Drawing Enhances Item Information but Undermines Sequence Information in Memory

Tanya R. Jonker
University of Waterloo and University of California, Davis

Jeffrey D. Wammes and Colin M. MacLeod
University of Waterloo

Drawing a picture of the referent of a word produces considerably better recall and recognition of that word than does a baseline condition, such as repeatedly writing the word, a phenomenon referred to as the drawing effect. Although the drawing effect has been the focus of much recent research, it is not yet clear what underlies the beneficial effects of drawing to memory. In 3 experiments, we explored the roles of item and order information following drawing versus silent reading and produced 2 important findings. First, the drawing effect in recall was substantially larger when the 2 conditions were intermixed in a single list compared to appearing in separate lists—in other words, drawing produced a design effect. Second, the studied order was better retained for silent reading compared to drawing in pure lists. These findings are consistent with the item-order account: Memory for the order of drawn lists is poorer because the elaborative act of drawing disrupts the encoding of interitem associations (i.e., relational information) between all list items. Thus, the selection of an encoding task should be informed by how that memory might be used in the future; if one wants to remember individual items, then an elaborative task such as drawing is recommended; if one wants to remember their sequence, it is likely better to read silently.

Keywords: item-order account, design effect, drawing, generation, interitem associations

The way information is processed when first encountered determines, at least in part, what is remembered. Indeed, there are numerous examples demonstrating that the characteristics of the task performed during encoding can have a marked influence on later recall of that information. In many cases, performing an active task, or put more plainly, "learning by doing," can serve to enhance one's later memory for studied information. For example, acting out a sentence improves long-term memory for the content compared to simply reading it to oneself or watching an experimenter act it out, a phenomenon referred to as the *enactment effect* (Cohen, 1989; Engelkamp & Dehn, 2000; Engelkamp & Krumnacker, 1980; Engelkamp & Zimmer, 1989). Similarly, reading aloud produces better retention than does reading silently, referred to as the *production effect* (MacLeod & Bodner, 2017; MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). It has been proposed that encoding tasks, such as enactment and production, improve

later recall due to the enhancement of item-specific information (McDaniel & Bugg, 2008), perhaps through the encoding of additional motor, perceptual, or semantic features.

An interesting characteristic of encoding tasks such as enactment and production is that their memory benefit is often considerably greater in magnitude when they are implemented in a *mixed list;* that is, when enactment or production occurs alongside another, more impoverished encoding task (e.g., reading silently). For example, Jonker, Levene, and MacLeod (2014) had participants encode lists of items by reading each item aloud (pure list), reading each item silently (pure list), or reading some items aloud and other items silently (mixed list). On a subsequent free-recall test, there was a much larger production effect (aloud > silent) within mixed lists than between pure lists.

The observation of a greater difference for mixed lists compared to pure lists is referred to as a *design effect* (McDaniel, Waddill, & Einstein, 1988; Serra & Nairne, 1993) and is commonly found in studies of the production effect (Hopkins & Edwards, 1972; Jones & Pyc, 2014; MacLeod et al., 2010), the generation effect (Slamecka & Graf, 1978), the testing effect (Karpicke & Roediger, 2008; Mulligan, Susser, & Smith, 2016), and many others (for a review, see McDaniel & Bugg, 2008). Design effects are interesting phenomena because they suggest that different cognitive processes are occurring in mixed versus pure encoding.

To link together many different encoding tasks, McDaniel and Bugg (2008) put forth a unifying framework of design effects in which they emphasized two potential cognitive mechanisms that might occur to different degrees during encoding. According to this item-order account, elaborative encoding requires more interpretation, which results in enhanced item-specific encoding but simultaneously disrupts the encoding of interitem associations

between sequential list items. Accordingly, lists that include an elaborative task (a) increase item strength for the elaborately encoded items and (b) decrease the encoding of relational associations between items in that list. It is important to note that this decrease in relational encoding will occur for both mixed lists and pure elaborative lists because the presence of elaborative processing in these lists promotes attention-capturing item-specific encoding. Consequently, the only list that benefits from the encoding of relatively strong interitem associations will be a pure list involving a common, nonelaborative encoding task (e.g., reading silently).

Both of these encoded sources of information—item information and relational information—can be used to guide and benefit retrieval (Hunt & Einstein, 1981). In the case of mixed lists, where a larger effect of elaborative encoding is typically observed, retrieval of the items encoded elaboratively benefits from enhanced item information, but retrieval of the items encoded through the common task (e.g., silent reading) is poor because relational information was disrupted by the presence of elaborative processing. Thus, the difference in memory between the elaborative task and the common task is maximal because the common task does not benefit from either source of memory-strengthening information (i.e., item or relational information). This is not the case, however, for pure lists. For pure lists, the elaborative task promotes item information and the common task promotes relational information, resulting in more balanced recall and an attenuated difference in memory as a function of the encoding task. This account does not hinge on any specific encoding task, and as such, design effects should occur across a number of elaborative encoding tasks, as has in fact been demonstrated.

The Drawing Effect

The drawing effect is a recent addition to the growing family of elaborative encoding tasks. It has now been shown many times that drawing a picture of the referent of a word produces considerably better recall and recognition of that word compared to repeated writing of the word (Wammes, Meade, & Fernandes, 2016; Wammes, Meade, & Fernandes, 2017a, 2017b; Wammes, Jonker, & Fernandes, 2018; see Paivio & Csapo, 1973; for a discussion of drawing in science education, see Ainsworth, Prain, & Tytler, 2011). Of interest, a recent study found that drawing the referent of a word resulted in a larger free-recall benefit for mixed lists than for pure lists, suggesting that the drawing effect might be another instance of a design effect. Wammes et al. (2016) instructed participants to perform one of three tasks while encoding a list of 66 words: (a) make a simple line drawing of the referent of the word that they were studying (pure draw list), (b) repeatedly write the word for an equal encoding duration (pure write list), or (c) do a mix of drawing and writing in response to prompts that preceded each word (mixed list; see Experiments 6 and 7 and Figure 5 of Wammes et al., 2016). The authors found a drawing effect—defined as the memory advantage for drawn words compared to written words—for both pure and mixed lists. It is important to note, however, that the magnitude of the drawing effect was greater in the mixed group than in the pure groups. This result suggests that the drawing effect might also be influenced by the differential use of relational versus item information, as has been found for many other elaborative encoding tasks.

