 4.



Note. Error bars represent the standard error. The black line indicates chance-level performance.

The ANOVA of proportion values across conditions was significant ($F(3,152) = 12.33$, $p < 0.001$). Results of the subsequent two-tailed paired samples *t*-tests are shown in Table 18. The only comparison that fails to reach significance is *Describe* vs. *Study*. This shows that describing is significantly better than drawing and tracing; studying is significantly better than drawing and tracing; and drawing is significantly better than tracing.

Table 18

The t-test results for the color judgment for Aim 4.

Comparison	*t*-test Results
Draw-Describe	$t(38) = -6.71$, $p < 0.001$, 95% CI [-0.23, -0.12]
Draw-Trace	$t(38) = 2.29$, $p = 0.027$, 95% CI [0.01, 0.13]

Table 18 – continued

Comparison	*t*-test Results
Draw-Study	$t(38) = -3.56$, $p = 0.001$, 95% CI [-0.21, -0.06]
Trace-Study	$t(38) = -6.25$, $p < 0.001$, 95% CI [-0.27, -0.14]
Describe-Trace	$t(38) = 7.90$, $p < 0.001$, 95% CI [0.18, 0.31]
Describe-Study	$t(38) = 1.59$, $p = 0.12$, 95% CI [-0.01, 0.10]

Subsequently, the Bayesian hierarchical model was used to analyze the log odds correct for each condition. The 95% credible intervals comparing the log odds correct for each condition can be seen in Table 19. In alignment with the frequentist analysis, the only interval that includes zero is *Describe* vs. *Study*.

Table 19

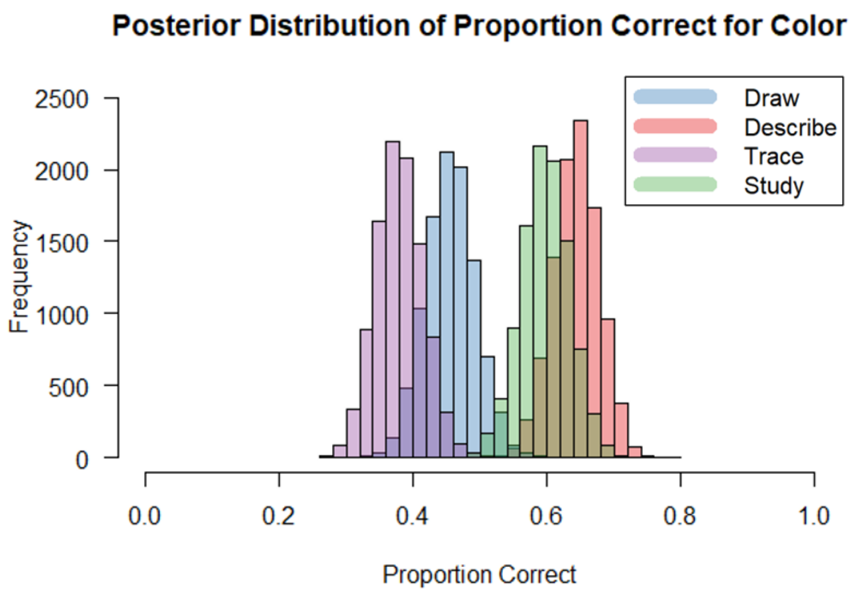
The 95% credible intervals of the log odds for the color judgment for each condition comparison for Aim 4.

Comparison	Log Odds 95% Credible Interval
Draw-Describe	-1.06 to -0.49
Draw-Trace	0.03 to 0.60
Draw-Study	-0.86 to -0.29
Trace-Study	-1.18 to -0.59
Describe-Trace	0.80 to 1.38
Describe-Study	-0.09 to 0.49

The log odds were then converted back into proportions to aid with the interpretation of the results. The posterior distribution of these across-participant average proportions is shown in Figure 26. In alignment with the frequentist analysis, describing is the best strategy followed closely by studying, then drawing, and then tracing.

Figure 26

Posterior distribution of the proportion values for the color judgment for each condition in Aim 4.

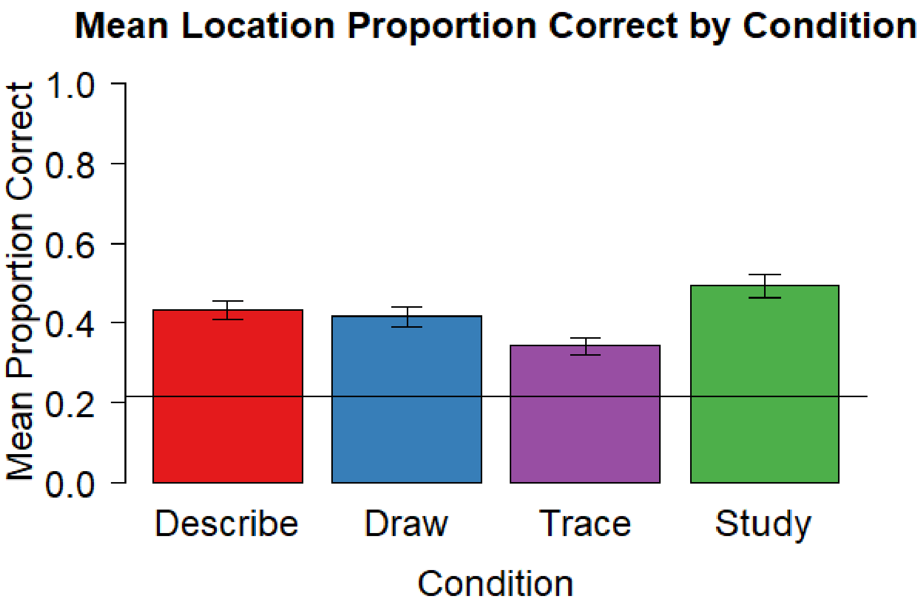


5.3.3 Location Judgment

As an initial check of the differences between conditions for the location judgment, frequentist analyses were conducted. A location placement was considered correct if the x location was within 87 pixels (25% of the range of the possible x locations) of the original x location and the y location was within 87 pixels (25% of the range of the possible y locations) of the original y location. The average proportion correct on the location judgment was calculated for each condition. These averages can be seen in Figure 27. Surprisingly, *Study* was the best condition followed by drawing and describing, and then tracing. The mean location proportion correct for described items was 43.2%, for drawn items was 41.5%, for traced items was 34.1%, and for studied items was 49.2%. Chance level on this task is 21.4%.

