Journal of Experimental Psychology:
Learning, Memory, and Cognition
2018, Vol. 44, No. 5, 734–751

0278-7393/18/$12.00 http://dx.doi.org/10.1037/xlm0000445

Creating a Recollection-Based Memory Through Drawing

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

Drawing a picture of to-be-remembered information substantially boosts memory performance in free-recall tasks. In the current work, we sought to test the notion that drawing confers its benefit to memory performance by creating a detailed recollection of the encoding context. In Experiments 1 and 2, we demonstrated that for both pictures and words, items that were drawn by the participant at encoding were better recognized in a later test than were words that were written out. Moreover, participants' source memory (in this experiment, correct identification of whether the word was drawn or written) was superior for items drawn relative to written at encoding. In Experiments 3A and 3B, we used a remember-know paradigm to demonstrate again that drawn words were better recognized than written words, and further showed that this effect was driven by a greater proportion of recollection-, rather than familiarity-based responses. Lastly, in Experiment 4 we implemented a response deadline procedure, and showed that when recognition responses were speeded, thereby reducing participants' capacity for recollection, the benefit of drawing was substantially smaller. Taken together, our findings converge on the idea that drawing improves memory as a result of providing vivid contextual information which can be later called upon to aid retrieval.

Keywords: memory, drawing, recollection, response deadline, mnemonic

Classic research (Paivio & Csapo, 1973) and our own more recent work (Wammes, Meade, & Fernandes, 2016) has demonstrated that creating one's own image through drawing can confer a benefit to memory beyond that gained from simply viewing an image. These findings indicate that drawing a picture of some to-be-remembered stimulus is a potent, while also simple and easily implemented, strategy through which one can improve later memory. However, uncertainty remains about how and why drawing imparts a benefit to memory, and whether its beneficial effects apply widely across testing methods and memory tasks.

In our previous work, we demonstrated that drawing pictures of words at encoding enhanced free recall performance more than writing words, semantic generation, visual imagery, and viewing pictures. In the current work, we address perhaps the most important of the questions that were left unanswered following this work: How does drawing exert its memorial benefit? We previously proposed, and provided preliminary evidence for the hypothesis that drawing produces a better-integrated memory trace (Wammes et al., 2016). That is, we suggested that drawing leads to better memory than other comparable encoding strategies because it facilitates the integration of a basic verbal memory trace with other multisensory information derived from the encoding experience. We argued that the memory trace created by drawing consists of three integrated components: the semantic/elaborative information required in deciding what to draw, motor information from physically producing the image, and pictorial information from the final drawing that was created. We also provided preliminary evidence to suggest that the estimated contributions of these three components, when each was isolated through subtractions and then added together, still fell slightly short of achieving the level of recall performance attained by drawing alone.

For this integrated trace hypothesis to be plausible however, participants must be able to effectively retrieve specific contextual information from the initial encoding experience to a greater extent when they had drawn items, relative to when they had written them. In other words, the participant must experience a "recollection" of the drawing experience. Within dual-process models of recognition memory, theorists posit that successful retrieval can be driven by two processes: recollection and familiarity (Gardiner, 2001; Perfect, Mayes, Downes, & Van Eljk, 1996; Tulving, 1983, 1985; Yonelinas, 2002). Recollection-based memory encompasses the rich, vivid experiences where one is able to consciously bring to mind specific contextual details about the initial encoding event. Conversely, familiarity-based memory is a general phenomenological feeling of familiarity or perceptual fluency, wherein one has the sense that they have been exposed to the item before, but cannot attribute any specific contextual details to that feeling (Yonelinas, 2002). While there are several different conceptualizations of dual-process models (e.g., Atkinson & Juola, 1973; Graf & Mandler, 1984; Jacoby & Dallas, 1981; Juola, Fischler,

This article was published Online First November 2, 2017.

Jeffrey D. Wammes, Melissa E. Meade, and Myra A. Fernandes, Department of Psychology, University of Waterloo.

This research was funded by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) awarded to MF. It was also supported by a postgraduate scholarship from NSERC, awarded to author JW. We would also like to thank Britt Anderson, for generously allowing us to use his laptop/tablet for data collection, and the research assistants who collected the data for these experiments: Rachel Cao, Robin Chang, Varsha Gobin, Erin Kim, Liat Kofler, Priya Mehta, Brady Roberts, and Grace Sim.

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

Wood, & Atkinson, 1971; Mandler, 1980; Yonelinas, 1994, 2002), it is not our aim to arbitrate. Rather, we focus on the commonalities between models as they pertain to the current work's goals. Most importantly for our study, it is generally agreed upon that recollection is accompanied by specific contextual details from the encoding experience, and that it is a slower-unfolding process than familiarity (Atkinson & Juola, 1973; Yonelinas, 2002).

