THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2016
Vol. 69, No. 9, 1752–1776, http://dx.doi.org/10.1080/17470218.2015.1094494

Routledge Taylor & Francis Group

The drawing effect: Evidence for reliable and robust memory benefits in free recall

Jeffrey D. Wammes, Melissa E. Meade, and Myra A. Fernandes

Department of Psychology, University of Waterloo, Waterloo, ON, Canada

(Received 23 December 2014; accepted 31 August 2015; first published online 16 February 2016)

In 7 free-recall experiments, the benefit of creating drawings of to-be-remembered information relative to writing was examined as a mnemonic strategy. In Experiments 1 and 2, participants were presented with a list of words and were asked to either draw or write out each. Drawn words were better recalled than written. Experiments 3–5 showed that the memory boost provided by drawing could not be explained by elaborative encoding (deep level of processing, LoP), visual imagery, or picture superiority, respectively. In Experiment 6, we explored potential limitations of the drawing effect, by reducing encoding time and increasing list length. Drawing, relative to writing, still benefited memory despite these constraints. In Experiment 7, the drawing effect was significant even when encoding trial types were compared in pure lists between participants, inconsistent with a distinctiveness account. Together these experiments indicate that drawing enhances memory relative to writing, across settings, instructions, and alternate encoding strategies, both within- and between-participants, and that a deep LoP, visual imagery, or picture superiority, alone or collectively, are not sufficient to explain the observed effect. We propose that drawing improves memory by encouraging a seamless integration of semantic, visual, and motor aspects of a memory trace.

Keywords: Subject-performed tasks; Memory; Drawing; Imagery; Levels of processing.

In the current study, we first sought to determine whether drawing was an efficacious strategy for boosting later retention and memory performance. Previous work has indicated that many subject-performed tasks, such as production (MacLeod, Gopie, Hourihan, Neary, & Ozbuko, 2010), generation (Slamecka & Graf, 1978), and enactment (Guttentag & Hunt, 1988), carried out during encoding can provide a memorial benefit relative to more passive encoding strategies such as silent reading. Though useful, these strategies may not be practical in a typical learning environment such as a classroom or lecture hall due to their disruptive nature. For this reason, there is a need to find practical unobtrusive techniques that people can apply in their everyday lives to remember important information, or that students can apply to enhance retention. One traditional study approach in the aforementioned circumstances would be to write detailed notes based on the professor's chosen lecture topic, or writing down a list of to-be-remembered information.

Correspondence should be addressed to Jeffrey Wammes, Department of Psychology, University of Waterloo, 200 University Ave W, Waterloo, ON, N2L 3G1, Canada. E-mail: jwammes@uwaterloo.ca

The authors would thank Grace Sim, Liat Kofler, and Terri Middleton for their dedication in completing this project, as well as insightful comments and suggestions for our work.

This work was funded by a scholarship from the Natural Science and Engineering Research Council (NSERC) awarded to author J.W.; and an NSERC Discovery grant awarded to author M.F.

Drawing and retention

Given that many studies show memorial benefits for information presented as pictures compared to words (e.g., Paivio, Rogers, & Smythe, 1968), it stands to reason that creating a drawing of to-be-learned information may lead to superior long-term retention relative to a more commonplace note-taking strategy. Pulling on a similar thread to that of previous researchers (Guttentag & Hunt, 1988; MacLeod et al., 2010; Slamecka & Graf, 1978) who explored the anecdotal notion that there was an advantage to learning by doing, we sought to determine whether drawing provided a measurable advantage over passive note-taking. Surprisingly, research corroborating or even testing this intuition is sparse. Some educational studies have touched on the matter, but often using educational paradigms in which drawing is confounded with a number of other additional learning tools, such as the provision of visual aids such as backgrounds and cutouts to assemble (Lesgold, De Good, & Levin, 1977; Schwamborn, Mayer, Thillmann, Leopold, & Leutner, 2010), existing images upon which to model their drawings, or additional time on task for the drawing relative to control conditions (see Van Meter & Garner, 2005, for a review). The latter example makes it difficult to ascertain whether the benefit of drawing was simply a function of the total time spent studying each item (Van Meter, 2001), a factor known to bolster memory (Cooper & Pantle, 1967). The goal of the current work was to first approach this question in its most basic form and then build to a comparison with other common learning techniques (visual imagery, elaborative encoding, and picture superiority). The reasoning behind this was twofold: first, to demonstrate that drawing may be a better alternative to more extensively researched mnemonic strategies, and second, because each could be argued as one potential mechanism through which drawing influences memory. Lastly, we examined potential boundary conditions, by testing whether any benefit of drawing was reduced by dramatically shortening the encoding duration, by increasing the length of the study list, or when the encoding strategy was manipulated between participants rather than intermixed and within participants. The latter case directly tests a distinctiveness account as an explanatory mechanism for any benefit conferred by drawing at encoding.

The first allusion to a potential benefit of drawing on later memory was suggested by the dual-coding hypothesis by Paivio and Csapo (1973). They showed a memorial advantage on a free-recall test for words that were drawn rather than written out during encoding. However, they gave participants under 5 s to create a rough sketch, which was probably too brief for participants to produce a complete drawing, and thus may have underestimated the extent to which drawing benefits memory performance. Further, because participants were asked to draw pictures, an unusual task, one cannot rule out a distinctiveness account to explain the advantage. This account, invoked as a partial explanation for production effects (MacLeod et al., 2010), holds that performing an encoding task for a select subset of items adds an extra layer of encoding that distinguishes these from others, therefore making them more memorable (Conway & Gathercole, 1987). This distinctiveness explanation is bolstered by the fact that participants were only asked to write the word once as their control condition (Paivio & Csapo, 1973), which in addition to being less distinct, can be completed quicker, is arguably easier to carry out, and is less engaging than drawing, leaving additional time for the mind to stray from the encoding task at hand. Accordingly, their task manufactured an inherent disadvantage to their writing condition in that the total time spent studying each word may not have been equated across their writing and drawing conditions.

Later work (Peynircioğlu, 1989) reincarnated the study of drawing as a facilitator of memory for scenes, though in lieu of more typical retrieval tasks, participants were required to later draw the studied scenes from memory. Scenes were graded as correct retrievals only if they were drawn exactly as initially presented. They found a benefit of drawing at encoding; however, it is difficult to determine whether the observed drawing benefit was due to the act of drawing, or simply attributable to transfer-appropriate processing (Morris, Bransford, & Franks, 1977). In addition, neither of the previously discussed studies compared drawing directly to any other competing encoding strategies.

Evidence from subject-performed tasks and picture superiority

Many subject-performed tasks, which involve some additional activity at encoding, have been shown to boost memory performance. As such, it is worth considering whether drawing would lead to the same memorial benefits. Encoding tasks requiring a deep level of processing (LoP), for example, were some of the first to be explored in detail. Craik and Lockhart (1972) suggested that encoding that required deep semantic processing (e.g., deciding whether a word fits in a sentence) would lead to superior later memory performance relative to shallower perceptual-based processing (e.g., tracking whether letters in word are in upper or lower case). In the first description of another subject-performed task, generation, Slamecka and Graf (1978) found better memory for the words generated from a cue, which was a synonym of the target word, and the target word's first letter (e.g., rapid-f___), relative to words that were simply read (see Bertsch, Pesta, Wiscott, & McDaniel, 2007, for a review). Previous work has also explored the benefits of reading a word aloud during encoding, relative to silently, and showed a robust "production effect" that manifests in both within- (MacLeod et al., 2010) and between-participants designs (though to a much lesser extent; Fawcett, 2013) and even extends to writing the word (Forrin, MacLeod, & Ozubko, 2012). Finally, one of the strongest subject-performed tasks known to greatly enhance memory is enactment. In this manipulation, participants are asked to perform a movement associated with the studied word, and this has been shown to provide a boost to subsequent memory performance (Guttentag & Hunt, 1988). Given the extensive research documenting these methods of enhancing memory, it is surprising that few have considered a potentially easy-to-use means to do so, such as drawing. Currently, however, there are no comprehensive explorations of the effectiveness of drawing as a viable encoding manipulation. Accordingly, our research aims to fill this void by empirically testing the hypothesis that drawing produces significant benefits to memory performance.

Beyond the efficacy of subject-performed tasks, there are other theoretical streams that provide encouraging evidence in support of our hypothesis. The finding that images are generally better remembered than words, termed the picture superiority effect, has been well supported and replicated in the literature, consistent across various methodologies, paradigms, and demographic groups (Hockley, 2008; Kinjo & Snodgrass, 2000; Maisto & Queen, 1992; Mintzer & Snodgrass, 1999; Paivio & Foth, 1970; Paivio et al., 1968; Rowe, 1972; Snodgrass & McClure, 1975; but see Boldini, Russo, Punia, & Avons, 2007; Vaidya & Gabrieli, 2000; Weldon & Roediger, 1987). Paivio's (1971, 1991) dual-code theory suggests that pictures are better remembered than words because they are represented both visually and verbally. In our paradigm, while an image is not directly provided to our participants, their encoding task is to create one. As such, this manipulation ought to lead to benefits to memory similar to that observed in the picture superiority effect, by encouraging visualization of the word.

Why would drawing boost memory?

The available evidence, though conflicted, foreshadows that the current work may demonstrate a memory benefit as a result of drawing. If the effect is reliable, the unsolved problem remains, of determining why this would be the case. As mentioned, many subject-performed tasks that result in improved item-specific memory (e.g., production, generation, enactment) are presumed to be driven by a distinctiveness mechanism (MacLeod et al., 2010; Mulligan, 2002), wherein the items that benefit from various such mnemonic strategies do so *relatively*, and at the expense of the other, less distinct items (e.g., read silently, passive viewing). This account was explored in Experiment 7. Prior to this, however, we examined the extent to which the benefit of drawing may be driven by one of the following individual components. Specifically, we argue that the process of drawing a verbal item engages at least four components: (a) elaboration, (b) visual imagery, (c) motor action, (d) pictorial representation. We hypothesize that in order to transfer a verbal item into a drawn visual representation, participants must first generate some physical characteristics of an item (elaboration), create a visual image of the item (visual imagery), engage in the actual hand movements required of drawing (motor action), and then are left with the picture as a memory cue for later retrieval.