Figure 27

The mean proportion correct by condition for the location judgment in Aim 4.



Note. Error bars represent the standard error. The black line represents chance-level performance.

The ANOVA of proportion values across conditions was significant ($F(3,152) = 6.37$, $p < 0.001$). Results of the subsequent two-tailed paired samples *t*-tests are shown in Table 20. All of the comparisons are significant except for *Draw-Describe* and *Describe-Study*. This shows that studying was significantly better than drawing and tracing; describing was significantly better than tracing; and drawing was significantly better than tracing.

Table 20

The t-test results for the location judgment for Aim 4.

Comparison	*t*-test Results
Draw-Describe	$t(38) = -0.73$, $p = 0.47$, 95% CI [-0.06, 0.03]
Draw-Trace	$t(38) = 2.55$, $p = 0.02$, 95% CI [0.02, 0.13]
Draw-Study	$t(38) = -2.52$, $p = 0.016$, 95% CI [-0.14, -0.02]

Table 20 – continued

Comparison	*t*-test Results
Trace-Study	$t(38) = -5.49$, $p < 0.001$, 95% CI [-0.21, -0.10]
Describe-Trace	$t(38) = 3.06$, $p = 0.004$, 95% CI [0.03, 0.15]
Describe-Study	$t(38) = -1.94$, $p = 0.060$, 95% CI [-0.12, 0.003]

Subsequently, the Bayesian hierarchical model was used to analyze the log odds correct for each condition. The 95% credible intervals comparing the log odds correct for each condition can be seen in Table 21. In alignment with the frequentist analysis, the intervals for *Draw* vs. *Describe* and *Describe* vs. *Study* include zero while the rest do not.

Table 21

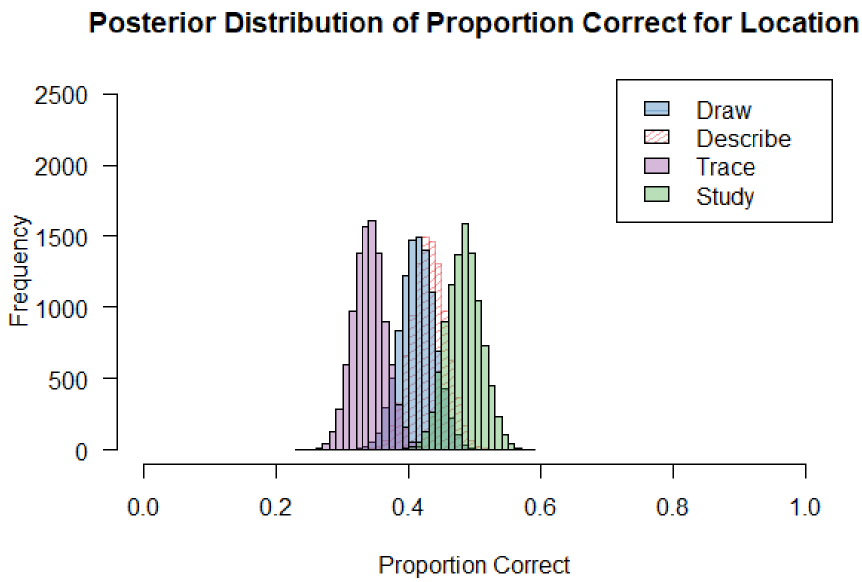
The 95% credible intervals of the log odds for the location judgment for each condition comparison for Aim 4.

Comparison	Log Odds 95% Credible Interval
Draw-Describe	-0.33 to 0.19
Draw-Trace	0.04 to 0.58
Draw-Study	-0.54 to -0.03
Trace-Study	-0.86 to -0.34
Describe-Trace	0.12 to 0.65
Describe-Study	-0.47 to 0.04

The log odds were then converted back into proportions to aid with the interpretation of the results. The posterior distribution of these across-participant average proportions is visible in Figure 28. As the frequentist analysis showed, *Study* was the best condition for subsequent location memory followed by describing and drawing, and then tracing.

Figure 28

Posterior distribution of the proportion values for the location judgment for each condition in Aim 4.