We proposed that drawing improves memory by integrating elaborative, motoric, and pictorial components of a memory trace. A necessary precondition for this integration hypothesis would be a demonstration that drawing improves recollection in particular. To inform our predictions, we can explore the relation these components might have with recollection and familiarity through the lens of their most analogous respective literatures (i.e., deep levels of processing [LoP], enactment effect, picture superiority). Deep LoP manipulations, which require that participants attend to and elaborate upon the *meaning* of the studied information, tend to selectively increase recollection (e.g., Gardiner, 1988; Gardiner, Java, & Richardson-Klavehn, 1996; Rajaram, 1993) as evidenced by a smaller generation effects when responses are speeded (Mulligan & Hirshman, 1995). There is evidence to suggest that enactment manipulations, which improve memory by instructing participants to perform an action in response to a presented item, also target recollection-based memory exclusively (Engelkamp & Dehn, 1997; Lövdén, Rönnlund, & Nilsson, 2002; Manzi & Nigro, 2008). Findings, however, are less clear with respect to whether picture superiority is driven by recollection or familiarity. Some findings suggest a role for both of these processes (e.g., Defeyter, Russo, & McPartlin, 2009; Dewhurst & Conway, 1994; Wagner, Gabrieli, & Verfaellie, 1997), while others indicate that recollection is most critical (e.g., Boldini, Russo, Punia, & Avons, 2007; Curran & Doyle, 2011; Rajaram, 1996). Taken together though, this constellation of findings suggests that drawing might serve to increase recollection primarily, but also could boost familiarity due to the influence of pictorial information.

While we make our predictions based on the foregoing, we do not yet know whether one's phenomenological experience when retrieving drawn items is consistent with a general feeling of familiarity or perceptual fluency, or with a rich, recollective experience, specific in time and place. To answer this question, in the present studies, we employed multiple variants of recognition memory tasks to determine whether drawing to-be-remembered words produces a contextually strong, recollective memory of an item, as opposed to a more familiar or gist-based memory.

To measure the contribution of recollection to the beneficial effects of drawing on memory, we chose three different paradigms that have been employed in the literature to either differentiate between recollection and familiarity, or to isolate recollection. We took such an approach because each paradigm independently carries its own strengths, weaknesses, and theoretical criticisms. To overcome these potential methodological issues, we examined the same phenomenon in three different ways, and predicted that the strength of evidence across tasks would support one common conclusion. In this way, any findings would not be an artifact of the failings of a particular paradigm. Specifically, we employed source memory decisions (Experiments 1 and 2), a remember-know/new (RKN) paradigm (Experiment 3), and a response-deadline method (Experiment 4). In so doing, we hope to determine whether drawing leads to a rich recollective memory trace, and to subsequently determine whether speeding responses, and thus reducing the influence of recollection, eliminates the drawing effect.

In Experiment 1, we contrasted the effect on memory of drawing, relative to our commonly used control of writing at encoding, using a source memory decision at test. We predicted both overall memory, and source memory would be better for drawn items as drawing likely produces a rich contextual memory trace. In Experiment 2, we changed the to-be-remembered information from words to pictures. Again we expected overall memory, and source memory, to be higher for items drawn at encoding. Because we had in essence added an otherwise absent visual trace to written items, but only a potentially redundant second visual trace in the drawn items, we speculated that the beneficial effects of drawing might be smaller in Experiment 2 than in Experiment 1. In Experiment 3A and 3B, we employed a traditional remember-know/new (RKN) paradigm. Similar to Experiment 1, our prediction was for better overall memory, as well as a greater proportion of recollection-driven responses for drawn items. In Experiment 4, we used a response deadline procedure in a basic old/new recognition task. Our prediction was that if memory for drawn items is driven largely by recollection-based responses, then speeding the decision with a response-deadline would lead to a smaller drawing effect than would more typical relaxed timing constraints.

Experiment 1

Because we are most interested in the potential contribution of recollection to memory for drawn, relative to written items, we began with a paradigm requiring participants to make source memory decisions. Our reason for choosing this is that in order to correctly identify the source (draw or write) of a studied item, one must remember specific details about what task was performed with the item during the encoding experience (i.e., recollection). Thus, while familiarity can be associated with above-chance source memory decisions, source memory is thought to rely primarily on recollection (Ranganath et al., 2004; Wais, Mickes, & Wixted, 2008; Wais, Squire, & Wixted, 2010; Yonelinas, 1999). In support of this idea, research has shown that there is an increase in the magnitude of the late positive component (LPC) for high confidence source memory decisions (Addante, Ranganath, & Yonelinas, 2012), that source memory decisions are associated with response patterns in the prefrontal cortex which are commonly associated with recollection (Dobbins, Foley, Schacter, & Wagner, 2002; Hayama, Vilberg, & Rugg, 2012; Leiker & Johnson, 2015), and that recollection-based responses are more often associated with correct source identifications than familiarity-based responses (Wais et al., 2008). If drawing indeed produces a rich recollection-based memory trace, then participants should be better able to correctly identify the source of drawn items than of written items. Thus, our main prediction for this experiment was that source memory would be better for drawn words than for written words.

A second prediction was that overall recognition performance (regardless of the correctness of source decisions) would be higher for drawn than written words. This is not a trivial prediction, as while we previously demonstrated that free recall was better for drawn than written items, we have not yet demonstrated this "drawing effect" in any variant of a recognition task.