In light of this literature, the goals in the present research were twofold. First, we aimed to establish whether the drawing effect is in fact a design effect. The previously mentioned finding of Wammes et al.'s (2016) was not the result of an intentional exploration of drawing as a design effect; instead, the finding was a byproduct of other goals. Some of the conditions of their experiments limit conclusions, the most important being that the mixed group and the two pure groups were run as two separate experiments, collected at different times of the year. Thus, here, we aimed to carry out a set of experiments that would directly provide several tests of the drawing effect as a design effect. To achieve this goal, we conducted two experiments using stimuli and methods that differed from those of Wammes et al.'s. Specifically, we measured free-recall accuracy following drawing versus reading silently in mixed and pure lists in a within-subjects design. We chose to eschew the definitional baseline condition of writing in favor of silent reading—a more typical baseline in the item-order literature—to render the effect more comparable to that of other encoding tasks whose benefits can be classified as design effects (e.g., perceptual interference: Burns, 1990; Mulligan, 1999, 2000; generation effect: Burns, 1990; production effect: Jonker et al., 2014; testing effect: Roediger & Karpicke, 2006).[1]

Our second goal was to explore the role of memory for sequence in the drawing effect to determine whether a potential design effect would be driven by the differential encoding of relational information. To achieve this goal, we analyzed the sequence information found within the recall data from our first two experiments and we directly tested memory for order in a final experiment using an order reconstruction test. By examining the potential role of relational encoding in the drawing effect, we sought to align the drawing effect with a theoretical framework already known to explain several encoding effects. Insight into the mechanisms that drive the drawing effect can then guide future explorations of the phenomenon. Furthermore, the present experiments have the potential to add drawing to the list of design effects that can be explained by the item-order account. This large and growing set of phenomena envelopes a diverse set of tasks (e.g., enactment, generation, production, bizarreness, perceptual interference), and yet all the tasks function in highly similar ways, suggesting that they might be driven by common cognitive mechanisms. By understanding these mechanisms, the field can move toward making predictions about the tasks best suited for memory enhancement, given a particular context. For example, in a business meeting, a manager might learn a series of disjointed facts about the health of the company, or over coffee, a mentor might share a list of several events that led to her getting a new job. In the former case, it is likely most advantageous to strengthen item information for each of the facts because sequence is not very important. In the latter case on the other hand, it might be most advantageous to strengthen memory for the sequence, to understand the order in which the events unfolded to lead to the mentor's success. Knowledge of the types of encoding that enhance item versus relational

---

[1] Furthermore, our recent work has demonstrated that active tasks of any kind, including a simple button press, might result in weak item and relational information (Jonker & MacLeod, 2015). Therefore, to examine the possibility that drawing impairs the encoding of relational information, we opted to include a baseline that is known to produce strong memory for order (i.e., silent reading).

information would help learners make informed decisions about how to strengthen the facets of a memory that they care most about.

Experiments 1 and 2

The first goal of Experiments 1 and 2 was to examine whether the drawing effect is indeed an instance of a design effect. To do so, we had participants encode multiple lists that were either mixed or pure in their composition. In both experiments, we also manipulated the number of items in a list to test whether the presence of the design effect is robust across a variety of list lengths. There has been some evidence to suggest that sequence information contributes less to retrieval as list length increases (Mulligan & Lozito, 2007), possibly because this information is less useful during retrieval with increasing length. Therefore, to obtain a more complete characterization of the drawing effect as a design effect, we examined mixed and pure lists that were either short or long.

Assuming that we would observe a design effect in these experiments, our second goal was to determine whether this effect was driven by differences in memory for order. To this end, we assessed memory for order with two commonly used metrics of sequence information.

Experiments 1 and 2 differed in only one respect: The long lists in Experiment 1 had 20 items, whereas the long lists in Experiment 2 had 34 items. By including a variety of list lengths, we were able to determine whether the drawing effect—and any associated sequence effect—is moderated by list length. Moreover, the similarity across these two experiments provided a valuable opportunity for an internal replication.

Method

Participants. Thirty-nine students from the University of Waterloo (32 female), with a mean age of 19.7 years, participated in Experiment 1 in exchange for partial course credit. Forty students (31 female), with a mean age of 20.7, participated in Experiment 2. All experiments reported in this and subsequent experiments were approved by an Institutional Ethics Board, and all participants reported fluency in English, normal or corrected-to-normal vision, and normal color vision.

Materials. A set of 276 common nouns with a word length of five or six letters was selected from the MRC psycholinguistic database (Coltheart, 1981), and all study lists were randomly drawn without replacement from this pool. Previous work has primarily chosen word lists for relative concreteness, to ensure that the words could in fact be drawn (e.g., Wammes et al., 2016). The group of words in the current work, however, was not constrained by any selection criteria beyond word length, to allow for a heterogeneous sample of words. This choice provided a more robust test of the drawing effect because the list contained words that may not be particularly well suited to the task. Any evidence of a drawing effect with a more diverse sample of words is a testament to the robustness of the phenomenon.

Experiment 1 involved 12 short lists of eight words and nine long lists of 20 words. Experiment 2 involved nine short lists of eight words and six long lists of 34 words. In each experiment, the three conditions (pure drawn, pure silent, and mixed) occurred equally often.

The words were presented via computer using E-Prime 2.0 software (Psychology Software Tools, 2016, Pittsburgh, PA). Participants were provided with a small writing tablet, equipped with a slider that could be used to quickly and cleanly erase their drawing, that they could use for drawing, as in previous work (Wammes et al., 2016).

Procedure. Participants completed as many study-test blocks as there were study lists (21 in Experiment 1, 15 in Experiment 2). Each block began with a study list, during which words appeared one at a time for 4 s each at the center of the screen. Participants were not informed of the list type prior to encoding. The order of the lists was fully randomized without constraint. When a word appeared in yellow, participants were to read it silently; when a word appeared in blue, participants were to create a drawing of the word's referent on their writing tablet. In the case of mixed lists, the order of the two encoding tasks was fully randomized without constraint. After each word disappeared, there was a 1-s interstimulus interval. A tone sounded at the start of this interval to signal that participants should stop drawing (if the trial involved drawing) and use the slider to erase their tablet in preparation for the next word. A research assistant was present throughout the experiment to ensure compliance.

Each study phase was followed by a 15-s distractor task, during which participants saw a series of single digits presented one at a time on the screen and were to make odd-even judgments with key presses. Finally, for each list, there was a free-recall test, during which participants were to recall the items studied in that list by writing them down. Critically, there was no mention of studied order in the instructions; instead, instructions emphasized the importance of recording words in the order in which they came to mind during recall. Upon completion of the test phase, the next study-test cycle began.

Prior to beginning the experiment, participants performed a practice round to familiarize themselves with the experimental procedure. In the practice phase, participants were exposed to all three tasks (study, distractor, test) and were given the opportunity to ask questions as needed.

Results and Discussion

Data were analyzed in R using base functions and the afex package for analysis of variance (ANOVA) models (Singmann, Bolker, Westfall, & Aust, 2016).

Free recall.

Experiment 1. A 2 × 2 × 2 repeated-measures ANOVA examined the effects of item type (draw vs. silent), list type (mixed vs. pure), and list length (short vs. long) on proportion recalled. All variables produced significant main effects: item type, $F(1, 38) = 4.82$, $MSE = .03$, $p = .03$, $\eta_p^2 = .11$; list type, $F(1, 38) = 10.93$, $MSE = .01$, $p < .01$, $\eta_p^2 = .22$; and list length, $F(1, 38) = 305.35$, $MSE = .01$, $p < .001$, $\eta_p^2 = .89$. Most important, list type interacted with item type, $F(1, 38) = 63.85$, $MSE = .01$, $p < .001$, $\eta_p^2 = .63$, demonstrating the overall design effect, which is displayed in Panel A of Figure 1. List length did not enter into any interactions (ps > .14), indicating that the design effect occurred for both long and short lists.