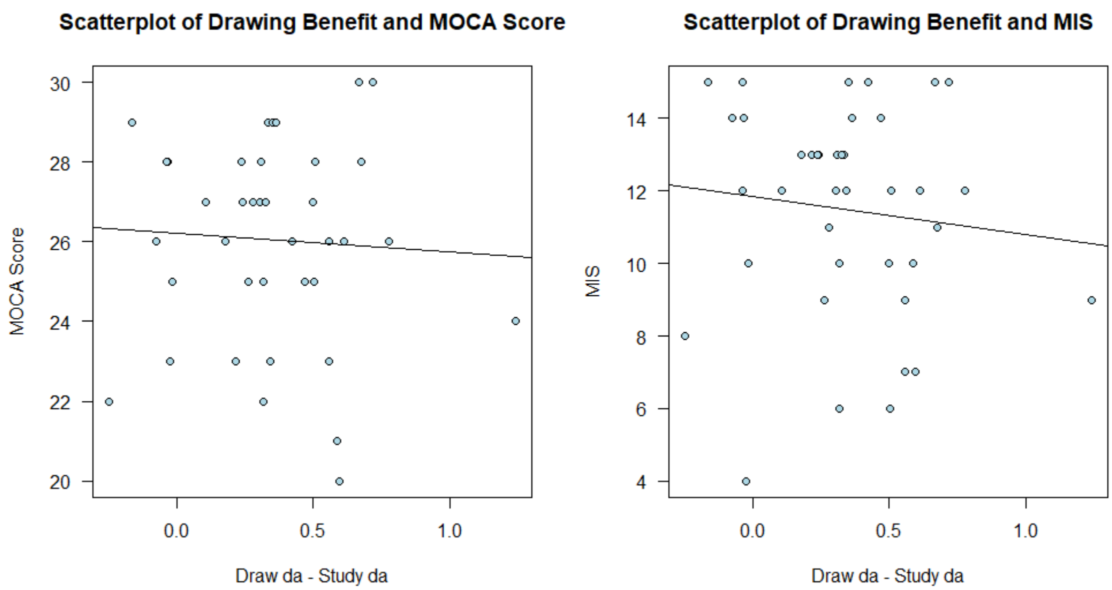
5.3.4 MoCA Analysis

We were interested in seeing whether these strategies (drawing, tracing, and describing) would be more beneficial for people with cognitive or memory impairments. Our indices of cognitive and memory abilities were from the MoCA test. MOCA scores are out of 30 points and are indicative of overall cognitive functioning. A higher score indicates better cognitive functioning. Memory index scores (MIS) are out of 15 points and measure memory abilities in particular. A higher score indicates better memory. To see if these strategies are beneficial relative to typical strategies, we used the *Study* condition as a baseline. The d_a values used in these analyses were from the simplified Bayesian recognition model that did not include cross-participant adjustments.

First, we looked at the benefit of drawing (*Draw* d_a – *Study* d_a) in relation to MOCA and MIS. Figure 29 shows scatterplots of the drawing benefit versus the MOCA score and MIS. The correlation between drawing benefit and MOCA score was weakly negative ($r(37) = -0.102$, $p = 0.54$), and the correlation between drawing benefit and MIS was weakly negative ($r(37) = -0.126$, $p = 0.44$). This suggests that drawing is marginally more beneficial for participants with more cognitive and memory impairments.

Figure 29

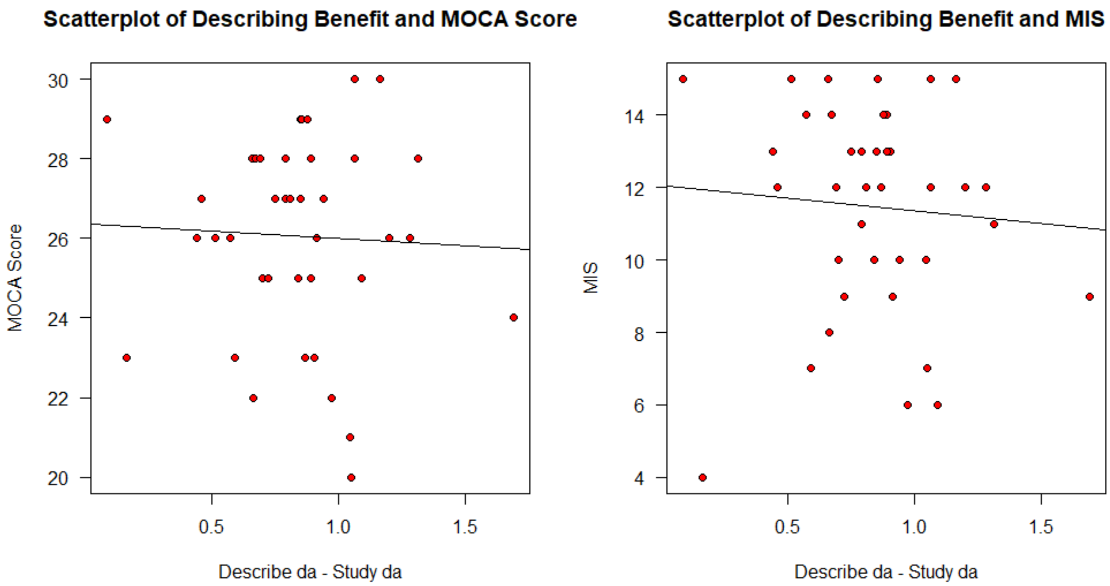
Scatterplots showing the relationship between the drawing benefit (Draw d_a – Study d_a) and MOCA scores (left) and MIS (right).



Next, we looked at the benefit of describing (*Describe* d_a – *Study* d_a) in relation to MOCA and MIS. Figure 30 shows scatterplots of the describing benefit versus the MOCA score and MIS. The correlation between the describing benefit and MOCA score was weakly negative ($r(37) = -0.123$, $p = 0.46$), and the correlation between the describing benefit and MIS was weakly negative ($r(37) = -0.131$, $p = 0.43$). This suggests that describing is marginally more beneficial for participants with more cognitive and memory impairments.