Method

Participants. Participants were 36 undergraduate students (27 female) at the University of Waterloo, who completed the experiment for course credit. Participants ranged in age from 18 to 27 (M = 20.00, SD = 1.97).[1]

Materials

Target items. A set of 80 concrete nouns used in previous work (Wammes et al., 2016) were target items on our memory test. This list of words ranged in frequency between one and 25 (M = 8.23, SD = 6.44), in length between three and 11 letters, (M = 5.56, SD = 1.79), and in number of syllables from one to four (M = 1.63, SD = .72).

Filler task. Audacity software (Mazzoni & Dannenberg, 2000) was used to create sine wave tones which were exactly 500-ms long, at low (350 Hz), medium (500 Hz), and high (650 Hz) frequencies. These were used during the retention interval in a continuous tone classification filler between study and test in the immediate testing condition.

Procedure. Participants were tested in groups in a lecture hall, as our previous work showed no impact of testing in group setting or individually on the drawing effect (Wammes et al., 2016). Stimulus presentation and timing was controlled using E-prime v2.0 software (Psychology Software Tools Inc., Pittsburgh, PA) and displayed via a projection screen. Instructions were presented visually on-screen and verbally by the experimenter. The experimenter informed the participants that they would be shown a prompt word conveying instructions for each trial, followed by the target word. The prompt word instructed participants to either "draw" or "write" the target word for that trial, using the pad of paper provided (14 cm × 21 cm).

When participants saw the "draw" prompt they were told to draw a picture illustrating the word on the screen. Each participant was provided with a sharpened pencil with an eraser on the end with which to complete their drawing. If a participant finished their drawing prior to seeing the next prompt word, they were told to continue adding detail until the trial time had elapsed. When participants saw the "write" prompt they were told to clearly and carefully write out the word multiple times. Participants were informed that they would hear a tone to alert them that the trial time had elapsed. This was done to ensure participants would stop drawing (or writing) and would be would have ample time to flip to the next page in their drawing pad, and be ready to view the prompt for the next trial. Participants were not informed that their memory would later be tested, to reduce the possibility that they might preferentially focus on drawn items in anticipation of a test.

Encoding. Prior to the experiment, participants completed one "draw" and one "write" trial type, in order to familiarize them with the task. Each participant, or group, was presented with a uniquely random subset of 40 words. The remaining 40 words were retained to serve as lures in the source memory task. In all conditions, half of the 40 target words were randomly assigned to the "draw" trial type and half to the "write" trial type.

For each of the 40 trials, a prompt ("draw" or "write") appeared in the center of the screen for 750 ms, followed by a 500-ms fixation. Next the participants viewed the study word for 750 ms. Participants were then given 40 s to perform the task indicated by the prompt, after which a tone alerted them that the trial was over, and they were to prepare for the next item. After the tone, participants had 3 s to turn over the next page in their pad of paper before the subsequent trial began.

Filler task. Participants completed a 2 min, continuous tone classification task immediately following encoding. Participants were presented with 60 tones, which they were instructed to classify as low, medium or high, by pressing the 1, 2, or 3 key on a small response pad. For each trial, participants heard a tone for 500 ms, and then had 1,500 ms within which to provide their response before the next tone was heard. This task was included for no other reason than to introduce a delay between encoding and retrieval. After classifying all tones, participants completed the retrieval task.

Retrieval. The retrieval phase consisted of a source memory recognition task. Participants were told that they would view a set of words one at a time, and that it would be their job to indicate whether or not each item was one they remembered from the study phase. They were given a Scantron sheet, typically used in multiple choice examinations, with which to indicate their responses. These sheets consist of a set of columns, each with 10 rows containing a number (e.g., 11–20), followed by the options A through E, each with a circle surrounding the letter. Participants fill in the corresponding circle to indicate their letter response to each numbered question. Rather than a typical old/new recognition task, participants indicated whether an item was drawn, written or not studied by shading in a bubble on their score sheet (A, B, or C, respectively).

The 40 studied words were presented among 40 lures. Each word was presented centrally on screen for 3 s. The recognition test trials were numbered consecutively from one to 80, and this number was presented in the top left of the screen during the test. This was done so that, in the event that a participant missed an item, they could reorient themselves to the appropriate numbered row on their Scantron sheet. The answer key was also presented on screen for every recognition test trial, indicating which letter corresponded to which response (i.e., Drawn [A], Written [B], Not Studied [C]). Following each trial, a blank screen was presented for 1 s.