Although there were no significant interactions with list length in this experiment, visual inspection of the data suggested a pure-list difference between drawing and silent reading for short lists
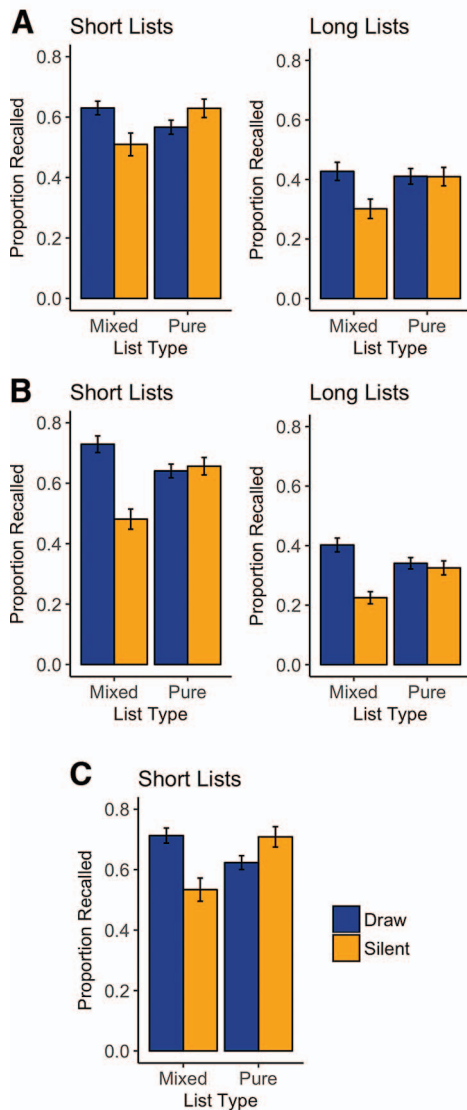
Figure 1. Proportion of items correctly recalled on all free-recall tests for each list type (Drawing, Silently Read, and mixed encoding of both Draw and Silent). Experiment 1 is displayed in Panel A, Experiment 2 in Panel B, and Experiment 3 in Panel C. Error bars represent one standard error of the condition's mean. See the online article for the color version of this figure.

but not for long lists (see Panel A of Figure 1). Paired-samples *t* tests confirmed this difference in pattern: There was no evidence for a recall difference between drawing and silent reading for long pure lists, $t(38) = .04$, $SE = .02$, $p = .97$, $d = .01$, but a notable silent reading advantage for short pure lists, $t(38) = 2.47$, $SE = .03$, $p = .02$, $d = .39$. In other words, we found a reverse drawing effect in the short lists. Indeed, a 2×2 ANOVA assessing the effects of list length (short vs. long) and item type (draw vs. silent) on free recall of pure lists revealed a significant interaction, $F(1, 38) = 5.51$, $MSE = .01$, $p = .02$, $\eta_p^2 = .13$, demonstrating that list length affected the drawing effect in pure lists. This reverse drawing effect for short pure lists was an unanticipated finding—

although similar phenomena have been observed with other elaborative encoding tasks (see Burns, 1990; Mulligan & Peterson, 2015a, 2015b). We return to this finding in a final cross-experiment analysis, so we defer discussion of it until that time.

A strong drawing effect was found for mixed lists at both list lengths: long, $t(39) = 4.64$, $SE = .03$, $p < .001$, $d = .74$; short, $t(39) = 3.67$, $SE = .03$, $p < .001$, $d = .59$.

Experiment 2. A $2 \times 2 \times 2$ repeated-measures ANOVA examined the effects of item type (draw vs. silent), list type (mixed vs. pure), and list length (short vs. long) on proportion recalled. Replicating Experiment 1, all variables produced significant main effects: item type, $F(1, 39) = 17.03$, $MSE = .05$, $p < .001$, $\eta_p^2 = .30$; list type, $F(1, 39) = 7.92$, $MSE = .01$, $p < .01$, $\eta_p^2 = .17$; and list length, $F(1, 39) = 487.84$, $MSE = .02$, $p < .001$, $\eta_p^2 = .93$. Moreover, replicating Experiment 1, list type interacted with item type, again demonstrating the overall design effect, $F(1, 39) = 78.91$, $MSE = .01$, $p < .001$, $\eta_p^2 = .67$ (see Panel B of Figure 1).

Unlike in Experiment 1, however, the three-way interaction was significant in Experiment 2, $F(1, 39) = 6.00$, $MSE = .01$, $p = .02$, $\eta_p^2 = .13$, demonstrating that the size of the design effect was dependent on list length. Assessment of this three-way interaction revealed that it was driven by a larger mixed-list drawing effect in short lists compared to long lists; 2×2 ANOVAs assessing the effects of list length and item type for mixed and pure lists separately revealed a significant interaction for mixed lists, $F(1, 39) = 4.12$, $MSE = .01$, $p = .05$, $\eta_p^2 = .10$, but not for pure lists, $F(1, 39) = 1.34$, $MSE = .01$, $p = .25$, $\eta_p^2 = .03$. Also, unlike in Experiment 1, in this experiment the short pure lists produced no reverse drawing effect. As can be seen in Panel B of Figure 1, recall was equivalent for silently read and drawn items from pure lists for both list lengths: long, $t(39) = .49$, $SE = .03$, $p = .63$, $d = .08$; short, $t(39) = .49$, $SE = .03$, $p = .62$, $d = .08$. For mixed lists, there was a drawing effect for both list lengths: long, $t(39) = 5.79$, $SE = .03$, $p < .001$, $d = .91$; short, $t(39) = 7.01$, $SE = .04$, $p < .001$, $d = 1.11$.

In summary, these two experiments produced four observations of the drawing effect as a design effect, seen in Panels A and B of Figure 1. In other words, although the drawing effect was consistently observed for mixed lists, it was never observed for pure lists, irrespective of list length. Thus, the drawing effect appears to be another clear instance of a design effect, and one that is not moderated by list length.

Memory for order. Our next analyses focused on the sequence in which participants recalled items. The item-order account postulates that elaborative encoding disrupts relational encoding. Accordingly, drawing—in both mixed and pure lists—should result in less orderly recall compared to silent reading. This account can explain why the drawing effect was absent in pure lists: Silently read lists benefited from relational encoding, which enhanced their free recall and made it comparable to that of the elaborative encoding afforded by drawing. To assess whether the lack of a pure-list difference in Experiments 1 and 2 was related to enhanced memory for sequence information for pure silent lists, we assessed temporal structuring of free-recall outputs via two different scoring methods used in previous work (Jonker et al., 2014).