Figure 30

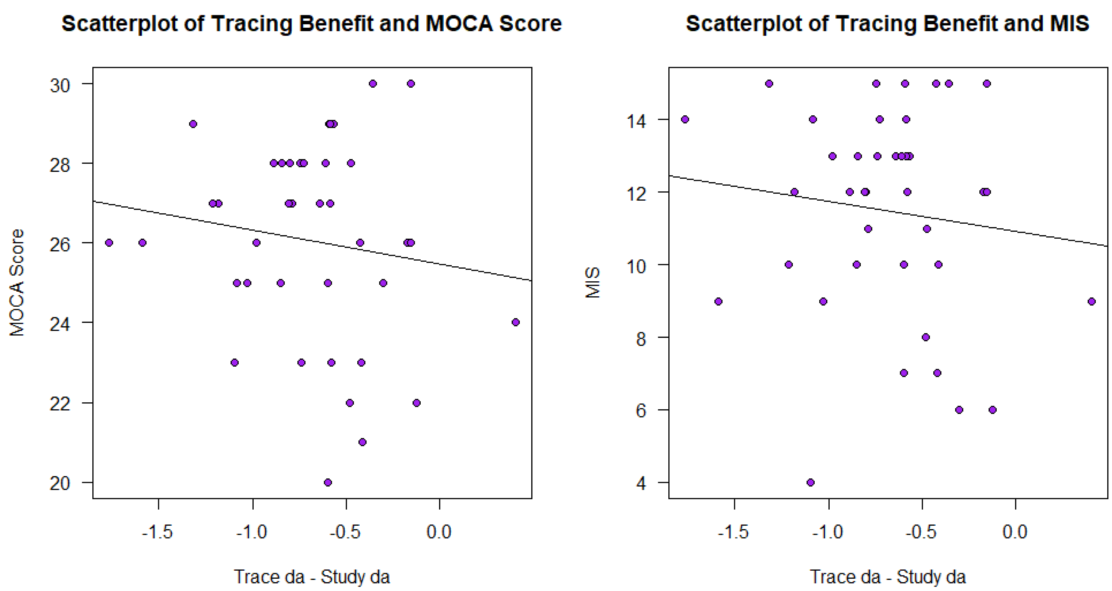
Scatterplots showing the relationship between the describing benefit (Describe d_a – Study d_a) and MOCA scores (left) and MIS (right).



Lastly, we looked at the benefit of tracing (*Trace d_a – Study d_a*) in relation to MOCA and MIS. Figure 31 shows scatterplots of the tracing benefit versus the MOCA score and MIS. The correlation between the tracing benefit and MOCA score was weakly negative ($r(37) = -0.096$, $p = 0.56$), and the correlation between the tracing benefit and MIS was weakly negative ($r(37) = -0.053$, $p = 0.75$). This suggests that tracing is marginally more beneficial for participants with more cognitive and memory impairments.

Figure 31

Scatterplots showing the relationship between the tracing benefit (Trace d_a – Study d_a) and MOCA scores (left) and MIS (right).



5.4 Discussion

The results of Aim 4 were surprisingly different from those of Aim 3 and emphasize the importance of testing out the learning strategy in the target population. In terms of overall recognition, describing and drawing were the best strategies for both younger adults (Aim 3) and older adults (Aim 4). For the younger adults (Aim 3), tracing was better than studying while for the older adults (Aim 4), studying was better than tracing. The worse performance for traced items compared to drawn items is in alignment with Tsutsui and colleagues (2017) who had older adults either draw or trace a diagram. On a subsequent memory test, the group that traced performed worse than the group that drew.

In terms of color judgments, both the younger adults (Aim 3) and older adults (Aim 4) had describing as the best strategy with drawing slightly worse than that and tracing slightly worse than that. The primary difference between the age groups for the color judgments was the studying performance. Color memory for studied items was much better relative to the other conditions for the older adults than the younger adults.

In terms of location judgments, drawing was better than describing for the younger adults (Aim 3) but a bit worse than describing for the older adults (Aim 4). In addition, studying was the 2^{nd} to worst condition for the younger adults (Aim 3) whereas it was the best condition for the older adults (Aim 4).

Overall, across all three test measures, the difference between *Study* and the other conditions varies by age group. Specifically, studying leads to better performance compared to the other conditions for the older adults (Aim 4) compared to the younger adults (Aim 3). This can be viewed as either a studying boost or an impairment to the other tasks. A studying boost may be due to differences in engagement with the studying task. The older adults were generally very motivated, taking the entire time period to study the item while the younger adults may have been less engaged with the task. An impairment with the other tasks (drawing, tracing, and describing) could be due to several factors. First, motoric challenges (e.g., tremors) which some older adults exhibited could be contributing. A tremor would make all of these motoric tasks more frustrating with difficulty writing legibly, drawing a clear picture, or staying on the lines while tracing. This frustration could have led to impaired memory for those items. In addition, Ovalle-Fresa and Martarelli (2024) used the Verbalizer-Visualizer Questionnaire to find that older adults disliked visual information more than younger adults. While we did not measure how much our participants enjoyed visual information and/or drawing, it is possible that this preference for verbal information over visual information was also present in our study and contributed to the relative impairments of drawing and tracing.

In summary, this study suggests that studying and describing are the best strategies for older adults to improve memory binding. When taken in conjunction with the overall recognition performance, describing is the best learning strategy for older adults.

CHAPTER 6

GENERAL DISCUSSION

6.1 Summary of Findings

This work explored whether drawing and tracing are useful learning strategies to effectively encode long-term memories of object images. To begin with, Aim 1 parsed out the conflicting claims between dual-code theory and verbal overshadowing. Dual-code theory would suggest that describing would be better than drawing since it creates a secondary representation (verbal). However, verbal overshadowing suggests that describing can be detrimental to visual memory. To test these conflicting claims, we investigated whether producing a verbal code (through describing) or gaining additional visual experience (through drawing) was more beneficial for learning visual materials. We found that performance was considerably better for images that were described and drawn compared to items that were studied, with a slight advantage of describing over drawing. Thus, this study did not find evidence of verbal overshadowing. Instead, creating dual representations seemed to be optimal.