Results

In all analyses, we report paired-samples, or when required, independent samples t tests. Effect sizes (d) were calculated based on Morris and Deshon's (2002) Equation 8 for within-subject designs, with the standard deviations of both trial types averaged, and correcting for the correlation among scores. However, because performance levels were so high in some cases as a result of the drawn trial type, the distribution of our data are often nonsymmetrical. Moreover, the draw and write trial types do not often have equal variances. Accordingly, we also include Wilcoxon's signed rank (WSR) tests, and paired-samples sign (PSS) tests in instances where paired-samples tests are used, and Wilcoxon's ranked sum (WRS) tests where independent-samples t tests are used. These are nonparametric tests that do not assume normality

[1] In all studies, we aimed to collect a sample of 35 participants. Because experimental sessions were often scheduled in blocks, and participants occasionally failed to attend their scheduled timeslot, there was some variation around this number.

or equal variances. In the Appendix, we also analyze the data using a multilevel logistic approach. While the outcome of these analyses are largely concordant with those presented here, we note any relevant differences in footnotes.

Hit rate. Trials were scored as hits if they were correctly identified as old, regardless of whether participants indicated that they were drawn or written. Paired-samples *t* tests indicated that hit rate was higher for words that were drawn than written at encoding, $t(35) = 6.22$, $SE = .02$, $d = 1.15$, $p < .001$. These results were corroborated by WSR, $Z = 4.74$, $p < .001$ and PSS tests, $Z = 4.53$, $p < .001$. Hit rates and the false alarm rate are reported in Table 1.

Correct source. For each participant, correct source was computed as the number of old items in which the encoding trial type was correctly identified, divided by the number of total hits, regardless of whether the trial type was correctly identified. While the comparison between drawn and written items was not conventionally significant and only marginal, $t(35) = 1.74$, $SE = .02$, $d = .34$, $p = .09$ (WSR $Z = 2.49$, $p < .05$, PSS $Z = 1.89$, $p = .056$), the effect did trend in the predicted direction, such that in drawn items, the source was correctly identified for a greater proportion of the hits than in written items. The comparison was also significant or marginal in both nonparametric tests. The raw distribution of responses for drawn, written, and unstudied trial types is displayed in Table 2. As can be seen, when not weighted as a proportion of hits, drawn items were more often correctly identified as drawn than written, $t(35) = 5.79$, $SE = .03$, $d = .1.07$, $p < .001$ (WSR $Z = 4.70$, $p < .001$, PSS $Z = 4.42$, $p < .001$).

Sensitivity (d'). We computed sensitivity in two different ways (Both are reported in Table 1). First, using the hit rate as outlined above, and the overall false alarm rate, collapsed across new trials that were incorrectly identified as *either* drawn or written. Similar to previous work (Forrin, Groot, & MacLeod, 2016) we also computed d' by source decision, such that our hit rate included only trials in which the source was correctly identified (i.e., responses of "drawn" to items that were indeed drawn), and only false alarms misidentified as a particular source (i.e., responses of "drawn" to items that were not studied). In this way, we had a separate false alarm rate for each trial type. In all d' calculations within this article, extreme values were corrected using the log-linear rule, which leads to less-biased and more conservative values than other commonly used corrections (Hautus, 1995). In either case, drawing led to greater sensitivity than did writing, $t(35) = 7.90$, $SE = .10$, $d = 1.34$, $p < .001$ (WSR $Z = 4.79$, $p < .001$, PSS $Z = 4.53$, $p < .001$); $t(35) = 7.44$, $SE = .13$, $d = 1.25$, $p < .001$ (WSR $Z = 4.78$, $p < .001$, PSS $Z = 4.42$, $p < .001$), respectively.

Table 1
Mean (Standard Error in Parentheses) Hit Rate, False Alarm Rate, Proportion Correct Source, Sensitivity, and Sensitivity by Source Decision for Each Item Type in Experiments 1 and 2

Experiment and measure	Draw		Write
Experiment 1			
Hit rate	.98 (.01)		.83 (.02)
Correct source	.98 (.01)		.94 (.02)
Sensitivity (d')	3.88 (.07)		3.07 (.09)
d' by source	3.86 (.08)		2.93 (.10)
False alarm rate		.01 (.00)	
Experiment 2			
Hit rate	.95 (.02)		.88 (.02)
Correct source	.97 (.02)		.92 (.02)
Sensitivity (d')	3.60 (.16)		3.22 (.15)
d' by source	3.58 (.17)		3.12 (.15)
False alarm rate		.03 (.02)	

Table 2
Mean Proportion (Standard Errors in Parentheses) of Each Possible Response (Draw, Written, not Studied) Given in Response to Each Actual Trial Type (Draw, Write, New) in Experiments 1 and 2

Experiment and response	Actual trial type		
	Draw	Write	New
Experiment 1			
Drawn	.96 (.01)	.05 (.02)	.003 (.001)
Written	.02 (.01)	.77 (.03)	.008 (.003)
Not studied	.02 (.01)	.17 (.02)	.988 (.004)
Experiment 2			
Drawn	.93 (.03)	.06 (.02)	.02 (.01)
Written	.02 (.01)	.82 (.03)	.01 (.01)
Not studied	.05 (.02)	.12 (.02)	.97 (.02)

Discussion

Our results from Experiment 1 demonstrate that drawing an item at encoding, relative to writing it out, leads to increased overall recognition memory performance, but more importantly, to increased accuracy in identifying the source of the encoded information at test. Accordingly, Experiment 1 provides the first direct evidence suggesting that drawing increases recollection.