Interitem associations. We first examined whether the order of one's recall output reflected interitem associations. For this analysis, each pair of sequentially recalled items was coded as

having been presented either back-to-back during study or separated during study. The order of the pair of sequentially recalled items was not considered. Referring to the example in Table 1, the only preserved interitem association was between the items that were recalled in Positions 2 and 3 of the test (*mug* and *bucket*), which occurred sequentially in the study list and were also recalled sequentially (had *mug* and *bucket* been recalled adjacently but in the reversed order, this would still have constituted a preserved interitem association). None of the other sequentially recalled pairs (*mouse*−*mug* or *bucket*−*cloud*) had occurred sequentially during the study. Using these scored pairs, we created a proportion score for each list (number of preserved associations/all possible pairs of sequentially recalled items). For the example in Table 1, this score would be .33 (1/3). We could not compute separate scores for drawn and silently read items in mixed lists, because the two item types were intermixed during study and recall; therefore, we computed one score for mixed lists, collapsing across item type.

For each experiment, a 2×3 ANOVA assessed the effects of list length (short vs. long) and list type (pure draw, mixed, and pure silent) on interitem associations. For Experiment 1, we observed main effects of list type, $F(1, 38) = 25.65$, $MSE = .04$, $p < .001$, $\eta_p^2 = .40$, and list length, $F(1, 38) = 95.54$, $MSE = .02$, $p < .001$, $\eta_p^2 = .72$, and a marginally significant interaction, $F(1, 38) = 3.07$, $MSE = .02$, $p = .05$, $\eta_p^2 = .07$. For Experiment 2, we observed main effects of list type, $F(1, 39) = 9.15$, $MSE = .03$, $p < .001$, $\eta_p^2 = .19$, and list length, $F(1, 39) = 216.94$, $MSE = .02$, $p < .001$, $\eta_p^2 = .85$, but no interaction, $F(1, 39) = .43$, $MSE = .02$, $p = .65$, $\eta_p^2 = .01$. In both experiments, pure silent reading led to a greater proportion of recalled interitem associations compared to pure drawing (see Figure 2), a result in keeping with the prediction from the item-order account. Pairwise comparisons for pure drawn and pure silent lists are reported in Table 2.

Distance. Our second metric of memory for order assessed the distance in serial position at study for all pairs of sequentially recalled items. For example, in Table 1, the successively recalled words *mouse* and *mug* were studied in serial Positions 3 and 6, resulting in a distance score of 3. The scores were absolute; forward versus backward distance was not considered in this analysis. A higher score indicates a greater average distance between successively recalled items.[2]

For Experiment 1, a 2×3 ANOVA assessed the effects of list length (short vs. long) and list type (pure draw, mixed, and pure silent) on distance scores. This analysis revealed significant main
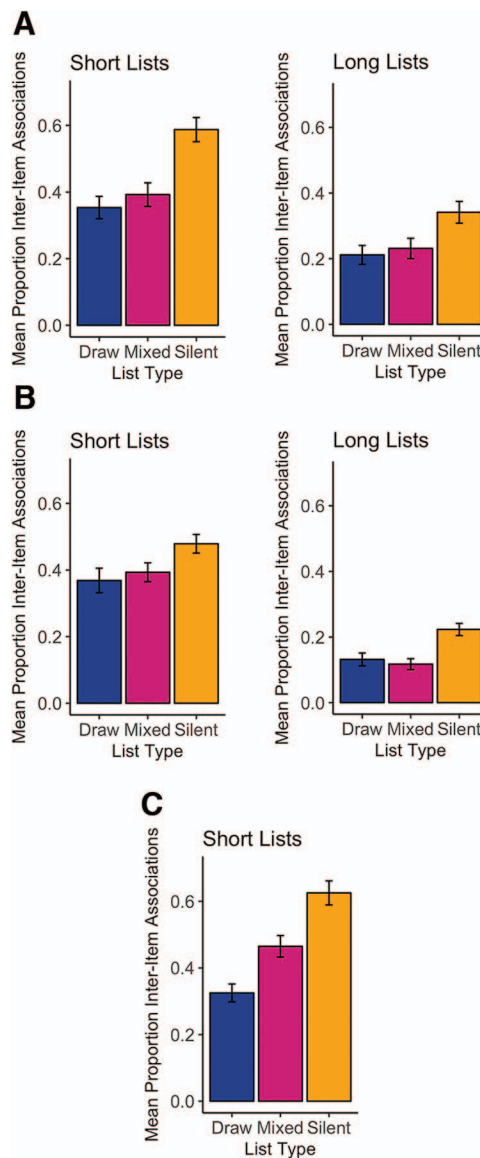

Figure 2. Proportion of pairs of recalled items that reproduced interitem associations for each list type (Drawing, Silently Read, and mixed encoding of both Draw and Silent). Experiment 1 is displayed in Panel A, Experiment 2 in Panel B, and Experiment 3 in Panel C. Error bars representation one standard error of the condition's mean. See the online article for the color version of this figure.

effects of list type, $F(1, 38) = 20.91$, $MSE = 1.16$, $p < .001$, $\eta_p^2 = .35$, and list length, $F(1, 38) = 277.92$, $MSE = 2.12$, $p < .001$, $\eta_p^2 = .88$, but, as with interitem associations, no interaction, $F(1, 38) = 1.48$, $MSE = .86$, $p = .23$, $\eta_p^2 = .04$. Similarly, for Experiment 2, we observed significant main effects of list type, $F(1, 39) = 4.67$, $MSE = 3.18$, $p = .01$, $\eta_p^2 = .11$, and list length,

Table 1

Sample Study List (Input) Sequence and Free-Recall Test (Output) Sequence

Study sequence	Recall test sequence
1. *Table*	1. *Mouse* (3)
2. *Grass*	2. *Mug* (6)
3. *Mouse*	3. *Bucket* (5)
4. *Dresser*	4. *Cloud* (7)
5. *Bucket*	
6. *Mug*	
7. *Cloud*	
8. *Soup*	

Note. Numbers in parentheses reflect the serial position in the study sequence.

[2] We did not analyze the last response if participants recalled all eight items, because the serial position of the eighth item would be determined because it was the last possible item in the list and was therefore not independent.

Table 2
Comparison of Pure Drawn Lists and Pure Silent Lists Across All Experiments

Experiment, metric, and list length	Statistic
Experiment 1	
Interitem associations	
Short	$t(38) = 5.13$, $SE = .04$, $p < .001$, $d = .82$***
Long	$t(38) = 3.67$, $SE = .04$, $p < .001$, $d = .59$***
Distance	
Short	$t(38) = 4.67$, $SE = .16$, $p < .001$, $d = .75$***
Long	$t(38) = 4.21$, $SE = .26$, $p < .001$, $d = .67$***
Experiment 2	
Interitem associations	
Short	$t(39) = 2.38$, $SE = .05$, $p = .02$, $d = .38$*
Extra long	$t(39) = 3.23$, $SE = .03$, $p < .01$, $d = .51$**
Distance	
Short	$t(39) = 2.56$, $SE = .14$, $p = .01$, $d = .40$*
Extra long	$t(39) = 1.89$, $SE = .49$, $p = .07$, $d = .30$†
Experiment 3	
Interitem associations	
Short	$t(36) = 7.92$, $SE = .04$, $p < .001$, $d = 1.30$***
Distance	
Short	$t(36) = 7.08$, $SE = .13$, $p < .001$, $d = 1.16$***

† $p < .10$.   * $p < .05$.   ** $p < .01$.   *** $p < .001$.