Aim 2 explored whether drawing, tracing, or describing is more beneficial for memory precision. In this experiment, participants viewed images of objects and either drew a copy of the image, traced the image, described the image, or simply studied the image. They then completed a 3-alternative forced choice test with very visually similar alternatives. We found that memory was much better for items that were drawn and described compared to items that were traced or studied, with a modest benefit of drawing over describing. This shows that, in comparison to Aim 1, when the goal is very detailed visual memory, drawing is slightly superior to describing.

Aim 3 and Aim 4 investigated whether drawing, tracing, or describing is more beneficial for binding object features in memory. In this experiment, participants viewed images of colored objects in a specific location in a box and either drew a copy of the image, traced the image, described the image, or simply studied the image. Participants first had an old/new recognition test. Then, to test their binding, participants were presented with an image of the object and had to select the color and location of it. Aim 3 was conducted with young adults (ages 18-24) and Aim 4 was conducted with older adults (ages 65-87). For both the young and older adults, recognition mirrored that of Aim 2 with better performance for drawn and described objects compared to traced and studied objects. For color judgments, describing was the best strategy followed by drawing for the younger adults. For the older adults, describing was the best strategy followed by studying. For location judgments, drawing was the best strategy followed by describing for the younger adults. For the older adults, studying was the best strategy followed by describing. Overall, for younger adults, describing and drawing lead to robust feature binding. For older adults, describing and studying lead to robust feature binding.

Four potential strategies were explored in this dissertation – drawing, tracing, describing, and our baseline of studying. A summary of the younger adult data for what learning strategy is optimal based on the memory goal can be seen in Table 22. Overall, drawing and describing are great learning strategies. Aside from binding, tracing is also more beneficial than simply studying.

Table 22

A summary of how effective each learning strategy was for younger adults, with #1 being the most effective and #4 being the least effective.

Condition	Gist Memory (Aim 1)	Gist Memory (Aim 3)	Detailed Memory (Aim 2)	Color Binding (Aim 3)	Location Binding (Aim 3)
Drawing	#1 (tie)	#1 (tie)	#1	#2	#1
Tracing	NA	#2	#3	#3 (tie)	#4
Describing	#1 (tie)	#1 (tie)	#2	#1	#2
Studying	#2	#3	#4	#3 (tie)	#3

Note. Non-significant differences between conditions are denoted as ties.

The patterns were different for older adults. See Table 23 for a summary of what learning strategy was optimal based on the memory goal for older adults. Overall, for older adults, describing is a great strategy. If binding is the goal, studying works well too. Tracing was the worst strategy across all test measures.

Table 23

A summary of how effective each learning strategy was for older adults, with #1 being the most effective and #4 being the least effective.

Condition	Gist Memory (Aim 4)	Color Binding (Aim 4)	Location Binding (Aim 4)
Drawing	#2 (tie)	#2	#1 (tie)
Tracing	#3	#3	#2
Describing	#1	#1 (tie)	#1 (tie)
Studying	#2 (tie)	#1 (tie)	#1 (tie)

Note. Non-significant differences between conditions are denoted as ties.

6.2 Limitations & Future Directions

Future work can be done to test these strategies with new stimuli, new methodologies, and new populations. A limitation of this work was that it only used images of common objects as the learning content, but there are many other visual stimuli that are important to remember. As such, it would be interesting to explore the potential benefits of drawing and tracing with other types of stimuli such as faces or maps. An additional limitation of this work was that it was conducted in a lab context, making it unclear how well it would translate to a classroom setting. As such, it would be interesting to see if having students draw or trace as part of their studying would lead to better performance on assessments compared to traditional study strategies. Another limitation of this work was having only a brief delay (up to two days) between learning and test. It would be important to see whether these strategies help to create lasting memories with tests weeks or months later. Lastly, a limitation of this work is that it was conducted primarily with white, highly educated individuals from Massachusetts. It would be important to explore whether these strategies are useful for other groups of people. Overall, there are many directions we can take this work.

6.3 Broader Implications

This work has clear implications for education, especially in fields that involve a lot of visual content such as botany or anatomy. There are many ways that these strategies could be implemented in class contexts. A simple application would be educators assigning drawing, describing, or tracing tasks for their students. For example, I had my introductory psychology students draw pictures of neurons and synapses during our neuroscience unit. A more complex application of this work would be integration with e-books. For instance, if the text is describing something visual, the e-book could have a pop-up that asks the student to trace the image or draw a copy of it.

However, it is important to note that drawing and tracing are not universally beneficial. Many sources share lists of good study strategies, implying that learning tools are universally beneficial. However, this dissertation emphasizes that this is not the case. The optimal study strategy depends on not only the type of content but also the testing goals and the target age group. The Meade et al. (2018) study found that drawing was significantly better than describing for learning verbal content while Aim 1 of this dissertation found that describing was slightly more beneficial than drawing for visual content. This suggests that the optimal strategy depends on the content. In addition, Aim 1 of this dissertation found a slight benefit of describing over drawing when tested on gist memory whereas Aim 2 of this dissertation found a slight benefit of drawing over describing when tested on precise memory. Thus, the optimal strategy depends on what aspects of the memory will be important. Lastly, Aim 3 and Aim 4 show how differently younger adults and older adults respond to these strategies with drawing and tracing being much more effective with younger adults than younger adults. So, educators and students need to take into consideration these factors when deciding on a study strategy.

This work is also informative for older adults who tend to have more memory deficits. It would be useful for older adults to know what strategies are effective for improving their memory. In particular, Aim 4 emphasized that if the individual wants to remember the overall object, describing and drawing are useful learning strategies. If binding object features is important, describing or studying the object can be good tools. Overall, tracing does not seem to be an effective learning strategy for older adults. Understanding what strategies are (and are not) useful can improve learning and memory outcomes for older adults. To increase awareness, perhaps doctors with older adult patients or nursing homes could share strategies such as describing or drawing that would be useful for older adults to be able to improve their memory.