Experiment 2

Experiment 1 showed a benefit both in raw memory performance and in recollection of source information for drawn, relative to written words. In Experiment 2, we sought to determine whether this benefit would remain when the to-be-remembered stimuli were pictorial. This was critical for two reasons: First, to test the generality of the benefit drawing affords to memory, as well as the generality of the boost it provides to recollection specifically, and second, to rule out a possible confound in our design. That is, it could be the case that the recollective benefit we had observed was driven simply by the presence of an image in the "draw" trial type (but not "write"), rather than the act of drawing itself. This is important to rule out in light of previous work which has shown that the picture superiority effect is driven mainly by recollection (e.g., Dewhurst & Conway, 1994; Rajaram, 1996). With this, in Experiment 2, we aimed to test whether the benefit in source memory would remain when *both* encoding orientations contained an image. The critical difference then between Experiment 1 and 2 is that the to-be-remembered stimuli were presented in picture rather than word format during encoding. If, as we have proposed here and in our previous (Wammes et al., 2016) work, drawn words are afforded better memory performance partially owing to contributions from dual-coding/picture superiority (in addition to elaborative coding and motor enactment), then the benefit should be smaller in magnitude for pictures than it was for words. If the benefit remains, the presence of an image alone could not be responsible for an increase in correct identifications of source. This is because pictorial stimuli should already inherently lead to memorial benefits due to picture superiority, thus minimizing any additional benefit resultant from actually producing a picture while drawing. As a result, only elaborative and motoric information would remain to improve retention and identification of the correct source for items drawn relative to written.

In this experiment, we again predicted that drawing would benefit source memory, and that overall recognition performance would be higher for drawn than written words. One might instead suggest an alternative hypothesis, such that the drawing effect would be greater for stimuli presented as pictures than as words, because the participant would benefit from having a stronger visuospatial representation, where one representation originates from the initial presentation format, and another from the drawing that they themselves created. However, we predicted the former, that the effect of drawing might be smaller here, as the written items, by virtue of being presented in pictorial form, will now *also* have the benefit of the visual trace, thus lessening the difference between "write" and "draw" trial types.

Method

Participants. Participants were 34 undergraduate students (26 female) at the University of Waterloo, who completed the experiment for course credit. Participants ranged in age from 18- to 25-years-old ($M = 19.00$, $SD = 1.44$).

Materials. Materials were the same as in Experiment 1, except for the words were replaced with images. Images corresponding to each of the 80 concrete nouns from Experiment 1 were compiled from a larger set of object images (Brady, Konkle, Alvarez, & Oliva, 2008), then converted to gray scale using Image Magick software.

Procedure. The procedure was identical to Experiment 1, except in the case of the "write" trial type. Because images were now presented instead of words, the instructions participants read asked them to write out the *verbal label* of the image (i.e., rather than word) multiple times.

Results

Hit rate. Paired-samples t tests indicated that hit rate was higher for drawn pictures than for written pictures, $t(35) = 4.20$, $SE = .02$, $d = 0.73$, $p < .001$. Hit rates and the false alarm rate are reported in Table 1.

Correct source. Participants identified the correct source on a greater proportion of the total hits for drawn items than for written items, $t(33) = 3.90$, $SE = .01$, $d = .72$, $p < .001$ (WSR $Z = 3.58$, $p < .001$, PSS $Z = 3.77$, $p < .001$). Not surprisingly, as a proportion of the *total* items in a trial type (i.e., not just total hits) the source was correctly identified more frequently for drawn items as well, $t(33) = 5.61$, $SE = .02$, $d = .99$, $p < .001$ (WSR $Z = 4.36$, $p < .001$, PSS $Z = 4.51$, $p < .001$; see Table 2).[2]

Sensitivity (d'). For both overall d' and d' by source, drawing led to greater sensitivity than writing, $t(33) = 4.13$, $SE = .09$, $d = 0.71$, $p < .001$ (WSR $Z = 3.45$, $p < .001$, PSS $Z = 2.84$, $p < .001$); $t(33) = 5.29$, $SE = .09$, $d = 0.92$, $p < .001$ (WSR $Z = 4.04$, $p < .001$, PSS $Z = 3.46$, $p < .001$), respectively.

Discussion

Results from Experiment 2 show that the actual act of drawing is important in increasing memory performance, as when the trial types were equated on the presence of an image, the benefit to overall memory, and memory for source remained. Taken together, the findings from Experiments 1 and 2 indicate that drawing leads to an increase in recollection. This conclusion follows from previous findings indicating that source memory, by definition, is mostly recollection-driven (Addante et al., 2012; Migo, Mayes, & Montaldi, 2012; Wais et al., 2008, 2010). However, research has determined that familiarity-based responses to some items can be associated with correct identification of source at a rate higher than chance (Kelley, Jacoby, & Hollingshead, 1989; Wais et al., 2008). The prevailing idea seems to be that most correct identifications of source are driven by recollection, but not all recollections are driven by information about source. To clarify, one could experience a recollection of some extra-item information that is *not* related to the source (e.g., someone sneezed in the hall, or that the word "dog" reminded them they were dog-sitting that weekend). In these examples, because the recollection was not source-related, it would not be captured and subsequently measured by a source memory decision (Yonelinas & Jacoby, 1996). Thus, source memory tasks fail to detect some aspects of recollection memory.