The current work delved first into whether drawing words at encoding would lead to improved later recall relative to writing, which has been described as a distinctiveness-based strategy in the production effect family (Forrin et al., 2012). The writing trial type was chosen in part to temper any distinctiveness-based explanations, and in part to control for the need for motor action (Component 3) involved in drawing. In Experiment 1, we measured the relative benefit of drawing using two different sets of drawing instructions. As our study was motivated in part by an interest in typical learning environments, in Experiment 2 we investigated whether any benefit of drawing would apply when the study phase was completed in a large lecture hall, where actions taken at encoding may be disruptive to others in the room. In Experiments 3 through 5, while maintaining our draw and write trial types to control for motor action (Component 3 described above), we directly contrasted these with three different encoding trial types. Each was presumed to play a fundamental role in the drawing process, and thus a likely candidate to explain any beneficial effects of drawing. In Experiment 3, we contrasted the drawing manipulation with a task requiring more elaborative semantic processing (Component 1), in order to rule out a deep level of processing as a potential explanatory factor. It is possible that drawing affords a participant memorial benefits because it encourages visualization of each word, and it is this imagery, a well-established mnemonic technique in its own right (e.g., Elliott, 1973), that is driving the effect rather than the physical act of drawing. To rule out this explanation, in Experiment 4, we compared drawing directly to a visual imagery encoding task (Component 2). In Experiment 5 we ruled out a dual-code or picture superiority mechanism for the drawing effect, by comparing drawing directly to viewing a picture of the word (Component 4) during encoding. Next we probed for potential limitations of the drawing effect by increasing the number of to-be-remembered items and decreasing the encoding time for each in Experiment 6. Lastly, in Experiment 7, we contrasted drawing and writing as between-participants conditions, to determine directly whether distinctiveness, invoked to explain the production effect (MacLeod et al., 2010), was also the mechanism driving the drawing effect.

We propose that in addition to encouraging deeper semantic processing, visual imagery, and picture superiority, the act of drawing provides some mechanical information, akin to when words are enacted. Accordingly, in all cases these motor contributions (Component 3) are controlled for by including repetitive writing as a control. Our proposed mechanism for this drawing effect, which is addressed in more detail in the General Discussion, is that by encouraging the cohesion of multiple modes of representation of target information, we effectively create a more resilient trace, which is robust in the face of changes in setting, instructions, or competing encoding strategies.

EXPERIMENT 1: INITIAL DEMONSTRATION

The aim of Experiment 1 was to determine whether, in a controlled and relatively simple paradigm, drawing would provide a benefit to later memory performance. The basic framework of the paradigm remains consistent across all experiments in this study, with only minor deviations in task instructions. In this initial experiment, recall of incidentally encoded words was tested after a brief retention interval. In Experiment 1A, during incidental encoding, words were either drawn in detail or written out multiple times. Based on previous work (Paivio, 1971; Paivio & Csapo, 1973), we predicted that despite the inherent advantage that production would afford written words (Forrin et al., 2012), drawn words would be better recalled than written ones.

We were concerned that instructing participants to add detail to their drawing, but not their writing, was implicitly biasing participants toward favouring retention of the drawn items. Further, it is possible that semantic satiation, or the loss of meaning that occurs with repeated exposure, would occur with repeated writing (Balota & Black, 1997). To control for this confound, the instructions were crossed over in Experiment 1B, such that the drawings were to be done multiple times, and the writing done once with the instruction to add detail to the writing. We predicted that the advantage of drawing would still manifest under these instructions.

Method

Participants

Participants for Experiment 1A were 30 undergraduate students (19 female), and for 1B were 25 undergraduate students (9 female) at the University of Waterloo, who completed the experiment for course credit or monetary remuneration. Participants ranged in age from 18 to 47 years ($M = 20.67$, $SD = 4.15$), with between 14 and 27 years of education ($M = 16.72$, $SD = 2.22$). All participants had normal or corrected-to-normal vision, and learned English before the age of seven.

Materials

Target items. An 80-item word list (see Appendix A) was created from a selection of the verbal labels for Snodgrass images (Snodgrass & Vanderwart, 1980), to ensure that all words could be easily drawn. Complex drawings were avoided (e.g., clown) in favour of simpler items (e.g., apple). This measure was taken to reduce the time it would take participants to create each of the drawings; every word could be drawn in the time provided, based on a pilot study, and no item required excessive visual detail to be discernable. Words ranged in frequency between 1 and 25 ($M = 8.23$, $SD = 6.44$), in length between 3 and 11 letters ($M = 5.56$, $SD = 1.79$), and in number of syllables from 1 to 4 ($M = 1.63$, $SD = 0.72$).

Filler task. A continuous reaction time task (CRT) was created by making sound files representing low-, medium-, and high-pitched tones. This was done using Audacity software (Mazzoni & Dannenberg, 2000), such that each sine wave tone was exactly 500 ms long, at frequencies of 350, 500, and 650 Hz, respectively.

Questionnaires. We also asked participants to complete the Vividness of Visual Imagery Questionnaire (VVIQ), created by Marks (1973), and three questions regarding participants' history of drawing. The VVIQ is a short questionnaire that assesses individual differences in ability to create a mental image of an item or scene. Individuals are provided with four scenarios and, for each, are asked to rate how clearly they can visualize nuanced aspects of each of the scenarios on a 5-point scale. There were no reliable correlations with either questionnaire throughout the course of study. Accordingly, more detailed methodological information and data regarding these correlations are presented in the Supplemental Material.

Procedure

Participants completed the experiment individually in a testing room. Stimulus presentation and response recording were controlled using E-prime v2.0 software (Psychology Software Tools Inc., Pittsburgh, PA) via an IBM computer with 17-inch monitor. Instructions were presented in English on screen and were also read aloud by the experimenter. Participants were told that, depending on the "prompt" word they saw, they were to either "draw" or "write" the subsequent word on the pad of paper (14 cm × 21 cm) provided. In Experiment 1A, a prompt of "draw" meant the participant was to draw a picture illustrating the word on the screen and to continue adding detail until their allotted time was exhausted. A prompt of "write" meant the participants were to clearly and carefully write out the word multiple times. In Experiment 1B, the instructions were crossed over, such that "draw" meant they were to repeatedly draw the item presented, while "write" meant they were to continue adding detail to a single iteration of writing the word. This instruction was unorthodox, but was included to potentially shift emphasis to the write condition. Participants' response to this instruction (see Appendix B) included using block letters, adding decorative flourishes to their letters, shading, or incorporating some elements of the concept into the writing (e.g., music notes on the word "harp"). They were informed of time constraints for each item and that they would hear a tone to warn them that the next item would appear. Participants were not told that their memory would be tested.

Encoding. Participants underwent a brief practice phase in order to familiarize them with the encoding phase, after which the experiment began. Participants were not informed that they would be required to complete a later memory test. This incidental encoding paradigm was selected to reduce the possibility that participants would develop a strategy of preferentially focusing on drawn items in anticipation of later testing. From the list of 80 words, 30 were randomly selected to be studied, a list unique for each participant. Of these 30, 15 were randomly selected to be drawn, and 15 written (see Appendix B for samples of the outcome of each trial type). This set of words was then presented in a randomized order, such that drawn and written items were randomly intermixed. On each trial, the prompt appeared in the centre of the screen for 750 ms, followed by a 500-ms fixation, after which the word to be encoded appeared for 750 ms. Participants then had 40 s to perform the encoding task, either draw or write. A 500-ms tone alerted them that the next item was forthcoming, after which they had 3 s to flip their pad of paper to the next page in preparation for the next prompt.

Retention. Following the encoding trials, participants were asked to perform the CRT as a filler task. Tones were to be classified as low, medium, or high, by pressing the 1, 2, or 3 key on a small response pad. After hearing samples of each kind of tone, participants proceeded to classify 60 tones, selected at random. For each trial, the tone was played for 500 ms, after which participants had 1500 ms to make their response, for a total of 2000 ms per trial. Thus the retention interval was two minutes.

Retrieval. In the next phase of the experiment, participants were asked to freely recall as many words as they could, in any order, either written or drawn, from earlier in the experiment. They were given 60 s to complete their recall, which was spoken aloud by the participant and recorded.

Questionnaires. Immediately following the retrieval phase, participants completed a version of the VVIQ (Marks, 1973) and the three questions pertaining to drawing experience and ability (see Supplemental Material).