$F(1, 39) = 978.24$, $MSE = 3.61$, $p < .001$, $\eta_p^2 = .96$, but again, no interaction, $F(1, 39) = 1.36$, $MSE = 2.53$, $p = .26$, $\eta_p^2 = .03$. Confirming the pattern found for interitem associations in the preceding analysis and again consistent with the item-order account, our analysis of distance scores indicated superior memory for order for pure silently read lists compared to pure drawn lists (see Figure 3; the effect for extralong lists in Experiment 2 was only marginal but corresponded with all other significant analyses). Pairwise comparisons are reported in Table 2.

Experiment 3

Having demonstrated that drawing produced a robust design effect and having found compelling evidence that differences in memory for order might underlie recall differences, we explicitly tested memory for order in our third experiment. To do so, we modeled Experiment 3 after the method used by Nairne, Riegler, and Serra (1991, Experiment 2; see also Jonker et al., 2014). All lists used in this experiment were short lists, with each followed by either a free-recall test (as in Experiments 1 and 2) or an order reconstruction test in which participants were provided with the study list in a scrambled sequence and were to reconstruct the studied order. By randomly assigning the test to each block, we were also able to reduce the predictability of the test demands; that is, participants could not predict how they would be tested and therefore presumably could not adjust their encoding strategy to meet the demands of that specific testing method.

Method

Participants.   Thirty-seven students from the University of Waterloo (23 female) with an average age of 20.4 participated in exchange for partial course credit.

Materials.   The materials were the same as those in Experiments 1 and 2. Twenty-four lists of eight items were sampled from the list of common nouns; these were equally distributed across mixed, pure drawn, and pure silent lists.

Procedure.   The study and distractor tasks were identical to those used in Experiments 1 and 2. The difference was at the time of test: Each test was either a free-recall test (identical to the one described in Experiments 1 and 2) or an order reconstruction test. During the order reconstruction test, participants were presented with a vertical list of the eight study words in a scrambled order, and they were to write (on a provided sheet of paper) the words in their studied sequence. Test words were presented in black font on a white background. Test type (free recall vs. order reconstruction) was assigned such that there was an equal number of free-recall and order reconstruction tests following each list type (pure silent, pure draw, mixed), with the order of these randomized for each participant.

Results and Discussion

We assessed the presence of sequence information following drawing versus silent reading for each test independently. We expected the results to converge with the item-order literature and with our Experiments 1 and 2. That is, we expected a free-recall advantage for drawing in mixed lists and poorer memory for order following drawing in pure lists.

Free recall.   We first assessed free recall as in Experiments 1 and 2. A 2 × 2 repeated-measures ANOVA revealed a significant main effect of list type, $F(1, 36) = 4.02$, $MSE = .02$, $p = .05$, $\eta_p^2 = .10$, but not of item type, $F(1, 36) = 2.73$, $MSE = .03$, $p = .11$, $\eta_p^2 = .07$. Critically, a significant interaction between list type and item type demonstrated that we again observed a design effect, $F(1, 36) = 60.12$, $MSE = .01$, $p < .001$, $\eta_p^2 = .63$. To follow up
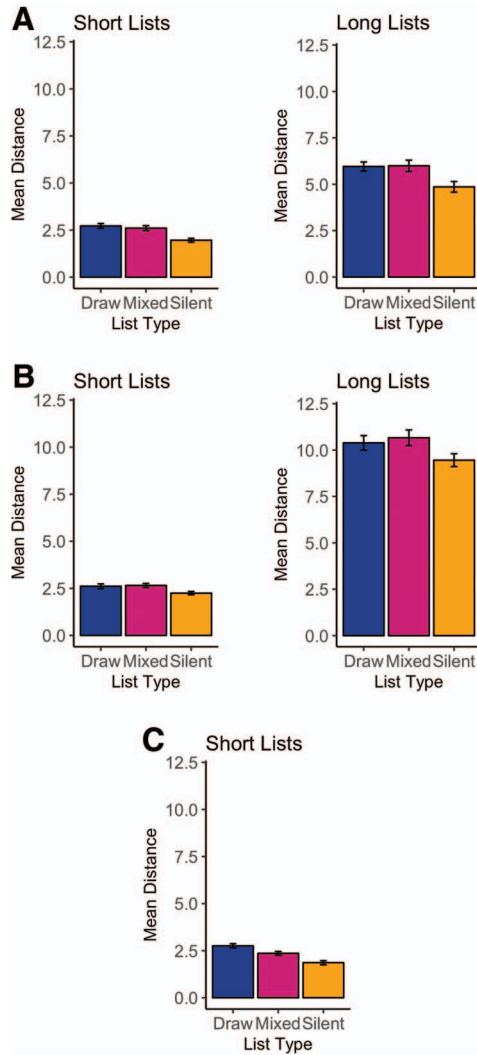
Figure 3. Mean serial position distance between all pairs of recalled items for each list type (Drawing, Silently Read, and mixed encoding of both Draw and Silent). Experiment 1 is displayed in Panel A, Experiment 2 in Panel B, and Experiment 3 in Panel C. Error bars representation one standard error of the condition's mean. See the online article for the color version of this figure.

on this interaction, we performed two paired-samples *t* tests, comparing drawing to silent reading separately for each list type. For mixed lists, participants recalled more drawn words than silent words, $t(36) = 4.74$, $SE = .04$, $p < .001$, $d = .78$. The reverse was true for pure lists: Participants recalled more silent words than drawn words from pure lists, $t(36) = 3.08$, $SE = .03$, $p < .01$, $d = .51$. In other words, as in Experiment 1 (but not in Experiment 2), we observed a reverse drawing effect for short pure lists.

Memory for order. To better understand our recall pattern, we analyzed memory for order in free-recall outputs using the same two methods as described for Experiments 1 and 2.

Interitem associations. Retention of interitem associations was better following silent reading in pure lists compared to both mixed lists, $t(36) = 4.72$, $SE = .03$, $p < .001$, $d = .78$, and pure drawn lists (see Table 2); omnibus ANOVA, $F(1.93, 69.62^3) =$

37.07, $MSE = .02$, $p < .001$, $\eta_p^2 = .51$. Furthermore, memory for interitem associations was superior following study of mixed lists compared to pure drawn lists, $t(36) = 4.29$, $SE = .03$, $p < .001$, $d = .71$.

Distance. The same pattern was observed for distance scores. Distance scores between recalled items for pure silent lists were significantly lower than were those for mixed lists, $t(36) = 4.54$, $SE = .11$, $p < .001$, $d = .75$, or pure drawn lists (see Table 2); omnibus ANOVA, $F(1.90, 68.34^4) = 26.53$, $MSE = .30$, $p < .001$, $\eta_p^2 = .42$. Furthermore, pure drawn lists resulted in larger distance scores compared to mixed lists, $t(36) = 3.02$, $SE = .13$, $p < .01$, $d = .50$. Thus, each pair of sequentially recalled items from pure silently read lists had been studied more closely together compared to recall from either pure drawn lists or mixed lists.