Experiment 3A

We sought to use Experiment 3 to more exhaustively sample the contribution of recollection to the subjective state of participants' memory using the remember-know paradigm (Tulving, 1985; Gardiner, 1988). One advantage of the remember-know paradigm is that it provides a more inclusive measure of recollection. Namely, it is not limited by what specific contextual information the experimenter requests, which is the critical shortcoming of the source memory decision employed in Experiments 1 and 2. The remember-know task was developed to allow participants to explicitly report whether their memory was based on recollection processes (i.e., they "remember" the item), or on familiarity processes (i.e., they "know" the item was present). Results from this paradigm typically converge on those using other methods of capturing recollection and familiarity (Yonelinas, 2002). The instruction encourages participants to focus on whether or not they remember specific episodic information about the encoding experience, and to answer "remember" if they do. Thus, unlike source memory decisions, which only capture recollections that are associated with knowing what task they performed in response to the studied item, the remember-know paradigm should capture nearly *all* recollection-based hits.

The remember-know paradigm has been used extensively as a proxy for the recollection and familiarity processes that underlie recognition memory. However, there are differing opinions on whether or not the paradigm actually measures separable psychological processes (Rotello & Zeng, 2008; Starns & Ratcliff, 2008). Specifically, the he exclusivity model holds that a given recogni-

[2] The outcome of the multilevel logistic regression (see Appendix) did not provide evidence consistent with the results of these more traditional analyses. When the relative proportions of source judgments in incorrect guesses are accounted for, the difference between trial types was not significant.

tion memory response can be based on recollection *or* familiarity, but not both (Gardiner & Parkin, 1990; Jones, 1987; cf. explicit vs. implicit information: Nelson, Schreiber, & McEvoy, 1992, but see Wais et al., 2008 for evidence to the contrary). The redundancy model suggests that all successfully recognized items are familiar, whether they are recollected or not (Joordens & Merikle, 1993; Knowlton & Squire, 1995; Tulving, 1985). Lastly, the independence model suggests that "remember" responses are based on recollection, while "know" responses are based on familiarity, but only the subset of instances of familiarity where there was no recollection (Jacoby, Yonelinas, & Jennings, 1997; Yonelinas & Jacoby, 1995).

Importantly however, we are more concerned with whether drawing improves recollection of studied items, a question for which remember-know is particularly well suited. Thus, in the current experiment, we used the remember-know paradigm under the assumption that "remember" responses contain more contextual detail than "know" responses, and aimed to determine the effects of drawing (relative to writing) during encoding on the subjective mnemonic experience of the rememberer. We predict that drawn items, relative to written items will be associated with an increase in "remember" responses, as well as an increase in overall memory performance.

Method

Participants. Participants were 35 undergraduate students (23 female) at the University of Waterloo, who completed the experiment for course credit. Participants ranged in age from 17- to 22-years-old ($M = 18.74$, $SD = 1.24$).

Materials. Materials were the same as in Experiment 1.

Procedure. The procedure was identical to Experiment 1, except for the retrieval stage, as the memory test was now an RKN paradigm. Participants were told that a "remember" (R) response meant that they had a conscious recollection of specific contextual information about their initial encounter with the word, such as hearing a sound in the hall, or what they did when the word was presented. They were told that a "know" (K) response meant that they had only a feeling of familiarity; that they believed that the word had been seen recently, but could not remember specific details from their initial viewing of the word. Lastly, a "new" (N) response meant that they did not encounter the word in the previous phase. Participants indicated whether an item was one that they would classify as R, K, or N by shading in a bubble on their score sheet (A, B, or C, respectively).

Results

Data from this experiment were analyzed in a few different ways. First, we compared overall recognition memory performance (collapsing remember and know responses), then compared R and K responses separately. Hit rates were calculated out of 20 possible items, while false alarm rates were out of 40 possible items. Accuracy (hit rate minus false alarm rate) and *d* prime were also computed overall and within each response type. Because hit rate and accuracy followed the same pattern as d' in all analyses, we only report d' in full, while the other values are presented in Table 3. We also computed and analyzed process estimates for recollection and familiarity based on calculations outlined in previous work (e.g., Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998). However, because the contextual quality of memory was so high in some participants (that is, their hits were exclusively "remember" responses), we were unable to reliably estimate the contribution of familiarity in these individuals. Thus, these analyses are based on only the subset of subjects wherein the estimates could be computed for familiarity in both trial types.