Results and discussion

Participants' recall output was sorted into the number of recalled words that were drawn at encoding and the number that were written. Each of these values was divided by the total number of words studied within each encoding trial type (15), to create a proportional recall score. Data were analysed in a 2 × 2 mixed measures analysis of variance (ANOVA), with Experiment (1A, 1B) as a between-participants variable and encoding (draw, write) as a within-participants variable. Analyses showed a significant main effect of encoding trial type, $F(1, 53) = 82.83$, $MSE = .02$, $p < .001$, $\eta^2 = .61$, such that drawn words were recalled better than written words (see Figure 1). The average number of times participants wrote out the word in Experiment 1A ($M = 16.71$, $SD = 5.19$) or drew the word in Experiment 1B ($M = 2.96$, $SD = 3.85$) was not correlated with the number of recalled drawn items, recalled written items, total recalled items, or the magnitude of the benefit of drawing, $p\text{s} > .21$. Twenty-six of the 30 participants in 1A showed the pattern of

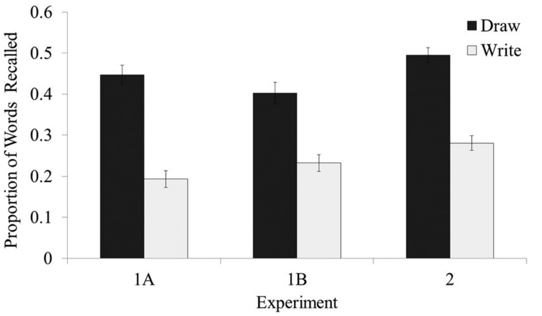

Figure 1. *Proportion of words recalled out of 15 in Experiments 1A, 1B, and 2, from the draw, list, and write trial types. Error bars represent the standard error of each mean.*

improved recall for drawn relative to written items, $t(29) = 8.29$, $SE = .03$, $p < .001$, $d = 1.51$, while 20 of the 25 did in 1B, $t(24) = 4.81$, $SE = .04$, $p < .001$, $d = 0.96$. Hereafter, we use the term "drawing effect" to refer to this distinct advantage of drawing words relative to writing them out in terms of later memory performance.

The main effect of experiment and the interaction were not significant, allowing us to conclude that first, there was no overall difference in memory performance depending on whether participants were instructed to add detail to one trial type or the other and, second, that the benefit of drawing did not differ depending on the instruction to add detail or not. Such a finding rules out the confound that the instruction to add detail on drawn trial types (Experiment 1A) could account for the beneficial effect of drawing on subsequent memory, since the instruction to add detail to the written trial types did not confer the same benefit (Experiment 1B).

First and most importantly, our results indicated that there was a significant recall advantage for words that were drawn during incidental encoding as compared to those that were written. Participants recalled more than two times as many drawn than written words, and most participants recalled more drawn than written words. This finding expands upon previous work suggesting some special advantage of drawing (Paivio & Csapo, 1973; Peynircioğlu, 1989). Contrary to previous work, however, we carefully controlled the time our participants were given to complete their drawings, thereby allowing sufficient time for the creation of said drawing, and also equated it with the time given for writing trial types. Unlike in previous work, in which participants wrote the word once then sat idle, we created trial types wherein the entirety of the time allotted for each trial type was most likely to be used. Our design thus increased the likelihood that participants were processing the word throughout the entire encoding trial, rather than engaging in task-unrelated thoughts.

Second, because the "draw" instruction in Experiment 1A was to add detail to drawings to fill the study time for each trial, while the "write" instruction was to simply rewrite each word on a given trial, we wanted to ensure that the drawing benefit was not just a result of this difference in instruction. For example, it is possible that the instruction to repeatedly write out the word on each instance of the "write" trial type was reducing access to semantic information about the words, thereby reducing subsequent memory (Balota & Black, 1997). In Experiment 1B, despite removing repetition in the "write" trial types, we still found a relative memorial benefit for drawn words. When participants were told to add detail to their writing, presumably shifting the emphasis to the "write" trial type, and removing repetition, the drawing effect was still evident, and the advantage was shown in the majority of participants.

EXPERIMENT 2: GROUP TESTING

In Experiment 1 we showed that drawing, compared to copying words during study, afforded participants a significant benefit in their later recall. In the next experiment we wanted to explore whether this was a useful strategy that students could use in a group classroom setting. Other mnemonic strategies, such as production and enactment, could be quite distracting in a group setting, making them less suited for such public environments. Accordingly, while the previous experiment was completed individually in a testing room devoid of distractions, we aimed to determine whether drawing could generalize to a group lecture hall setting in Experiment 2.

It is possible that in this context, the drawing effect would still manifest, but that it might be smaller due to distraction from the activity or movement of other participants in a classroom setting. As well, due to the group setting, recall in this experiment was written rather than aloud. The transfer-appropriate processing theory (Morris et al., 1977) would dictate that since words were to be output using the same format as that at study, a written memory test should lead to a more robust reexperiencing of the encoding experience for target words that were written than for those that were drawn, providing them with a memory boost. The end result of the change in test format might therefore be a reduction in our reported drawing effect. In addition to looking at whether the drawing effect was maintained even in the classroom setting, we analysed the combined data from Experiment 1 and 2 in order to determine whether there were any possible interactions based on test setting.

Method

Participants

Participants in Experiment 2 were 49 undergraduate students (38 female) at the University of Waterloo, who completed the experiment in return for course credit or monetary remuneration. Participants ranged in age from 17 to 24 years (M = 19.10, SD = 1.26), with between 13 and 18 years of education (M = 14.78, SD = 1.12). All participants had normal or corrected-to-normal vision and learned English before the age of seven.

Materials

Study words and tones used for the filler task were identical to those in Experiment 1.

Procedure

Participants were tested in two different group sessions in a large lecture hall with stadium-style seating for up to 126 students. One session consisted of 15 and another of 34 undergraduate students. Stimulus presentation and response recording was controlled using E-prime v2.0 software (Psychology Software Tools Inc., Pittsburgh, PA) and displayed via projection screen at the front of the lecture hall. Given the group setting, some additional preliminary instructions were given, to ensure participants did not interact with one another or look at one another's responses: Participants were always separated by at least one empty seat and were instructed to treat the experiment as they would an examination.

The instructions from Experiment 1A were used here, as the data from Experiment 1B ruled out any differential effects due to instruction. Due to the group setting, some obvious changes were necessary to avoid participants' responses interfering with others' focus. Specifically, responses in the tone classification task were done using pen and paper, as was their recall output and response to the questionnaires. Apart from this, the procedure was identical to that of Experiment 1A.

Results and discussion

Data were analysed in a paired-samples t test, with encoding trial type (draw, write) as a within-participants variable. Analyses showed a significant main effect of encoding trial type, $t(48) = 9.11$, $SE = .03$, $p < .001$, $d = 1.30$, such that drawn words were better recalled than written words (see Figure 1). Forty-three of the 49 participants showed this pattern. Our results indicate that the drawing effect reported in Experiment 1 is resilient to changes in environment and is thus generalizable to a group lecture setting. What they do not tell us, however, is whether the change in setting altered the size of the drawing effect. For this reason, data from Experiments 1 and 2 were compared for the following analyses.

Comparison with Experiment 1

Given that the group lecture hall setting is a much more efficient way of testing participants and is probably more analogous to an actual learning environment in the real world, additional analyses were conducted to ensure that our effect of drawing was not interacting with the experimental setting. Data from Experiments 1A and 1B, in which participants were tested individually, and Experiment 2, in which participants were tested in groups, were combined into a pooled analysis. A 2 × 2 mixed measures ANOVA was conducted with setting (individual, group) as a between-participants variable, and encoding trial type (draw, write), as a within-participants variable. There was a significant main effect of encoding trial type, $F(1, 102) = 165.06$, $MSE = .02$, $p < .001$, $\eta^2 = .62$, such that drawn words were better recalled than written words. Interestingly, the main of setting was also significant, $F(1, 102) = 16.24$, $MSE = .02$, $p < .001$, $\eta^2 = .14$, such that those participants who were tested in groups performed much better than those tested individually, recalling nearly two more words than their individually tested counterparts. The Encoding Trial Type × Setting interaction, however, was not significant, $F(1, 102) = 0.001$, $MSE = .02$, $p > .05$, $\eta^2 = .00$. In other words, the memory boost afforded by drawing an item as compared to writing it out was stable when the task was moved from an individual to a group lecture setting. This finding should serve to assuage any concerns about the venue of testing. Accordingly, our remaining empirical questions were often addressed in a group setting.

The aim of the subsequent three experiments was to directly contrast drawing with what we believe to be primary components of the progression from a verbal item to a participant-created drawing of that item, its visual referent. As a reminder, these components include (a) elaboration, (b) visual imagery, (c) motor action, (d) pictorial representation. In all cases, writing was included as a baseline partially to control for distinctiveness, but primarily to control for the motor action (Component 3) that drawing necessitates.

EXPERIMENT 3: DEEP LEVEL OF PROCESSING

Having ruled out the alternative account that our instruction to add detail was producing our drawing effect (in Experiment 1B), the aim of Experiment 3 was to rule out another possible mechanism driving the drawing effect. We sought to determine whether the observed benefit resulting from drawing relative to writing words during a "study phase" occurred because drawing encourages a deeper, more elaborative level of processing. As outlined by Craik and Lockhart (1972) in their LoP framework, a stimulus that is encoded at a "deeper" level is one that evokes a greater degree of semantic analysis. This framework suggests that items that undergo deeper encoding promote enrichment or elaboration that leads to greater success during later retrieval than for items that were encoded shallowly. It follows that creating a drawing of an item would probably require recognition of the stimulus, then some elaboration in order to recreate one's verbal representation in a sketch. It has been suggested that this elaboration, even in expert artists, draws first from denotative semantic information, including knowledge of physical characteristics of the to-be-drawn item rather than visual imagery (McMahon, 2002).

To rule out LoP as an explanation for the drawing advantage, we considered the influence of a "list" trial type during encoding, which required participants to write a list of semantic characteristics of the presented target word, thereby encouraging more elaborative processing. It is important to note that while most previous work investigating the effect of a deep LoP had a companion "shallow" LoP condition (e.g., indicating whether a word has the letter "e" or "g"; Walsh & Jenkins, 1973), our paradigm included only the deep LoP "list" trial type, along with our original "draw" and "write" trial types. Notably, the "write" trial type is far from a shallow LoP condition, as even writing a word once resulted in 13% more hits in a recognition task than passive reading (Forrin et al., 2012).