In summary, our free-recall results entirely replicated those of Experiments 1 and 2. We expected that the results from the order reconstruction task would converge with the recall patterns, thereby providing strong evidence for the differential use of order information favoring silent reading over drawing.

Order reconstruction. To assess performance on the order reconstruction test, we used the same strict scoring criterion that has been employed in previous work (Jonker et al., 2014; Nairne et al., 1991): Items were considered to be correct only when placed in their exact serial position; any deviation was scored as incorrect. Thus, an item studied in Position 3 would be scored as correct on the order reconstruction test only if it was written in the third space on the test. A single proportion correct score was produced for each pure list; two proportion scores were produced for each mixed list—one for silently read items and one for drawn items.

A 2×2 repeated-measures ANOVA assessing the effects of list type (mixed vs. pure) and item type (draw vs. read) on order reconstruction scores revealed no effect of list type, $F(1, 36) = .04$, $MSE = .02$, $p = .84$, $\eta_p^2 < .01$, but a significant effect of item type, $F(1, 36) = 6.32$, $MSE = .02$, $p = .02$, $\eta_p^2 = .15$. Critically, there was a significant interaction between list type and item type, $F(1, 36) = 16.23$, $MSE = .02$, $p < .001$, $\eta_p^2 = .31$, as shown in Figure 4. Two paired-samples *t* tests comparing drawing to silent reading for each list type revealed no significant difference for mixed lists, $t(36) = 1.53$, $SE = .02$, $p = .13$, $d = .25$, but a significant advantage following silent reading compared to drawing for pure lists, $t(36) = 3.95$, $SE = .03$, $p < .001$, $d = .65$. These results parallel the order results from Experiments 1 and 2 despite substantial methodological and analytical differences, thus providing strong conceptual replication of the basic design-driven pattern. Together, these analyses all converge on the same conclusion: Compared to silent reading, drawing disrupts the encoding of relational information. These findings are entirely consistent with the item-order account, and they confirm that the drawing effect fits with McDaniel and Bugg's (2008) class of design-influenced encoding techniques.

---

[3] Due to a violation of the assumption of homogeneity of variance, a Greenhouse-Geisser correction to degrees of freedom was applied.
[4] Due to a violation of the assumption of homogeneity of variance, a Greenhouse-Geisser correction to degrees of freedom was applied.
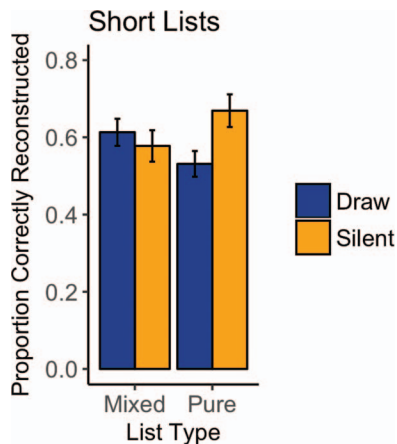
Figure 4. Order reconstruction performance for Experiment 3. See the online article for the color version of this figure.

Analysis Across Experiments

Thus far, we have provided support for the role of relational and item information in the drawing effect by examining differences across *conditions*. However, we also conducted exploratory analyses to probe for converging evidence across *experiments* and *individuals*. Specifically, we wondered whether participants who produced a stronger drawing effect would also tend to produce more orderly recall from lists that had been drawn compared to lists that had been silently read. This investigation was guided by the interesting and unanticipated observation of a reverse drawing effect in pure lists for two of the three cases involving our short eight-item lists (see Experiments 1 and 3). Neither of our long lists produced a reverse drawing effect (see Figure 1). Upon closer inspection of the data for our short lists, we noticed that—across experiments—the size of this reverse drawing effect in recall positively correlated with evidence of greater sequence memory for silently read versus drawn pure lists, as shown in Panel A of Figure 5. In other words, when pure-list recall was poorer for drawn items than for silent items at the group level, it also happened that memory for order was greater for these silently read items. Although the pattern was not as clear as in the measure of interitem associations, we observed a similar trend in distance scores (Panel C; short vs. long list lengths should be considered separately because the range of possible distance scores increases with list length); the trend was negative, as expected, given that greater distance scores index poorer memory for order. Taken together, the pattern across experiments suggests that the size of the drawing effect in recall is associated with the drawing effect in measures of relational order information.

Given that the short lists in all three experiments were encoded under identical conditions, we speculated that these differences across experiments might be due to individual differences within our samples. To examine the association between the drawing effect in recall and differences in order across drawn and silently read pure lists, we collapsed the three experiments to examine individual differences. Across participants from all experiments, the drawing effect in pure-list recall correlated reliably with pure-list differences in interitem associations, $r(193) = .33$, $p < .001$. This pattern is clearly evident in Panel B of Figure 5. In other

words, participants who produced a strong drawing effect in recall tended to produce equivalent or greater order in their recall of drawn lists compared to silently read ones. In contrast, those with a reverse drawing effect tended to use sequence information more in their output for pure silent lists than for pure drawn lists. The same pattern was observed for the distance metric: The drawing effect correlated negatively with differences in distance scores, $r(193) = -.23$, $p = .001$, as shown in Panel D,[5] indicating that better memory for order on pure drawn lists compared to pure silent lists was linked to greater recall of the drawn items.

These results converge with the conclusions already reached from our across-condition analyses, and they are consistent with the predictions of the item-order account (McDaniel & Bugg, 2008). Recall can be driven by item information and by relational information. Thus, when item information is strong for drawn lists and relational information is strong for silently read lists, then recall of the two lists will be similar and perhaps even equivalent. However, when relational information was greater for the common encoding task, then a reversed drawing effect was observed. Although these analyses are post hoc, they are a valuable demonstration of convergence with the primary analyses of Experiments 1, 2, and 3, demonstrating that the drawing effect is influenced strongly by the encoding of sequence information.

General Discussion

In three experiments, the drawing effect was substantially attenuated when drawing occurred in pure lists (see also Wammes et al., 2016) as opposed to mixed lists. In fact, across the three experiments reported here, we found no evidence for a significant benefit of drawing over silent reading in short pure lists; in some cases, we even found a reversed effect. In contrast, this was not the case for mixed lists: Every case resulted in a strong memory advantage for drawing over silent reading. Thus, the drawing effect is a clear instance of a design effect.

Our assessments of sequence memory in free-recall outputs illuminate this design effect. Specifically, we found that drawing in pure lists or in mixed lists disrupted memory for order for all items in these lists—even silently read ones in the mixed lists—when compared to silent reading in pure lists. A direct investigation of sequence memory using an order reconstruction test corroborated these findings (see Experiment 3).