Overall recognition. Paired-samples *t* tests indicated that overall d' was higher for drawn words than for written words,

Table 3
Mean (Standard Error in Parentheses) Accuracy, Hit Rate, False Alarm Rate, Proportion Remember (R) Responses, Sensitivity (d′), Overall, for R Responses Only, and for K Responses Only for Each Item Type, and Process Estimates for Each Item Type in Experiment 3A and 3B

Experiment and measure	Trial Type and Response					
	Draw			Write		
	Overall	Remember	Know	Overall	Remember	Know
Experiment 3A						
Accuracy	.87 (.04)	.89 (.03)	−.02 (.04)	.70 (.04)	.58 (.05)	.12 (.05)
Hit rate	.97 (.01)	.91 (.03)	.06 (.03)	.80 (.03)	.60 (.05)	.20 (.03)
False alarm rate	.10 (.04)	.02 (.01)	.08 (.03)	.10 (.04)	.02 (.01)	.08 (.03)
d'	3.49 (.17)	3.53 (.16)	.18 (.18)	2.65 (.17)	2.35 (.16)	.84 (.18)
Experiment 3B						
Accuracy	.84 (.03)	.85 (.04)	−.02 (.03)	.60 (.04)	.51 (.04)	.10 (.02)
Hit rate	.96 (.01)	.90 (.03)	.06 (.03)	.73 (.03)	.55 (.03)	.18 (.02)
False alarm rate	.13 (.03)	.05 (.02)	.08 (.02)	.13 (.03)	.05 (.02)	.08 (.02)
d'	3.11 (.15)	3.25 (.18)	−.08 (.13)	2.00 (.14)	2.02 (.14)	.61 (.09)
		Recollection	Familiarity		Recollection	Familiarity
Process estimates						
Experiment 3A		.82 (.04)	1.77 (.42)		.55 (.05)	1.66 (.30)
Experiment 3B		.87 (.03)	1.51 (.25)		.52 (.04)	1.18 (.12)

Note. Process estimates are based on the subset of participants (3A = 18/35, 3B = 33/37) for whom these estimates could be computed. If one scores 100% "remember" hits, there is no way to estimate familiarity.

$t(34) = 8.00$, $SE = .10$, $d = 1.35$, $p < .001$ (WSR $Z = 4.64$, $p < .001$, PSS $Z = 4.93$, $p < .001$). Accuracy and hit rate were also higher for drawn than written words, $ps < .001$.

Remember responses. d' for R responses was significantly higher for drawn words than written words, $t(34) = 8.40$, $SE = .14$, $d = 1.42$, $p < .001$ (WSR $Z = 4.86$, $p < .001$, PSS $Z = 5.39$, $p < .001$). Accuracy and hit rate were also higher for drawn than written words, $ps < .001$.

Know responses. d' for K responses was significantly *lower* for drawn words than written words, $t(34) = 5.41$, $SE = .12$, $d = 0.92$, $p < .001$ (WSR $Z = 3.99$, $p < .001$, PSS $Z = 4.35$, $p < .001$). Accuracy and hit rate were also lower for drawn than written words, $ps < .001$.

Process estimates. Nearly half of our sample gave exclusively "remember" responses during recognition of items that were drawn at encoding. As such, there was no opportunity for "know" responses to be given, making it impossible to calculate the influence of familiarity, in line with the underlying assumptions of the independence model upon which the equations are based (Yonelinas et al., 1998). Accordingly, we present here only the data from the 18 participants who provided at least one "know" response, thereby rendering the contribution of familiarity calculable. Because it is based on only a subset of the sample, the results should be interpreted with caution. However, we do not believe familiarity to be making a *large* contribution, especially given that "know" responses were given so infrequently to drawn items. In any case, results indicated that the contribution of recollection was greater for drawn, relative to written items, $t(17) = 2.11$, $SE = .05$, $d = 1.34$, $p < .001$ (WSR $Z = 3.52$, $p < .001$, PSS $Z = 3.87$, $p < .001$), while there was no difference in the contribution of familiarity, $t(17) = 0.27$, $SE = .42$, $d = 0.07$, $p = .79$ (WSR $Z = 0.31$, $p = .76$, PSS $Z = 0.53$, $p = .59$).[3]

Discussion

In line with the findings from Experiments 1 and 2, drawing improved overall recognition memory, but this increase seemed to be driven by recollection more than familiarity. Specifically, d' for R responses was significantly larger for drawn than for written items. The opposite pattern occurred in the analysis of K responses, such that d' was higher for written items. The remember-know procedure, however, likely underestimates the actual contribution of familiarity to recognition memory. This is because K responses, as taken from the instructions, typically indicate that an item is familiar *and* not recollected (Yonelinas, 2002), rather than just familiar. Accordingly, using the independence remember-know method (e.g., Koen & Yonelinas, 2010; Yonelinas & Jacoby, 1995; Yonelinas et al., 1998), we were able to compute the estimated contribution of familiarity to overall recognition, based on extrapolation from the subset of items that were not recollected. Using this method, the estimated contribution of familiarity was similar across drawn and written items, though our multilevel logistic analyses (see Appendix) suggested familiarity was playing a small role in driving the benefit drawing affords memory. Still, the contribution of recollection to the benefit of drawn over written items was substantially larger than that of familiarity.