In a sense, we were also stacking the odds against our drawing trial types in this experiment. The words in the "list" trial type were to be encoded at a deep level, and in line with the levels of processing framework, they should enjoy memorial benefits. Second, as in previous experiments, transfer-appropriate processing (Morris et al., 1977) and distinctiveness (Forrin et al.,

2012) should provide some benefit to written items. Lastly, previous work showed that a semantic encoding task led to words being better remembered than pictures (Vaidya & Gabrieli, 2000). It follows from this result that because our encoding task in the "list" trial type required semantic processing, those words would show a stronger memorial advantage than those in the "draw" trial types at encoding. If the drawing effect still occurs, which we hypothesize it will based on our prior experiments, it would indicate firmly that drawing is an extremely robust and reliable encoding manipulation, and one that cannot be explained by elaborative encoding alone.

Method

Participants

Participants were 47 undergraduate students (40 female) at the University of Waterloo, who completed the experiment in return for course credit or monetary remuneration. Participants ranged in age from 17 to 51 years ($M = 20.17$, $SD = 5.07$), with between 13 and 18 years of education ($M = 15.26$, $SD = 1.37$). All participants had normal or corrected-to-normal vision and learned English before the age of seven.

Materials

Word stimuli and tones were the same as those in previous experiments.

Procedure

The procedure was identical to that used in Experiment 2, except that a third trial type was added during encoding. Participants were instructed to "draw", "list", or "write" the word being presented to them. Instructions for "draw" and "write" were identical to those in previous experiments. For the "list" prompt, participants were instructed to write out a list of physical characteristics of the word presented until the time for that trial type ended. Because there were now three different prompts instead of two, the randomly selected list of 30 words was divided into three lists of 10 words each (10 to be drawn, 10 to be visualized, 10 to be written) instead of two lists of 15 words of each trial type as in the prior experiments. This set of words was then presented in a randomized order, such that drawn, visualized, and written items were randomly intermixed. Apart from these modifications, the experimental protocol was identical to that of Experiment 2.

Results and discussion

Data from one participant were excluded as they performed the incorrect encoding instruction for 9 of the 30 words. Data from the remaining participants were analysed using repeated measures ANOVA, with encoding trial type (draw, list, write) as a within-participants variable. Results indicated a significant main effect of encoding trial type, $F(2, 92) = 21.20$, $MSE = .02$, $p < .001$, $\eta^2 = .32$. Paired-samples t tests using Bonferroni adjusted alpha levels of .0167 per test (.05/3) revealed that this was driven by significantly better recall for words in the draw than in both the list, $t(46) = 5.16$, $SE = .03$, $p < .001$, $d = 0.75$, and the write trial types, $t(46) = 5.89$, $SE = .03$, $p < .001$, $d = 0.86$. The difference in recall between words from the list and write trial types was not significant, $t(46) = 0.39$, $SE = .03$, $p = .70$, $d = 0.06$ (see Figure 2). In the "list" trial types, participants listed a mean of 6.37 characteristics ($SD = 1.65$) per word, but the number of characteristics listed was not correlated with the number of

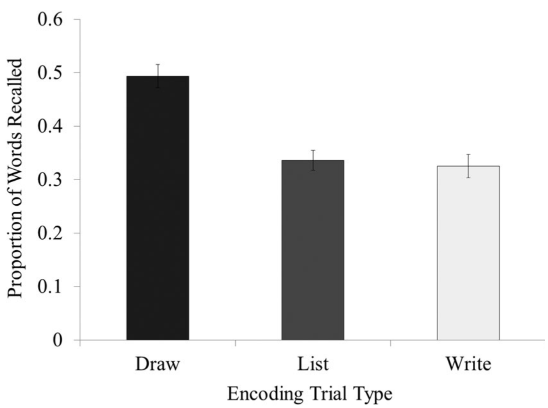

Figure 2. *Proportion of words recalled out of 10 in Experiment 3 from the draw, list, and write trial types. Error bars represent the standard error of each mean.*

words recalled overall, or in any of the three encoding trial types, $ps > .11$.

Results indicated again that drawing was superior to both of the other encoding instructions (listing physical characteristics or writing out the word). Such a finding suggests that the effect of drawing cannot be dismissed as another iteration of a classic deep LoP manipulation. What is more, our deep LoP manipulation, which required "listing" semantic characteristics describing the study word, showed no significant advantage over a "writing" instruction. Some might argue that our deep LoP trial type ought to have improved memory performance relative to writing out the studied word repeatedly. As mentioned previously, deep LoP manipulations typically are compared to shallow LoP counterparts (Walsh & Jenkins, 1973), something that was absent in our paradigm. The aforementioned research showed that deep processing is superior to shallow, but our "write" trial type was not shallow in the traditional sense, and this may explain why "list" was not superior to "write". To clarify, writing a word out multiple times over the span of 40 s may have allowed participants to think more deeply about the word or visualize the item, as the task of writing a word repeatedly is probably not as cognitively demanding as the other trial types. Accordingly, our "write" trial type cannot be categorically considered to be a shallow LoP trial type, which may explain the absence of a significant difference in recall of words from the "list" relative to "write" trial types. That the drawing effect was larger than the benefit conferred to words receiving a deep LoP orientation at encoding, known to reliably boost memory (Walsh & Jenkins, 1973), is a testament to the strength of the effect. Further, listing a greater number of characteristics for any particular word did not lead to better memory for that word, suggesting that regardless of how deeply words were encoded in the listing task, the benefit of drawing as an encoding orientation remained superior.

EXPERIMENT 4: VISUAL IMAGERY

It was worth considering that the benefit from drawing at encoding occurred as a result of adding visual imagery to the memory trace, as dual-code theory suggests is the case when creating a memory for a picture (Paivio, 1971). As discussed, drawing requires the translation of material from a verbal into a visual code. While retrieving physical characteristics, as in Experiment 3, is a likely contributor to this translation process, it seems intuitive that participants may also engage visual imagery as a subsequent or even concurrent step. One must create a mental image of their prototypical visual representation of that word, so that they can attempt to reproduce this prototype through drawing. Visual imagery alone has been shown to improve memory performance relative to rote repetition (Elliott, 1973; Kieras, 1978; Lupiani, 1977; Winnick & Brody, 1984), so it is very possible that visual imagery is a contributor to, and potential explanatory mechanism for, any benefit documented as a result of drawing.

As discussed, Paivio (1971) posited that memory for pictures was superior to that of words due to the fact that they invoke both the readily apparent visual representation and also the representation of the picture's verbal referent. In other words, upon seeing a picture of a horse, one readily retrieves the word "horse" as well, providing two codes for memory. Thus, this theory suggests that disparate memory performance for target pictures and words is driven by differences in visual imagery. By encouraging participants to draw a studied word, we are coaxing a detailed visual representation in addition to their existing verbal code, thereby asking participants to create a (parallel) dually coded memory trace.

In order to rule out the possibility that the drawing effect is simply a residual outcome of differential invocation of visual imagery, we introduced an additional competing encoding instruction to "visualize" or engage in mental imagery of a study word, to compare to the trial types in which "draw" or "write" instructions were given. The dual-code theory (Paivio, 1971) led us to hypothesize that the instruction to create a mental image would boost subsequent memory, relative to the instruction to simply write out the word.

Importantly, though we speculated that creating a mental image would boost memory, we predicted that this boost would be smaller than that from our drawing manipulation. This is because with the drawing instruction, participants must not only create a mental image of the word they are presented with and process the meaning of this word, but also undergo the mechanistic process of moving the pencil to create the image, which may be akin to a muted enactment effect. In other words, whereas the "visualize" trial type encourages mental imagery, the "draw" trial type not only encourages imagery, but also a deeper level of processing and potentially a form of enactment. Additionally, the participant is left with an actual picture to use as a cue once they are finished drawing. By requiring integration of these aspects of the memory trace, we expect that drawing will have an advantage over typical writing or visual imagery alone.

Method

Participants

Participants were 28 undergraduate students (21 female) at the University of Waterloo, who completed the experiment in return for course credit or monetary remuneration. Participants ranged in age from 19 to 23 years ($M = 20.29$, $SD = 1.01$), with between 14 to 18 years of education ($M = 15.86$, $SD = 1.04$). All participants had normal or corrected-to-normal vision and learned English before the age of seven.

Materials

Word stimuli and tones were the same as those in previous experiments.

Procedure

The procedure was the same as that of Experiment 3, except for one substitution. The prompt "list" was replaced with "visualize", such that the three prompts were "draw", "visualize", and "write". A prompt of "visualize" meant the participant was to create a mental image of what the word represents. They were instructed to maintain their focus on the image, adding detail to it until that trial type time had elapsed.

Results and discussion

Data were analysed using repeated measures ANOVA, with encoding trial type (draw, visualize, write) as a within-participants factor. Results indicated a significant main effect of encoding trial type, $F(2, 54) = 8.23$, $MSE = .04$, $p < .005$, $\eta^2 = .23$. Paired-samples t tests using Bonferroni adjusted alpha levels of .0167 per test (.05/3) revealed that this was driven by significantly better recall for drawn trial types than for write trial types, $t(27) = 4.62$, $SE = .04$, $p < .001$, $d = 0.87$, and marginally better recall for drawn than for visualized trial types, $t(27) = 2.20$, $SE = .06$, $p = .037$, $d = 0.42$. The difference in recall between visualize and write trial types was not significant, $t(27) = 1.58$, $SE = .05$, $p = .126$, $d = 0.30$ (see Figure 3).