These sequence findings accord with the item-order account of design effects: In the case of elaborative encoding (i.e., drawing), item information is encoded well but relational information is not, whereas in the case of common encoding (i.e., silent reading), relational information is encoded well but only when no elaborative task is present to disrupt it (i.e., in pure lists but not in mixed lists). Accordingly, drawing should result in strong item information irrespective of list type, whereas silent reading should result in strong relational information in only pure silent reading lists. Table 3 outlines all the predictions derived from McDaniel and Bugg's (2008) item-order account and shows that each of them was upheld by our data.

---

[5] This analysis revealed a clear outlier, visible in Panel D of Figure 5. We reran the analysis excluding this outlier participant and the same pattern was observed, $r(193) = -.17$, $p = .02$.
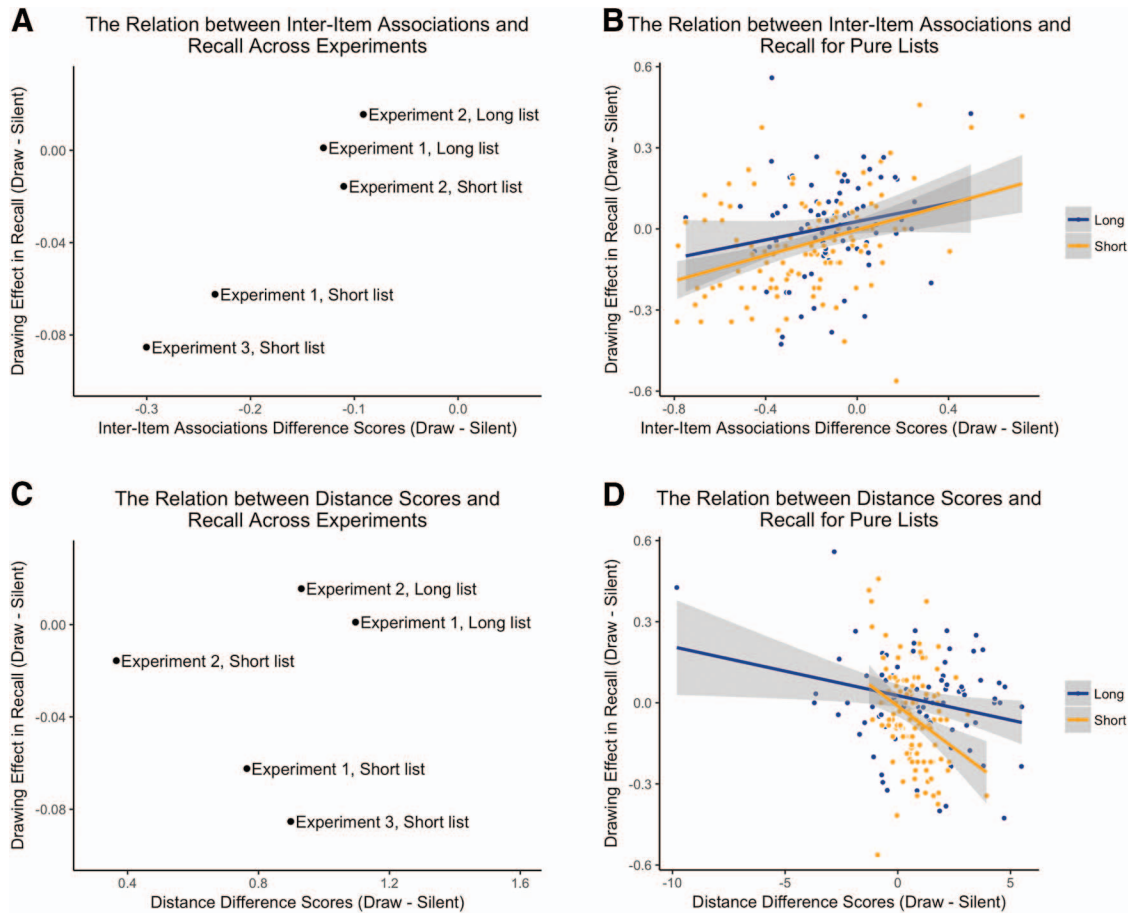
Figure 5. Relation between the drawing effect in recall and differences in memory for order. Panels A and C display the mean differences across experiments, whereas Panels B and D plot each individual participant, collapsed across experiments. The gray areas in Panels B and D indicate a 95% confidence interval for the prediction from the fitted linear model. See the online article for the color version of this figure.

To provide evidence converging with our primary findings, we ran an additional analysis examining individual differences. We found that the size of the drawing effect for pure lists was strongly related to the use of sequence memory during recall: When individuals exhibited greater memory for order following silent read-

ing compared to drawing, the drawing effect in recall was attenuated or even reversed. This result links the size of the pure-list drawing effect to the use of sequence information, highlighting the important role of relational information in driving differences in recall in the drawing effect.

Table 3

Results Provide Support for All Predictions Made by the Item-Order Account (McDaniel & Bugg, 2008)

Prediction	Supported?
1. Memory for order worse in pure + relative to pure − lists.	Yes
2. Memory for item better in pure + relative to pure − lists.	Yes[a]
3. In mixed lists, memory for order equal across + and −.	Yes
4. In mixed lists, memory for order between pure + and pure −.	Yes
5. In mixed lists, memory for item better for + than −.	Yes
6. In pure lists, input−output correspondence higher in − relative to +.	Yes
7. Memory for − items better in pure relative to mixed lists.	Yes
8. Memory for + items better in mixed relative to pure lists.	Yes

Note. + represents elaborative encoding; − represents common or passive encoding.
[a] When memory for order was equivalent, a pure-list drawing effect occurred, presumably due to item information.

The observed correspondence between the attenuation of the drawing effect and the use of order has precedence in the literature from studies of group-level effects. There are some cases demonstrating that list length moderates the size of design effects across mixed and pure lists (e.g., Mulligan et al., 2016) and that as list length increases, sequence information contributes less to free recall (Mulligan & Lozito, 2007). Thus, when probing for the influence of relational information in pure silently read lists, its influence is likely to be larger in short lists than in longer lists. Relative to drawing, stronger relational information for silently read items would presumably drive up recall, which would attenuate and even possibly reverse the drawing effect, as was evident here for our manipulations of list length (i.e., short lists in Experiments 1 and 3).

We want to highlight a notable difference between our findings and those of Wammes et al.'s (2016), who provided the first report of the drawing effect as a possible design effect. In their study, they found a significant pure-list drawing effect when comparing drawing to simply writing the study word repeatedly (drawing $M = .23$, writing $M = .14$), but here we found no significant drawing effect for pure lists in any of our experiments. There are several key differences between our work and this earlier work by Wammes et al. that might underlie this pattern difference. One possibility is list length. Wammes et al. used long lists of 66 items, whereas we used lists of 20 and 34 items. Previous research has revealed that recall retains less of the studied order as list length increases (Mulligan & Lozito, 2007), which could bring down the memory performance rates for the more common encoding type—writing or reading—leading to a larger pure-list drawing effect. Thus, perhaps Wammes et al. found a significant pure-list drawing effect because their lists were nearly twice as long as our longest list. However, we did not observe any prominent difference between Experiments 1 and 2, despite differing list lengths (20 items vs. 34 items, respectively). This finding suggests that another factor may play a greater role.