Beyond education and individuals with memory impairments, this work has practical implications for all people. Everyone wants to remember where they left their keys, what their childhood home looked like, or the faces of their loved ones. Knowing about simple learning strategies that can improve subsequent visual memory would benefit everyone.

BIBLIOGRAPHY

Abrams, R. A., Davoli, C. C., Du, F., Knapp III, W. H., & Paull, D. (2008). Altered vision near the hands. *Cognition*, 107(3), 1035-1047.

Agostinho, S., Tindall-Ford, S., Ginns, P., Howard, S. J., Leahy, W., & Paas, F. (2015). Giving learning a helping hand: Finger tracing of temperature graphs on an iPad. *Educational Psychology Review*, 27, 427-443.

Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science*, 333(6046), 1096-1097.

Ainsworth, S. E., & Scheiter, K. (2021). Learning by drawing visual representations: Potential, purposes, and practical implications. *Current Directions in Psychological Science*, 30(1), 61-67.

Alesandrini, K. L. (1981). Pictorial–verbal and analytic–holistic learning strategies in science learning. *Journal of Educational Psychology*, 73(3), 358.

Alsaid, B., & Bertrand, M. (2016). Students' memorization of anatomy, influence of drawing. *Morphologie*, 100(328), 2-6.

Askov, E. N., & Greff, K. N. (1975). Handwriting: Copying versus tracing as the most effective type of practice. *The Journal of Educational Research*, 69(3), 96-98.

Backhouse, M., Fitzpatrick, M., Hutchinson, J., Thandi, C. S., & Keenan, I. D. (2017). Improvements in anatomy knowledge when utilizing a novel cyclical "observe-reflect-draw-edit-repeat" learning process. *Anatomical sciences education*, 10(1), 7-22.

Bäckman, L., & Nilsson, L. G. (1985). Prerequisites for lack of age differences in memory performance. *Experimental Aging Research*, 11(2), 67-73.

Bäckman, L., Nilsson, L. G., & Chalom, D. (1986). New evidence on the nature of the encoding of action events. *Memory & Cognition*, 14, 339-346.

Balemans, M. C., Kooloos, J. G., Donders, A. R. T., & Van der Zee, C. E. (2016). Actual drawing of histological images improves knowledge retention. *Anatomical Sciences Education*, 9(1), 60-70.

Bara, F., Gentaz, E., & Colé, P. (2007). Haptics in learning to read with children from low socio-economic status families. *British Journal of Developmental Psychology*, 25(4), 643-663.

Bara, F., Gentaz, E., Colé, P., & Sprenger-Charolles, L. (2004). The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. *Cognitive development*, 19(3), 433-449.

Baratz, G., Buaron, B., & Mukamel, R. (2023). The influence of active tracing on shape representation in visual pathways. *Journal of Vision*, 23(9), 5201-5201.

Berman, A. (1939). The influence of the kinaesthetic factor in the perception of symbols in partial reading disability. *Journal of Educational Psychology*, 30(3), 187.

Bobek, E., & Tversky, B. (2016). Creating visual explanations improves learning. *Cognitive research: principles and implications*, 1, 1-14.

Brady, T. F., Konkle, T., Alvarez, G. A. and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, USA, 105 (38), 14325-14329.

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological science*, 24(6), 981-990.

Brod, G. (2021). Generative learning: Which strategies for what age?. *Educational Psychology Review*, 33(4), 1295-1318.

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, *24*(4), 403-416.

Cromley, J. G., Du, Y., & Dane, A. P. (2020). Drawing-to-learn: does meta-analysis show differences between technology-based drawing and paper-and-pencil drawing?. *Journal of Science Education and Technology*, 29, 216-229.

Dean, R. S., & Kulhavy, R. W. (1981). Influence of spatial organization in prose learning. *Journal of Educational Psychology*, 73(1), 57.

Edens, K. M., & Potter, E. (2003). Using descriptive drawings as a conceptual change strategy in elementary science. *School Science and Mathematics*, 103(3), 135-144.

Engelkamp, J., & Zimmer, H. D. (1989). Memory for action events: a new field of research. *Psychological research*.

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica*, 96(1-2), 43-60.

Fan, J. E. (2015). Drawing to learn: How producing graphical representations enhances scientific thinking. *Translational Issues in Psychological Science*, 1(2), 170.

Fan, J. E., Bainbridge, W. A., Chamberlain, R., & Wammes, J. D. (2023). Drawing as a versatile cognitive tool. *Nature Reviews Psychology*, 2(9), 556-568.

Fan, J. E., Yamins, D. L., & Turk-Browne, N. B. (2018). Common object representations for visual production and recognition. *Cognitive science*, 42(8), 2670-2698.

Fan, J. E., Wammes, J. D., Gunn, J. B., Yamins, D. L., Norman, K. A., & Turk-Browne, N. B. (2020). Relating visual production and recognition of objects in human visual cortex. *Journal of Neuroscience*, 40(8), 1710-1721.

Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science*, 27(5), 302-308.

Fiorella, L., & Zhang, Q. (2018). Drawing boundary conditions for learning by drawing. *Educational Psychology Review*, 30, 1115-1137.

Ginns, P., Hu, F. T., Byrne, E., & Bobis, J. (2016). Learning by tracing worked examples. *Applied Cognitive Psychology*, 30(2), 160-169.

Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on psychological science*, 8(5), 573-585.

Gobert, J. D., & Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 36(1), 39-53.

Gonzalez, C., Anderson, J., Culmer, P., Burke, M. R., Mon-Williams, M., & Wilkie, R. M. (2011). Is tracing or copying better when learning to reproduce a pattern?. *Experimental Brain Research*, 208, 459-465.

Gowen, E., & Miall, R. C. (2006). Eye–hand interactions in tracing and drawing tasks. *Human movement science*, 25(4-5), 568-585.