Results from Experiment 4 suggest that instructions to create a mental image of a word during encoding do not confer a memorial benefit relative to writing or drawing the word. We acknowledge, however, that visual imagery is a manipulation that presents difficulties in verifying participant compliance. As a reviewer pointed out, this could result in lazy imaging or writing, akin to the lazy reading hypothesis presented to explain the production effect. This is a limitation inherent in the duration provided for encoding, but a confound we address, at least for the general drawing effect,

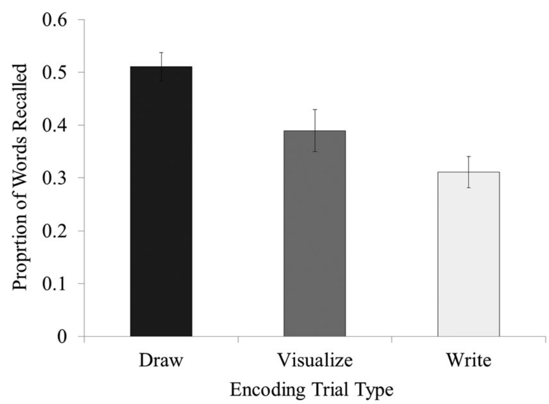

Figure 3. *Proportion of words recalled out of 10 in Experiment 4 from the draw, visualize, and write trial types. Error bars represent the standard error of each mean.*

in Experiment 6. Importantly though, assuming participants were compliant with our instructions, drawing led to much better memory performance than visualizing, Thus, we can rule out visual imagery alone as a potential mechanism by which drawing exerts its memorial benefit and move forward in exploring picture superiority as a potential mechanism.

EXPERIMENT 5: PICTURE SUPERIORITY

Picture superiority, the reliable effect wherein pictures are better remembered than words (Paivio et al., 1968), was detailed earlier as a theoretical reason for pursuing drawing as an encoding strategy. It is perhaps the most obvious alternative explanation for our reported drawing benefit. After having retrieved a set of physical characteristics describing the to-be-drawn item, and generating a visual image of that item, participants must produce the action necessary to produce the drawing. Participants are then left with an image to associate with its encoded verbal referent. Accordingly, one could argue that any benefit observed as a result of drawing is simply due to the creation of pictures, which we know are better remembered than words. If this were the case, one would expect that memory for drawn words would be no different from memory for pictures of those words.

In order to more directly assess whether picture superiority could be the mechanism through which drawing exerts its influence, we replaced the "visualize" condition in Experiment 4 with a "view" condition, wherein participants were instructed to simply view an image of the studied word.

Method

Participants

Participants were 37 undergraduate students (29 female) at the University of Waterloo, who completed the experiment in return for course credit or monetary remuneration. Participants ranged in age from 18 to 23 years ($M = 19.14$, $SD = 1.40$), with between 14 and 19 years of education ($M = 14.78$, $SD = 1.16$). All participants had normal or corrected-to-normal vision and learned English before the age of seven.

Materials

Word stimuli and tones were the same as those in previous experiments. A picture was chosen for each word stimulus from a picture set containing thousands of unique images of objects (Brady, Konkle, Alvarez, & Oliva, 2008). If the object was not found in this stimulus set, one was retrieved using a Google image search. All images were cropped to the same dimensions (256 × 256 pixels) and were converted to greyscale using ImageMagick software (ImageMagick Studio LLC, 1999–2013).

Procedure

The procedure was the same as that of Experiment 4, except for one minor change. The prompt "visualize" was replaced with "view", such that the three prompts presented during encoding were "draw", "view", and "write". For the view prompt, participants were instructed to view a picture representative of the to-be-remembered word. In this condition, after the prompt and word disappeared, a picture of what the word represents appeared on screen, and remained on screen for the remainder of the encoding time.

Results and discussion

Data were analysed using repeated measures ANOVA, with encoding trial type (draw, view, write) as a within-participants variable. Results indicated a significant main effect of encoding trial type, $F(2, 72) = 7.24$, $MSE = .03$, $p < .005$, $\eta^2 = .17$. Paired-samples t tests using Bonferroni adjusted alpha levels of .0167 per test (.05/3) revealed that this was driven by significantly better recall for words in the draw than in both the view, $t(36) = 2.56$, $SE = .04$, $p < .0167$, $d = 0.42$, and the write trial types, $t(36) = 4.08$, $SE = .04$, $p < .001$, $d = 0.67$. The difference in recall between words from the view and write trial types

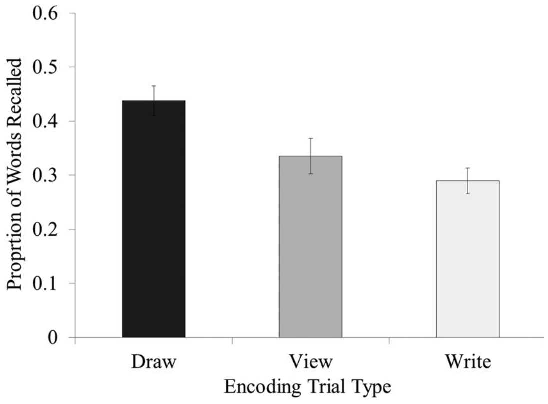

Figure 4. *Proportion of words recalled out of 10 in Experiment 5 from the draw, view, and write trial types. Error bars represent the standard error of each mean.*

was not significant, $t(36) = 1.07$, $SE = .04$, $p = .293$, $d = 0.18$ (see Figure 4).

Results indicate that drawing led to better recall than viewing images of the to-be-remembered words, and, therefore, the effects of picture superiority alone cannot be invoked as an explanation for the observed benefit of drawing on memory.

EXPERIMENT 6: BRIEF ENCODING TIME

All experiments thus far have required that participants study 30 words, with 40 s to perform the required encoding strategy for each word. Within these 40 s, it is possible, as a reviewer pointed out, that participants may not be engaging in the required encoding strategy for the entirety of the allotted encoding time. This has been described as the problem of lazy reading in the production effect, though when this concern was controlled for, the effect still remained (MacLeod et al., 2010). This is certainly a concern with our prior experiments, especially in our less directly observable manipulations (visualize, view), but also arguably in the writing encoding trial types as well. Given this, the goal of Experiment 6 was to address some of these concerns experimentally, as well as to search for potential boundary conditions of the drawing effect. Some might argue that the 40 s allotted for drawing, and the various encoding trial types in the first five experiments, was excessive and thus not as practical as other mnemonic strategies that can be completed in periods of time under 5 s. Further, we used relatively short lists in the foregoing experiments as a direct result of our lengthy encoding duration. To address these concerns, in Experiment 6, the number of studied items was more than doubled, and the encoding duration for each item was reduced from 40 to 4 s. If the drawing effect was due to this long encoding duration, or due to "lazy writing" over the course of 40 s, then the drawing effect should not be present when encoding duration is drastically reduced. With this methodological change, the drawing effect might still occur; although we predicted that the effect may be smaller, due to the fact that the drawing created within a 4-s time period would probably contain fewer details than a drawing created in 40 s.

Method

Participants

Participants were 28 undergraduate students (22 female) at the University of Waterloo, who completed the experiment in return for course credit or monetary remuneration. Participants ranged in age from 18 to 25 years ($M = 20.64$, $SD = 1.90$), with between 13 and 22 years of education ($M = 16.41$, $SD = 1.96$). All participants had normal or corrected-to-normal vision and learned English before the age of seven.

Materials

Word stimuli were the same as those in previous experiments. Using a notepad and flipping the pages between each stimulus at encoding was not conducive to this new rapid presentation of stimuli. Accordingly, we switched to using a Fisher-Price Doodle Pro®. This small drawing pad has roughly the same drawable surface as did our drawing pads (11 cm × 16 cm), but uses magnetic drawing technology similar to that of an Etch-a-Sketch®. Unlike an Etch-a-Sketch®, the Doodle Pro® comes with a small stylus to draw, rather than knobs, making it more analogous to the paper and pencil technique used in the first five experiments. This product features a sliding knob on the top, which quickly wipes clean the drawing pad in preparation for the next trial.

Procedure
The procedure was very similar to Experiment 1A, with the exception of some timing and presentation changes to support the more rapid encoding time. Additionally, in order to facilitate comparison with investigations of other encoding strategies thought to depend on distinctiveness (McDaniel & Bugg, 2008), this experiment employed intentional, rather than incidental, encoding. Like Experiment 1A, participants were tested individually in a testing room. Participants studied 66 words instead of 30, split equally between the write and draw encoding trial types. These 66 words were then presented with drawn and written items randomly intermixed. During encoding, rather than the typical protocol of showing a participant a prompt (750 ms), a fixation (500 ms), and the word (750 ms), and then asking them to complete the task (40 s), participants were simply shown the word. Words were presented on a black background in either white or red font, and participants were instructed that the font color indicated whether they were to draw or write the item (with colors counterbalanced across participants). Each word was presented for 4 s, during which time the participant was to complete the task indicated by the font color of the word. Between trials, a blank screen was shown for 1 second to give participants time to reset their drawing pads.

The retention task from the previous experiment was replaced with a 5-min tone classification and a 5-min visual CRT task, in which participants identified within which three numbered boxes (left, middle, or right) an asterisk was presented, using the keys corresponding to the numbers presented above the boxes (1, 2, 3).

Participants were then given two minutes to recall as many words as they could from the encoding phase, by typing them in an input field. Participants heard a tone to notify them when there were 20 s remaining in the recall test. Participants did not complete the VVIQ or the drawing-related questions after this experiment.

Results and discussion

First, the data were analysed with cue colour (red or white) as a between-participants factor. There were no main effects or interactions with cue colour, so the data were collapsed across this variable. Data were analysed in a paired-samples *t* test, with encoding trial type (draw, write) as a within-participants variable. Analyses showed a significant main effect of encoding trial type, $t(27) = 12.01$, $SE = .02$, $p < .001$, $d = 2.27$, such that drawn words were better recalled than written words (see Figure 5).

These results indicate that even with a longer study list, a much shorter encoding duration, and a slightly longer retention interval between study and test, drawn words were remembered much better than written ones. Further, the effect was substantially more pronounced (see Figure 5) than in previous experiments (see Figures 1–4), indicating that the drawing encoding strategy is potentially even more potent at shorter encoding durations. It is worth noting, however, that it is also possible that writing was simply a *less* potent strategy at shorter encoding strategies, or that this larger effect size was an artefact of the change from incidental to intentional encoding.