A second possibility, then, involves design. In the present research, every participant experienced all list types multiple times. Conversely, Wammes et al. (2016) used a between-subjects design, wherein each participant studied just a single list. It is possible that our participants developed a strategy over the course of the experiment (or perhaps even as early as the second list) that ultimately attenuated or reversed the drawing effect. It is difficult to assess this possibility because we do not have enough power to compare only List 1 performance from each participant in a between-subjects analysis.

Finally, the most readily apparent and supported possibility is related to differences in the choice of control condition. In our experiments, the control condition was passive silent reading; in the study by Wammes et al. (2016), the control condition involved repeated writing of the study word. Prior work has revealed that an active response task that is not relational in nature severely disrupts memory for order (Jonker & MacLeod, 2015), even if that task is as nonelaborative as a simple key press. Thus, it is quite plausible that Wammes et al. observed a pure-list effect because the active task of repeatedly writing a word wiped out any relational encoding in the control condition as well. In the present work, we chose a passive common encoding task (i.e., silent reading) to align our work with other design effects (e.g., generation, production, perceptual interference, testing effect, and en-

actment in some cases). And with this passive encoding task, we indeed found that the drawing effect functions quite like other design effects.

Taken together, these three experiments confirm that drawing produces a reliable design effect, joining a large and growing class of phenomena. Furthermore, we provide compelling evidence that drawing disrupts memory for order relative to silent reading. This evidence comes from three different classes of analyses: distance and interitem association analyses of free-recall outputs, an experimentally controlled test of memory for order, and an individual differences analysis across experiments. All of these converged on the same conclusion: Memory for item order is poorer following drawing compared to silent reading, despite a substantial mnemonic benefit for the items themselves—the drawing effect.

In an educational context, drawing shows potential for both learning of visual materials (e.g., depiction of cellular structures; diagram of a machine engine) and, as preliminary evidence suggests, learning of textbook definitions (Wammes et al., 2017b). However, the present work suggests an important caveat: If sequence is a critical factor in learning the material (e.g., a set of steps for a biology assay), then perhaps it is better to simply read or listen to the materials. Or, ideally, the learner might profitably combine strategies and perform one pass over the material using each technique to optimize representations of the material. It is certainly known from prior research (e.g., Nelson & Hill, 1974) that the availability of multiple retrieval paths can be a valuable contribution to successful remembering.

References

Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science, 333,* 1096–1097. http://dx.doi.org/10.1126/science.1204153

Burns, D. J. (1990). The generation effect: A test between single- and multifactor theories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1060–1067. http://dx.doi.org/10.1037/0278-7393.16.6.1060

Cohen, R. L. (1989). Memory for action events: The power of enactment. *Educational Psychology Review, 1,* 57–80. http://dx.doi.org/10.1007/BF01326550

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A,* 497–505.

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,* 671–682. http://dx.doi.org/10.1037/0278-7393.26.3.671

Engelkamp, J., & Krumnacker, H. (1980). Imaginale und motorische prozesse beim behalten verbalen materials [Image- and motor-processes in the retention of verbal materials]. *Zeitschrift fiir Experimentelle und Angewandt Psychologie, 27,* 511–533.

Engelkamp, J., & Zimmer, H. D. (1989). Memory for action events: A new field of research. *Psychological Research, 51,* 153–157. http://dx.doi.org/10.1007/BF00309142

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20,* 497–514. http://dx.doi.org/10.1016/S0022-5371(81)90138-9

Hopkins, R., & Edwards, R. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior, 11,* 534–537. http://dx.doi.org/10.1016/S0022-5371(72)80036-7

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,* 300–305. http://dx.doi.org/10.1037/a0033337

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,* 441–448. http://dx.doi.org/10.1037/a0034977

Jonker, T. R., & MacLeod, C. M. (2015). Disruption of relational processing underlies poor memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 831–840. http://dx.doi.org/10.1037/xlm0000069

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319,* 966–968. http://dx.doi.org/10.1126/science.1152408

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science, 26,* 390–395. http://dx.doi.org/10.1177/0963721417691356

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 671–685. http://dx.doi.org/10.1037/a0018785

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review, 15,* 237–255. http://dx.doi.org/10.3758/PBR.15.2.237

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language, 27,* 521–536. http://dx.doi.org/10.1016/0749-596X(88)90023-X

Mulligan, N. W. (1999). The effects of perceptual interference at encoding on organization and order: Investigating the roles of item-specific and relational information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 54–69. http://dx.doi.org/10.1037/0278-7393.25.1.54

Mulligan, N. W. (2000). Perceptual interference and memory for order. *Journal of Memory and Language, 43,* 680–697. http://dx.doi.org/10.1006/jmla.2000.2721

Mulligan, N. W., & Lozito, J. P. (2007). Order information and free recall: Evaluating the item-order hypothesis. *Quarterly Journal of Experimental Psychology, 60,* 732–751. http://dx.doi.org/10.1080/17470210600785141

Mulligan, N. W., & Peterson, D. J. (2015a). Negative and positive testing effects in terms of item-specific and relational information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 859–871. http://dx.doi.org/10.1037/xlm0000056

Mulligan, N. W., & Peterson, D. J. (2015b). The negative testing and negative generation effects are eliminated by delay. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 1014–1025. http://dx.doi.org/10.1037/xlm0000070

Mulligan, N. W., Susser, J. A., & Smith, S. A. (2016). The testing effect is moderated by experimental design. *Journal of Memory and Language, 90,* 49–65. http://dx.doi.org/10.1016/j.jml.2016.03.006

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,* 702–709. http://dx.doi.org/10.1037/0278-7393.17.4.702

Nelson, T. O., & Hill, C. C. (1974). Multiple retrieval paths and long-term retention. *Journal of Experimental Psychology, 103,* 185–187. http://dx.doi.org/10.1037/h0036840

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology, 5,* 176–206. http://dx.doi.org/10.1016/0010-0285(73)90032-7

Psychology Software Tools, Inc. [E-Prime 2.0]. (2016). Retrieved from http://www.pstnet.com

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x

Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition, 21,* 34–40. http://dx.doi.org/10.3758/BF03211162

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). afex: Analysis of Factorial Experiments [Computer software]. Retrieved from https://cran.r-project.org/package=afex

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592–604. http://dx.doi.org/10.1037/0278-7393.4.6.592

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *Quarterly Journal of Experimental Psychology, 69,* 1752–1776. http://dx.doi.org/10.1080/17470218.2015.1094494

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2017a). Creating a recollection-based memory through drawing. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. http://dx.doi.org/10.1037/xlm0000445

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2017b). Learning terms and definitions: Drawing and the role of elaborative encoding. *Acta Psychologica, 179,* 104–113. http://dx.doi.org/10.1016/j.actpsy.2017.07.008

Wammes, J. D., Jonker, T. R., & Fernandes, M. A. (2018). *Drawing improves memory: The importance of multimodal context.* Manuscript under review.