Hall, V. C., Bailey, J., & Tillman, C. (1997). Can student-generated illustrations be worth ten thousand words?. *Journal of educational psychology*, 89(4), 677.

Hirsch, E., & Niedermeyer, F. C. (1973). The effects of tracing prompts and discrimination training on kindergarten handwriting performance. *The Journal of Educational Research*, 67(2), 81-86.

Hoffman, H. G. (1997). Role of memory strength in reality monitoring decisions: Evidence from source attribution biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(2), 371.

Hu, F. T., Ginns, P., & Bobis, J. (2014). Does Tracing Worked Examples Enhance Geometry Learning?. *Australian Journal of Educational & Developmental Psychology*, 14, 45-49.

Hu, F. T., Ginns, P., & Bobis, J. (2015). Getting the point: Tracing worked examples enhances learning. *Learning and Instruction*, 35, 85-93.

Hulme, C. (1979). The interaction of visual and motor memory for graphic forms following tracing. *The Quarterly journal of experimental psychology*, 31(2), 249-261.

Hulme, C., Monk, A., & Ives, S. (1987). Some experimental studies of multi-sensory teaching: the effects of manual tracing on children's paired-associate learning. *British Journal of Developmental Psychology*, 5(4), 299-307.

Jalava, S. T., Wammes, J. D., & Cheng, K. (2023). Drawing your way to an A: Long-lasting improvements in classroom quiz performance following drawing. *Psychonomic Bulletin & Review*, 1-7.

Kalenine, S., Pinet, L., & Gentaz, E. (2011). The visual and visuo-haptic exploration of geometrical shapes increases their recognition in preschoolers. *International Journal of Behavioral Development*, 35(1), 18-26.

Kato, C., Isoda, H., Takehara, Y., Matsuo, K., Moriya, T., & Nakai, T. (1999). Involvement of motor cortices in retrieval of kanji studied by functional MRI. *Neuroreport*, 10(6), 1335-1339.

Kim, B. S., Tran, S. H., & Fernandes, M. A. (2024). Comparing the effectiveness of encoding techniques on memory for vocabulary in a second language. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of experimental Psychology: general*, 139(3), 558.

Kulhavy, R. W., Lee, J. B., & Caterino, L. C. (1985). Conjoint retention of maps and related discourse. *Contemporary educational psychology*, 10(1), 28-37.

Lansing, K. M. (1981). The effect of drawing on the development of mental representations. *Studies in Art Education*, 22(3), 15-23.

Leopold, C., & Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learning and Instruction*, 22(1), 16-26.

Leopold, C., Sumfleth, E., & Leutner, D. (2013). Learning with summaries: Effects of representation mode and type of learning activity on comprehension and transfer. *Learning and Instruction*, 27, 40-49.

Leutner, D., Leopold, C., & Sumfleth, E. (2009). Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. *Computers in Human Behavior*, 25(2), 284-289.

Levin, J. R., Ghatala, E. S., DeRose, T. M., & Makoid, L. A. (1977). Image tracing: An analysis of its effectiveness in children's pictorial discrimination learning. *Journal of Experimental Child Psychology*, 23(1), 78-83.

Levin, J. R., Ghatala, E. S., DeRose, T. M., Wilder, L., & Norton, R. W. (1975). A Further Comparison of Imagery and Vocalization Strategies in Children's Discrimination Learning. *Journal of Educational Psychology*, 67(1), 141.

Longcamp, M., Anton, J. L., Roth, M., & Velay, J. L. (2003). Visual presentation of single letters activates a premotor area involved in writing. *Neuroimage*, 19(4), 1492-1500.

Longcamp, M., Boucard, C., Gilhodes, J. C., Anton, J. L., Roth, M., Nazarian, B., & Velay, J. L. (2008). Learning through hand-or typewriting influences visual recognition of new graphic shapes: Behavioral and functional imaging evidence. *Journal of cognitive neuroscience*, 20(5), 802-815.

Longcamp, M., Boucard, C., Gilhodes, J. C., & Velay, J. L. (2006). Remembering the orientation of newly learned characters depends on the associated writing knowledge: A comparison between handwriting and typing. *Human movement science*, 25(4-5), 646-656.

Longcamp, M., Zerbato-Poudou, M. T., & Velay, J. L. (2005). The influence of writing practice on letter recognition in preschool children: A comparison between handwriting and typing. *Acta psychologica*, 119(1), 67-79.

Macken, L., & Ginns, P. (2014). Pointing and tracing gestures may enhance anatomy and physiology learning. *Medical teacher*, 36(7), 596-601.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Mason, L., Lowe, R., & Tornatora, M. C. (2013). Self-generated drawings for supporting comprehension of a complex animation. *Contemporary Educational Psychology*, 38(3), 211-224.

Meade, M. E., Ahmad, M., & Fernandes, M. A. (2020a). Drawing pictures at encoding enhances memory in healthy older adults and in individuals with probable dementia. *Aging, Neuropsychology, and Cognition*, 27(6), 880-901.

Meade, M. E., Klein, M. D., & Fernandes, M. A. (2020b). The benefit (and cost) of drawing as an encoding strategy. *Quarterly Journal of Experimental Psychology*, 73(2), 199-210.

Meade, M. E., Wammes, J. D., & Fernandes, M. A. (2018). Drawing as an encoding tool: Memorial benefits in younger and older adults. *Experimental aging research*, 44(5), 369-396.

Meade, M. E., Wammes, J. D., & Fernandes, M. A. (2019). Comparing the influence of doodling, drawing, and writing at encoding on memory. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 73(1), 28.

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 15(6), 603-616.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, 16(5), 519-533.

Naka, M. (1998). Repeated writing facilitates children's memory for pseudocharacters and foreign letters. *Memory & cognition*, 26, 804-809.