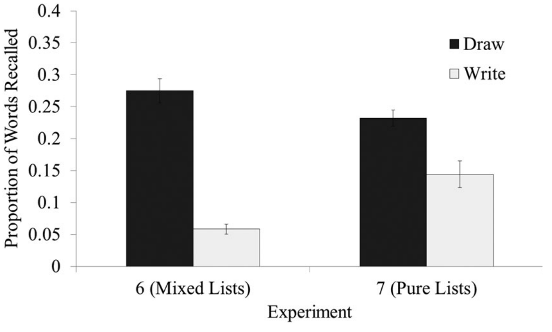

Figure 5. *Mean proportion of words recalled out of 33 for each trial type in Experiment 6 (mixed lists) and out of 66 for each trial type in Experiment 7 (pure lists). Error bars represent the standard error of each mean.*

Comparisons between the results of Experiment 6 and those of the prior experiments may be problematic because the paradigms differ in more ways (e.g., list length, intentional learning) than simply encoding duration.

EXPERIMENT 7: BETWEEN PARTICIPANTS

Having ruled out emphasis on adding detail to a particular encoding trial type (the instruction to continue adding detail to drawing, or writing in Experiment 1), and each of the three components proposed to be integral to the drawing process (elaboration, visual imagery, and picture superiority in Experiments 3–5), we turned our focus to examining whether the drawing effect can be accounted for, like many other item-specific encoding strategies (McDaniel & Bugg, 2008; Nairne, Riegler, & Serra, 1991), by a distinctiveness account. We maintain our proposal that drawing exerts its memorial benefits through integration of a number of different types of memory codes into one cohesive trace. Alternatively, though, drawing could be exerting its effects by making some items more distinct than others. The current work draws obvious parallels with the production effect, wherein item memory for words read aloud is superior to that for words read silently (MacLeod et al., 2010). The mechanism commonly proposed to explain this, and other item-specific mnemonic strategies, is a distinctiveness account (Bodner & Taikh, 2012; MacLeod et al., 2010; McDaniel & Bugg, 2008; Ozubko & MacLeod, 2010). In other words, in a mixed-list paradigm, some items are read aloud (or generated from a cue, or enacted), while others are simply read silently (or enacted by experimenters while the participant passively views; Engelkamp & Zimmer, 1997; Slamecka & Graf, 1978). Accordingly, the trials that are more unique, in which a participant completes a novel or bizarre encoding task, have some distinctive information that can later be used as an extra retrieval cue. The empirical support for this account comes from a commonly observed absence or reversal of most of these effects when trial types are completed between participants, in pure lists, as opposed to within participants. This is true of production (MacLeod et al., 2010; but see Bodner, Taikh, & Fawcett, 2014), generation (Slamecka & Katsaiti, 1987), enactment (Engelkamp & Dehn, 2000), and many other commonly cited mnemonic strategies (McDaniel & Bugg, 2008; Mulligan, 2002). While most of the foregoing findings used recognition, the experiments in the current work used free recall. Some previous work has explored how the production effect is modulated depending on whether trial types are compared within or between subjects. In contrast to recognition, where some previous work has found a significant between-participants production effect (e.g., Bodner et al., 2014), when employing free recall, there is no evidence for the production effect in pure lists (Jones & Pyc, 2014; Jonker, Levene, & MacLeod, 2014). What this tells us is that the benefit of most of these effects manifests as a *relative* enhancement, rather than an absolute one. Specifically, distinctive items are better encoded at the expense of the less distinctive ones. In an attempt to counter this potential account of the observed drawing effect, throughout all of our experiments thus far, we employed writing as a baseline condition, with which to compare drawing. This baseline was chosen as writing has been previously described as an iteration of the production effect, and as a task that is distinctive in its own right (Forrin et al., 2012). Further, we compared drawing with alternate encoding trial types of visual imagery, elaborative encoding, and picture superiority, which could also be categorized under the distinctive encoding umbrella.

However, in order to conduct the most direct experimental exploration of a distinctiveness account, in Experiment 7 we compared drawing and writing encoding strategies in pure lists, a manipulation that typically eliminates the effects of other encoding strategies that have been described as relying on distinctiveness (McDaniel & Bugg, 2008; Mulligan, 2002). With this, if the benefit of drawing is driven exclusively by distinctiveness of drawn relative to written items, we would not observe a benefit of drawing when compared between participants. If, however, distinctiveness can only partially explain our findings, which we expect will be the case given the large effect size we observed thus far, then results should indicate an extant, albeit smaller, boost as a result of drawing items, relative to writing them.

Method

Participants

Participants were 47 undergraduate students (37 female) at the University of Waterloo, who completed the experiment in return for course credit or monetary remuneration. Participants ranged in age from 18 to 24 years ($M = 19.51$, $SD = 1.51$), with between 13 and 21 years of education ($M = 15.55$, $SD = 1.58$). All participants had normal or corrected-to-normal vision and learned English before the age of seven. Participants were randomly assigned to be in the pure draw ($n = 24$) or pure write ($n = 23$) conditions.

Materials

All materials were the same as those in Experiment 6.

Procedure

The procedure was identical to that in Experiment 6, including the intentional encoding, except that participants were randomly assigned to one of two conditions, instructing them to either draw all words (pure draw) or write all words (pure write) presented during study. Accordingly, there was no need for colour coding of stimuli, and all study words were presented in white font on a black background. Participants still studied 66 total words, but rather than 33 falling under each encoding trial type, all 66 were either drawn or written.

Results and discussion

Data were analysed using an independent-samples *t* test, with encoding trial type (draw, write) as a between-participants variable. Analyses showed a significant main effect of encoding trial type, $t(45) = 3.63$, $SE = 0.02$, $p < .005$, $d = 1.08$, such that those in the pure draw condition recalled significantly more words than those in the pure write condition (Figure 5). It is worth noting, however, that the effect is actually likely to be larger than indicated, as one participant in the pure write condition recalled a striking 33 words, which is over 3.5 standard deviations above the mean. When analysed without this outlier, the effect was substantially larger, $t(44) = 5.54$, $SE = .02$, $p < .001$, $d = 1.63$. Overall, these results suggest that a distinctiveness account is not sufficient to explain the benefit that drawing affords recall. Other effects (e.g., production, generation) depend on a relative boost at the expense of the less distinct words, resulting in a lack of any memorial effect when encoding strategy is presented in pure lists, between participants (Mulligan, 2002). Our data indicate that even when compared between subjects, drawing allows for a significant boost in memory performance.

It should be noted that by inspection of the means, we see that the proportion of written words that were recalled is substantially lower in mixed list presentation than in pure list presentation; however, the increase in proportion of drawn words recalled in mixed lists relative to pure lists was only marginal. An analysis designed to contrast within- and between-participant comparisons of the same conditions was performed, as outlined by Erlebacher (1977). This analysis is designed to explore the interaction of design type with a dependent variable of interest.[1] Results indicated that there was a significant Design Type × Encoding Trial Type interaction, $F(1, 44) = 19.63$, $p < .001$, $\eta^2 = .09$. Such a result indicates that the effect of drawing is slightly augmented in mixed lists, but this is due to lower recall of written words, rather than a massive boost in recall of drawn words. Because of well-documented issues with output interference in free-recall paradigms (Roediger, 1974), it is also unclear whether the poorer memory for written

[1]The analysis detailed in Erlebacher (1977) requires that sample sizes in all three groups be equal. Our within-subjects design had 28 participants, and the between-subjects groups had 23 and 24 participants, respectively. In order to complete the analysis, 23 participants were randomly selected from the larger groups, and the analysis was conducted on this truncated data set.

words is due to slight distinctiveness advantages of drawn words, or due to the fact that output interference, especially with long lists, prevents output of written words later in the allotted recall time. Our results may be comparable to the findings of Bodner et al. (2014) in their between-participants production effect experiments. Similar to our work, these researchers found that memory for words encoded aloud was not substantially improved in a within-participants relative to a between-participants design, as would be expected given the distinctiveness account (MacLeod et al., 2010), indicating that a distinctiveness account might not be applicable. It is worth noting, though, that these researchers used a recognition paradigm, while we used free recall. In previous explorations of the benefits of production on free recall, there is no evidence to support a between-participants production effect (Jones & Pyc, 2014; Jonker at al., 2014). Thus, it is clear that the drawing effect cannot be fully explained by a relative distinctiveness account and can be differentiated from other item-specific encoding strategies thought to rely on distinctiveness (McDaniel & Bugg, 2008). In order to comprehensively explore the applicability of a distinctiveness account, and because of the discussed issues of output interference, future work should test the pure- versus mixed-list paradigms using a recognition task. Most importantly, though, there remains a 9% benefit of drawing in our pure-lists design, such that 23% of drawn words, but only 14% of written words, were later recalled. This shows that even if distinctiveness can account for some of the advantage provided by drawing, it is far from an exhaustive mechanism.

GENERAL DISCUSSION

Our results showed unequivocally that drawing pictures of words presented during an incidental study phase provides a measurable boost to later memory performance relative to simply writing out the words, once or repeatedly. Across all seven experiments, memory for drawn words was superior to all other instructional manipulations during encoding (see Table 1 for effect sizes). Specifically, drawing led to better later memory performance than adding detail to written words, listing physical characteristics of words, creating mental images of words, and viewing pictures of the words. We also demonstrated that the drawing effect was a feasible mnemonic strategy, even at much shorter (4 relative to 40 s) encoding durations, with much longer (66 relative to 30 words) lists, and that the benefit

Table 1. *Summary of results from Experiments 1–7*

Experiment	*Description*	*Results*	*Cohen's* d *(draw vs. write)*
1A	Studied items are drawn (draw) in detail or written (write) repeatedly; later free recall	Draw > write	1.51
1B	Same as 1A, except drawings are repeated, and writing is in detail	Draw > write	0.96
2	Same as 1A, except participants were tested in large groups in lecture halls	Draw > write	1.30
3	Drawing compared to a deep LoP condition (list), in addition to writing	Draw > list = write	0.86
4	Drawing compared to a visual imagery condition (visualize), in addition to writing	Draw > visualize[a] = write	0.87
5	Drawing compared to a picture superiority condition (view), in addition to writing	Draw > view = write	0.67
6	Same as 1A, but with shortened encoding time, and longer list of items	Draw > write	2.27
7	Same as 6, except between subjects; one group only drew items, other group only wrote	Draw > write	1.63

Note: LoP = level of processing.
[a]The difference between draw and visualize was marginal.

afforded by drawing items was robust in a pure-list, between-participants design. The drawing effect was stable despite both subtle (emphasis on detail) and drastic (addition of a third encoding trial type) changes to our paradigm. Our results show that drawing should be considered among the ranks of production (MacLeod et al., 2010), generation (Slamecka & Graf, 1978), and enactment (Engelkamp & Zimmer, 1997) as a robust encoding manipulation that can, and does, improve memory performance dramatically.

Experiments 1 and 2 showed that drawing is far superior to simply writing out a word, which has been documented as an effective iteration of the production effect (Forrin et al., 2012). We next turned our attention to the exploration of a number of potential mechanisms by which drawing could exert its effect. We reasoned that the act of creating a drawing based on a presented word would invoke (a) elaboration, (b) visual imagery, (c) motor action, and (d) a picture memory. Accordingly, it was prudent to rule out each of these individually as mechanisms driving the drawing effect.

Components of the drawing process

In Experiment 3, participants were asked to list physical characteristics of the studied words, as an elaborative, deep LoP instruction manipulation during encoding (Craik & Lockhart, 1972). In this, as well as in Experiment 4, results showed that drawing resulted in enhanced memory, and was even was superior to visual imagery, which probably contributes to the picture superiority effect (Paivio, 2014). Such results suggest that the act of drawing a word during study engages a participant more than does a simple visual imagery encoding orientation, and accordingly drawing produces a more powerful memorial benefit. In Experiment 5, participants' memory for drawn items was better than that for words presented with their associated pictures during study; this suggests that the benefit of drawing was beyond that provided by the picture superiority effect alone (Paivio et al., 1968). This finding is especially compelling given the recent finding that participants' discrimination of 2500 studied images was extraordinary (almost 90%; Brady et al., 2008). As participants were asked to choose between image pairs with as little as small perspective changes to differentiate them, these results suggest that people have a massive capacity for remembering the detail in images. The "list" (Experiment 3), "visualize" (Experiment 4), and "view" (Experiment 5) encoding orientations led to roughly similar recall performance as did the writing encoding orientation in our various experiments. Such a finding suggests that the differences we have observed, in memory for words drawn compared to written at encoding, were not due to a cost in memory from writing, but rather due to a boost in memory for the drawn words. Experiments 3 through 5 also served the added purpose of ruling out these three individual components of the drawing effect (elaboration, visual imagery, picture memory), as explanations for the effect of drawing on memory. Our results indicate that not only do LoP, imagery, and picture superiority fail to fully account for our data, but that the drawing effect is significantly more effective than these reliable alternative encoding strategies.

Distinctiveness accounts

The question remained: What is driving the "drawing effect"? Most effects (e.g., production, generation) from item-specific encoding strategies tend to be conspicuously absent in between-participants designs, leading to a compelling distinctiveness account to explain them. In other words, when produced and silently read words are compared in a mixed list, the produced words are more salient than read words, leading to better recall of produced words (MacLeod et al., 2010). This distinctiveness results in a reliable production effect in within-participant designs, but often no effect in between-participants designs. Accordingly, we tested the drawing effect in pure lists to directly explore the notion of a distinctiveness account. We found a substantial (9%) boost in recall (23% of drawn, relative to 14% of written words were later recalled), even in a between-participants pure list paradigm, suggesting that distinctiveness fails to account for the observed benefit, though we acknowledge it may contribute, since the effect was reduced in our pure compared to mixed design. It is important to note that some studies have shown that reliable, albeit smaller, production effects can be observed with recognition scores compared in between-participants designs (Bodner et al., 2014), and meta-analyses of between-participants production effect experiments do show a reliable benefit (Bodner et al., 2014; Fawcett, 2013). This slightly weakens the case for distinctiveness as an explanation for production. In comparison with the current work, however, it should be noted that the observed between-participants production effects are only roughly 4% differences, while the between-participants drawing effect showed a much larger 9% benefit, indicating that distinctiveness is probably not a mechanism that could fully explain our foregoing findings.

When looking at the recall proportions from our between-participants data (Experiment 7), compared with the same conditions within participants (Experiment 6), it is clear that the effect is smaller, due in part to an increase in proportion of drawn items recalled in mixed relative to pure lists, but mainly driven by a decrease in written items output in mixed relative to pure lists. Though this appears to point toward a small role of distinctiveness in driving our effect, it is difficult to interpret this reduction. It could be due to distinctiveness, in which case this account can partially, but far from fully, explain our data. Alternatively, written items could be suppressed due to output interference on recall later in the allotted retrieval time. This type of output interference has been well documented in the literature (e.g., Roediger, 1974). If this is the reason for the suppression, the implication is an even smaller or nonexistent role of distinctiveness, in favour of an alternative account, peripherally related to strength accounts (Ratcliff, Clark, & Shiffrin, 1990).

Synergistic interaction

There is evidence to suggest that the combination of various encoding strategies can provide a benefit that together is better than the sum of each individually. For example, Fawcett, Quinlan, and Taylor (2012) showed that the production effect was larger when naming pictures than words. Similarly, studies have shown that the well-replicated generation effect is larger when the generated word is read aloud, in a form of the production effect (MacLeod et al., 2010, 2012), suggesting that integrating two of these effects can lead to additive performance. Research has also shown a generation effect for pictures (Kinjo & Snodgrass, 2000), indicating that even though memory for images is clearly very detailed, there are still encoding manipulations that can further drive performance upward. We propose that drawing, through the seamless integration of its constituent parts, produces a synergistic effect, whereby the whole benefit is greater than the sum of the benefit of each component. To reiterate, we reasoned that drawing relies on the integration of (a) elaboration, captured by the benefit of the "list" trial type in Experiment 3, (b) visual imagery, captured by the benefit of the "visualize" trial type in Experiment 4, (c) motor action, controlled for inherently in the "write" trial type, and (d) creation of a pictorial representation, captured by the benefit of the "view" trial type in Experiment 5. The mechanism driving our drawing effect may be one that integrates these traces into a more cohesive unit.

The benefit of list (.01), visualize (.08), and view (.05) conditions (computed as proportion recall in these trial types minus proportion recall in their respective write conditions), added on top of the average proportion of written recall across all experiments (.310) falls short of the performance levels achieved, on average, from the draw encoding trial types (.48). This unexplained difference, however small, provides some very preliminary support for a synergistic effect of each of the constituent components of drawing that leads to greater memory than from the sum of the individual contributions. The consistent superiority of drawing relative to other encoding strategies throughout the experiments in the current work is also consistent with this account (see Table 1). Specifically, we believe that because drawing results in more interconnected memory cues to draw upon at recall, the memory trace for drawn words is much more likely to be effectively retrieved than when it was simply written, listed, visualized, or viewed at encoding.

The evidence cited above in favour of our synergistic account is indirect and certainly not conclusive. Accordingly, before this account can be invoked to explain the drawing effect, further experimental work is required. Future research could examine this proposed synergistic account more directly in one of two ways. First, memory for drawn versus written items could be compared using functional magnetic resonance imaging (fMRI), and functional connectivity analyses could help determine whether drawn items activate more diverse brain regions associated with visual imagery, motor cortex, and prefrontal areas associated with greater depth of processing. Functional connectivity analyses would allow the exploration of whether there is greater connectivity between the aforementioned regions and the hippocampus (known to be active for rich, detailed retrieval; Wiltgen et al., 2010) when drawn relative to written items are retrieved. Second, one could address the feasibility of the synergistic account more directly in a carefully controlled behavioural paradigm. This could be achieved by systematically varying the presence or absence of various proposed components of the drawing process (e.g., a condition where participants are unable to see their drawing, or where they must trace an existing image). Thus, direct comparisons of conditions in which the only difference is the presence or absence of one of the components would give an estimate of the contribution to recall that that component alone provides. Doing similar subtractions for each component (e.g., isolating the elaborative component by subtracting recall in a "trace the picture" condition, from recall in a "draw the picture" condition) would provide some indication of roughly how much of a benefit each component provides to memory on its own. If the total benefit of drawing surpassed the sum of these components, this would provide a compelling case for our synergistic account of drawing.

In our Experiments 1–5, participants were provided with one minute for free recall. Previous work exploring cumulative recall functions have shown that relational tasks reach asymptote faster than item-specific tasks, which reach asymptote later, when more time is given for recall (Burns & Hebert, 2005; Burns & Schoff, 1998). While all of the tasks used in the current work were item-specific, it is still a possibility that if provided with more time for recall, the number of recalled written words would approach the number of recalled drawn words. Given the large effect size, and that the effect was even larger in Experiment 6, when more time (2 minutes) was provided for recall, it seems unlikely that this would be the case. Given this concern, it is important for future work to explore a number of methodological variations to rule out the influence of the amount of time allotted for recall, as well as any potential effects of output interference. To address the former issue, further explorations of the drawing effect could allow 10 minutes for free recall, to determine whether the effect might decrease in magnitude as a result of slower recall of written words. A possible iteration to rule out the latter output interference issue (Roediger, 1974) could instruct participants to recall exclusively written words first, followed by drawn words.