Naka, M., & Naoi, H. (1995). The effect of repeated writing on memory. *Memory & cognition*, 23, 201-212.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695-699.

Nissen, M. J. (2016). Accessing features and objects: Is location special?. In *Attention and performance XI* (pp. 205-219). Routledge.

Ofman, W., & Shaevitz, M. (1963). The kinesthetic method in remedial reading. *The Journal of Experimental Education*, 31(3), 317-320.

Ovalle-Fresa, R., & Martarelli, C. S. (2024). Drawing as an efficient encoding tool in younger but not always older adults: The case of associative memory. *Memory & Cognition*, 1-26.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 45(3), 255.

Peynircioğlu, Z. F. (1989). The generation effect with pictures and nonsense figures. *Acta Psychologica*, 70(2), 153-160.

Ploetzner, R., & Fillisch, B. (2017). Not the silver bullet: Learner-generated drawings make it difficult to understand broader spatiotemporal structures in complex animations. *Learning and Instruction*, 47, 13-24.

R Core Team (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reed, C. L., Grubb, J. D., & Steele, C. (2006). Hands up: attentional prioritization of space near the hand. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1), 166.

Roberts, R. W., & Coleman, J. C. (1958). An investigation of the role of visual and kinesthetic factors in reading failure. *The Journal of Educational Research*, 51(6), 445-451.

Roberts, B. R., Meade, M. E., & Fernandes, M. A. (2024). Brain regions supporting retrieval of words drawn at encoding: fMRI evidence for multimodal reactivation. *Memory & Cognition*, 1-17.

Roberts, B. R., & Wammes, J. D. (2021). Drawing and memory: Using visual production to alleviate concreteness effects. *Psychonomic Bulletin & Review*, 28, 259-267.

Scheiter, K., Schleinschok, K., & Ainsworth, S. (2017). Why sketching may aid learning from science texts: Contrasting sketching with written explanations. *Topics in cognitive science*, 9(4), 866-882.

Schmeck, A., Mayer, R. E., Opfermann, M., Pfeiffer, V., & Leutner, D. (2014). Drawing pictures during learning from scientific text: Testing the generative drawing effect and the prognostic drawing effect. *Contemporary Educational Psychology*, 39(4), 275-286.

Schmidgall, S. P., Eitel, A., & Scheiter, K. (2019). Why do learners who draw perform well? Investigating the role of visualization, generation and externalization in learner-generated drawing. *Learning and Instruction*, 60, 138-153.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive psychology*, 22(1), 36-71.

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology*, 102(4), 872.

Schwamborn, A., Thillmann, H., Opfermann, M., & Leutner, D. (2011). Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *Computers in human behavior*, 27(1), 89-93.

Snowman, J., & Cunningham, D. J. (1975). A comparison of pictorial and written adjunct aids in learning from text. *Journal of Educational Psychology*, 67(2), 307.

Tang, M., Ginns, P., & Jacobson, M. J. (2019). Tracing enhances recall and transfer of knowledge of the water cycle. *Educational Psychology Review*, 31, 439-455.

Tirre, W. C., Manelis, L., & Leicht, K. L. (1979). The effects of imaginal and verbal strategies on prose comprehension by adults. *Journal of Reading Behavior*, 11(2), 99-106.

Tran, S. H., & Fernandes, M. A. (2024a). Age differences in effectiveness of encoding techniques on memory. *Aging, Neuropsychology, and Cognition*, *31*(3), 479-495.

Tran, S. H., & Fernandes, M. A. (2024b). Which encoding techniques facilitate comprehension?. *Experimental Psychology*.

Tsutsui, M., Notoya, M., Kimura, D., Nakatani, K., Fujita, T., Sunahara, N., & Inoue, K. (2017). Utility of Tracing as a Memory Storage Method. *World Journal of Neuroscience*, 7(02), 216.

Tzeng, O. J., Hung, D. L., Chen, S., Wu, J., & Hsi, M. S. (1986). Processing Chinese logographs by Chinese brain damaged patients. *Advances in Psychology* (Vol. 37, pp. 357-374).

Van Essen, G. V., & Hamaker, C. (1990). Using self-generated drawings to solve arithmetic word problems. *The journal of educational research*, 83(6), 301-312.

Van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of educational psychology*, 93(1), 129.

Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology*, 31(2), 142-166.

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational psychology review*, 17, 285-325.

Wammes, J. D., & Fernandes, M. A. (2017). The residual protective effects of enactment. *Cognition*, 164, 87-101.

Wammes, J. D., Jonker, T. R., & Fernandes, M. A. (2019). Drawing improves memory: The importance of multimodal encoding context. *Cognition*, 191, 103955.

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *Quarterly Journal of Experimental Psychology*, 69(9), 1752-1776.

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2017). Learning terms and definitions: Drawing and the role of elaborative encoding. *Acta Psychologica*, 179, 104-113.

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2018a). Creating a recollection-based memory through drawing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(5), 734.

Wammes, J. D., Roberts, B. R., & Fernandes, M. A. (2018b). Task preparation as a mnemonic: The benefits of drawing (and not drawing). *Psychonomic Bulletin & Review*, 25, 2365-2372.

Zemlock, D., Vinci-Booher, S., & James, K. H. (2018). Visual–motor symbol production facilitates letter recognition in young children. *Reading and Writing*, 31, 1255-1271.

Zhang, Z. H., & Linn, M. C. (2011). Can generating representations enhance learning with dynamic visualizations?. *Journal of research in science teaching*, 48(10), 1177-1198.

Zhang, Z. H., & Linn, M. C. (2013). Learning from chemical visualizations: Comparing generation and selection. *International Journal of Science Education*, 35(13), 2174-2197.