While we did show that the drawing effect is reliable in group testing in our experiments, the content was still only single words and hardly representative of an academic setting. As mentioned, there is a body of work that has examined the efficacy of drawing for retaining information in an academic setting, though manipulations were often complex and frequently involved confounds or additional tools beyond basic drawing (e.g., Schwamborn et al., 2010; Van Meter, 2001; Van Meter & Garner, 2005). With this, follow-up work from the present study could address whether the advantage of drawn relative to written words will apply to memory for more complex conceptual concepts across a cross-section of educational domains, as has been suggested by previous work discussing the influence of drawing on science learning (Ainsworth, Prain, & Tytler, 2011). Future work could explore drawing as a strategy for learning information from video lectures, presented to groups of participants in large lecture halls. Such research would highlight whether drawing is a viable technique for the improvement of recall of more representative scholarly material. In educational settings, students often doodle, which is essentially drawing, but of information unrelated to the presented material. Andrade (2009) asked participants to listen to phone messages and assigned half of the participants to shade in shapes while listening to the message. The "doodling" group outperformed the other group on a later memory test. The authors suggest that this could be due to a reduction in mind wandering. Future work could explore whether drawing or doodling leads to better later performance by reducing the mind-wandering rates of participants, thereby improving later retention.

Conclusion

In the first systematic exploration of the efficacy of drawing as an encoding strategy, we showed a large and reliable advantage in memory performance for items that were previously drawn relative to those that were written, which we label the "drawing effect". Drawing a to-be-remembered stimulus was superior to writing it out (Experiment 1), regardless of whether the encoding orientation instructed the participant to emphasize either detail or repetition. Further, this effect was replicated in an ecologically valid large group setting (Experiment 2). Next, we determined that the drawing effect was superior to a deep level of processing (Experiment 3), visual imagery (Experiment 4), and viewing pictures (Experiment 5) as encoding orientations, thereby ruling these out as possible alternative explanations for the benefit of drawing on subsequent memory. Experiment 6 showed that drawing effects were still observable with long lists and short encoding durations, while Experiment 7 effectively ruled out a distinctiveness account for the drawing effect. We argue that the mechanism driving the effect is that engaging in drawing promotes the seamless integration of many types of memory codes (elaboration, visual imagery, motor action, and picture memory) into one cohesive memory trace, and it is this that facilitates later retrieval of the studied words.

Supplemental material

Supplemental content is available via the "Supplemental" tab on the article's online page (http://dx.doi.org/10.1080/17470218.2015.1094494).

REFERENCES

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

Andrade, J. (2009). What does doodling do? *Applied Cognitive Psychology*, *24*, 100–106.

Balota, D. A., & Black, S. (1997). Semantic satiation in healthy young and older adults. *Memory & Cognition*, *25*, 190–202.

Bertsch, S., Pesta, J. P., Wiscott, R., McDaniel, A. M. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*, 201–210.

Bodner, G. E., & Taikh, A. (2012). Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1711–1719.

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review*, *21*(1), 149–154.

Boldini, A., Russo, R., Punia, S., & Avons, S. E. (2007). Reversing the picture superiority effect: A speed-accuracy trade-off study of recognition memory. *Memory and Cognition*, *35*(1), 113–123.

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

Burns, D. J. & Hebert, T. (2005). Using cumulative-recall curves to assess the extent of relational and item-specific processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 189–199.

Burns, D. J., & Schoff, K. (1998). Slow and steady often ties the race: Effects of item-specific and relational processing on cumulative recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1041–1051.

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*(3), 341–361.

Cooper, E. H., & Pantle, A. J. (1967). The total-time hypothesis in verbal learning. *Psychological Bulletin*, *68*(4), 221–234.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Elliott, L. (1973). Imagery versus repetition encoding in short-and long-term memory. *Journal of Experimental Psychology*, *100*(2), 270–276.

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 671–682.

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

Erlebacher, A. (1977). Design and analysis of experiments contrasting the within-and between-subjects manipulation of the independent variable. *Psychological Bulletin*, *84*(2), 212–219.

Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica*, *142*(1), 1–5.

Fawcett, J. M., Quinlan, C. K., & Taylor, T. L. (2012). Interplay of the production and picture superiority effects: A signal detection analysis. *Memory*, *20*(7), 655–666.

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory and Cognition*, *40*(7), 1046–1055.

Guttentag, R. E., & Hunt, R. R. (1988). Adult age differences in memory for imagined and performed actions. *Journals of Gerontology*, *43*(4), P107–P108.

Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory and Cognition*, *36*(7), 1351–1359.

Jones, A. C., & Pyc, M. A. (2014). The production effect: Costs and benefits in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 300–305.

Jonker, T. R., Levene, M., & MacLeod, C. M. (2014). Testing the item-order account of design effects using the production effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 441–448.

Kieras, D. (1978). Beyond pictures and words: Alternative information-processing models for imagery effect in verbal memory. *Psychological Bulletin*, *85*(3), 532–554.

Kinjo, H., & Snodgrass, J. G. (2000). Is there a picture superiority effect in perceptual implicit tasks? *European Journal of Cognitive Psychology*, *12*(2), 145–164.

Lesgold, A. M., De Good, H. & Levin, J. R. (1977). Pictures and young children's prose learning: A supplementary report. *Journal of Reading Behavior*, *9*(4), 353–360.

Lupiani, D. A. (1977). The facilitative effects of imagery instructions and stimulus characteristics on immediate and long term free recall and clustering. *The Journal of General Psychology*, *97*(1), 73–87.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozbuko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 671–685.

MacLeod, C. M., Pottruff, M. M., Forrin, N. D., & Masson, M. E. (2012). The next generation: The value of reminding. *Memory & cognition*, *40*(5), 693–702.

Maisto, A. A., & Queen, D. E. (1992). Memory for pictorial information and the picture superiority effect. *Educational Gerontology*, *18*, 213–223.

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

Mazzoni, D., & R. Dannenberg. 2000. *Audacity*. Retrieved from http://audacity.sourceforge.net/

McDaniel, M., & Bugg, J. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, *15*(2), 237–255.

McMahon, J. A. (2002). An explanation for normal and anomalous drawing ability and some implications for research on perception and imagery. *Visual Arts Research*, *28*(1), 38–52.

Mintzer, M. Z., & Snodgrass, J. G. (1999). The picture superiority effect: Support for the distinctiveness model. *American Journal of Psychology*, *112*(1), 113–146.

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.

Mulligan, N. (2002). The emergence of item-specific encoding effects in between-subjects designs:

Perceptual interference and multiple recall tests. *Psychonomic Bulletin & Review*, *9*(2), 375–382.

Nairne, J. S., Riegler, G. L., & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 702–709.

Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 15–43.

Paivio, A (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(3), 255–287.

Paivio, A. (2014). Intelligence, dual coding theory, and the brain. *Intelligence*, *47*, 141–158.

Paivio, A. & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*, 176–206.

Paivio, A., & Foth, D. (1970). Imaginal and verbal mediators and noun concreteness in paired-associate learning: The elusive interaction. *Journal of Verbal Learning and Verbal Behavior*, *9*(4), 384–390.

Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, *11*(4), 137–138.

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

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 163–178.

Roediger, H. L. (1974). Inhibiting effects of recall. *Memory & Cognition*, *2*, 261–269.

Rowe, E. J. (1972). Discrimination learning of pictures and words: A replication of picture superiority. *Journal of Experimental Child Psychology*, *14*(2), 323–328.

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), 872–879.

Slamecka, N., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604.

Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language*, *26*(6), 589–607.

Snodgrass, J. G., & McClure, P. (1975). Storage and retrieval properties of dual codes for pictures and words in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, *1*, 521–529.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, *6*(2), 174–215.

Vaidya, C. J., & Gabrieli, J. D. E. (2000). Picture superiority in conceptual memory: Dissociative effects of encoding and retrieval tasks. *Memory and Cognition*, *28*(7), 1165–1172.

Van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology*, *93*(1), 129–140.

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, *17*(4), 285–325.

Walsh, D. A., & Jenkins, J. J. (1973). Effects of orienting tasks on free recall in incidental learning: "Difficulty," "effort," and "process" explanations. *Journal of Verbal Learning and Verbal Behavior*, *12*(5), 481–488.

Weldon, M. S., & Roediger, H. L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition*, *15*(4), 269–280.

Wiltgen, B. J., Zhou, M., Cai, Y., Balaji, J., Karlsson, M. G., Parivash, S. N., … Silva, A. J. (2010). The hippocampus plays a selective role in the retrieval of detailed contextual memories. *Current Biology*, *20*(15), 1336–1344.

Winnick, W. A., & Brody, N. (1984). Auditory and visual imagery in free recall. *The Journal of psychology*, *118*(1), 17–29.

APPENDIX A

List of target items used in Experiments 1 through 7

airplane	couch	kite	ruler
ant	cow	knife	sailboat
axe	desk	ladder	scissors
balloon	doll	lamp	screwdriver
banana	door	lemon	sheep
bee	drum	lion	shoe
beetle	duck	lips	skirt
blouse	ear	monkey	spider
boot	elephant	mushroom	spoon
broom	flute	owl	stool
butterfly	fork	pants	stove
camel	frog	peanut	strawberry
cannon	giraffe	pear	sweater
carrot	glove	penguin	toaster
cat	grapes	pepper	trumpet
caterpillar	guitar	pig	turtle
cherry	hammer	pineapple	violin
clock	harp	pumpkin	wagon
coat	jacket	rabbit	whistle
corn	kettle	rooster	wrench

APPENDIX B

Samples from participants