Experiment 3B

Because participants' level of recollection was so high in drawn items, it is difficult to interpret the process estimates from Experiment 3A, which were derived from only a nonrandomly selected subset of the sample. In Experiment 3B, we sought to bring performance down by increasing the number of studied items and decreasing the encoding time, in order to derive a more complete set of process estimates.

Method

Participants. Participants were 37 undergraduate students (31 female) at the University of Waterloo, who completed the experiment for course credit. Participants ranged in age from 18- to 27-years-old ($M = 19.74$, $SD = 1.79$).

Materials. The materials were the same as in Experiment 1.

Procedure. For this experiment, we eschewed the pen and paper format, in favor of an entirely computer-based paradigm, and participants were all tested individually. Stimulus presentation and response recording were controlled using Python and displayed via an Acer One 10 touchscreen 2-in-1 notebook with 10.1-inch monitor and a stylus. Rather than 40 items, as in previous experiments, 80 items (40 drawn and 40 written) were now studied, and the encoding time was reduced from 40 s to 20 s per item. In addition, there were now 60, rather than 40 lures. Because of our change in apparatus, participants now used a stylus to draw directly on the same screen where the words were presented. The computer was converted into tablet mode during the encoding phase so that the screen was flat on the desk, analogous to the sheet of paper which was lying flat on the desk in previous experiments. Because page-flipping was no longer required, the idle time that was previously filled by page-flipping between trials was omitted. Instead, when the trial was over, the screen was cleared of all content, and the next prompt simply appeared where the participant had previously been drawing or writing.

The filler task and retrieval phase were also completed using the computer. Rather than classifying the filler task tones on a sheet of paper, participants classified tones as low medium or high using number keys (1, 2, and 3, respectively). The retrieval task required participants to press one of three keys to indicate their response, instead of filling in a bubble sheet as in Experiments 1 through 3A.

Results

Overall recognition. Paired-samples t tests indicated that overall d' was higher for drawn words than for written words, $t(36) = 12.57$, $SE = .09$, $d = 2.07$, $p < .001$ (WSR $Z = 5.23$, $p < .001$, PSS $Z = 5.83$, $p < .001$). Accuracy and hit rate were also higher for drawn than written words, $ps < .001$.

Remember responses. d' for R responses was significantly higher for drawn words than written words, $t(36) = 8.91$, $SE = .14$, $d = 1.52$, $p < .001$ (WSR $Z = 4.75$, $p < .001$, PSS $Z = 5.50$,

[3] The analogous multilevel logistic regression analysis suggested that familiarity contributed more to memory for drawn than written items (see Appendix). However, the contribution of recollection to the benefit of drawn over written items was still substantially larger than the contribution of familiarity.

$p < .001$). Accuracy and hit rate were also higher for drawn than written words, $ps < .001$.

Know responses. d' for K responses was significantly *lower* for drawn words than written words, $t(36) = 5.81$, $SE = .12$, $d = 1.01$, $p < .001$ (WSR $Z = 4.47$, $p < .001$, PSS $Z = 4.50$, $p < .001$). Accuracy and hit rate were also lower for drawn than written words, $ps < .001$.

Process estimates. In this iteration, only four participants gave exclusively "remember" responses during recognition of items that were drawn at encoding, leaving 33 subjects where process estimates were calculable. The contribution of recollection was greater for drawn, relative to written items, $t(32) = 8.76$, $SE = .04$, $d = 1.53$, $p < .001$ (WSR $Z = 4.73$, $p < .001$, PSS $Z = 5.22$, $p < .001$), while the difference in the contribution of familiarity was not significant, $t(32) = 1.28$, $SE = .26$, $d = 0.26$, $p = .21$ (WSR $Z = 0.74$, $p = .46$, PSS $Z = 0.35$, $p = .73$).[4]

Discussion

Our data here replicate the findings from Experiment 3A, demonstrating that drawing leads to significantly better memory performance when looking at overall recognition. This difference was driven primarily by recollection, a pattern which was mirrored in the process estimates. While the contribution of familiarity was not significantly different between trial types, it was numerically larger for drawn than written items in both Experiment 3A and 3B, and the multilevel logistic analyses (see Appendix) pointed toward a small but significant contribution of familiarity to the benefit of drawn over written items. While this difference was not as pronounced as was the difference for the recollection estimate, it indicates that familiarity may be contributing in some way to the benefit that drawing affords memory. Nevertheless, the current work is concerned with testing whether drawing improves memory by creating a recollection-based memory trace. The evidence, from this and the experiments presented thus far, suggests that while familiarity may contribute, recollection is more critical in driving the benefit.

Experiment 4

If, as the evidence thus far indicates, the benefit that drawing affords memory is most strongly driven by recollection, it follows that this benefit should be reduced or eliminated when the opportunity to engage in the recollection process is limited or taken away. In Experiment 4, we employed a response-deadline procedure to test this prediction. The prevailing view in dual-process models is that familiarity is relatively fast, while recollection is slower (e.g., Yonelinas, 2002). Evidence for the temporal dissociation of these processes comes from event-related potential (ERP) studies demonstrating that each is uniquely associated with dissociable components. Specifically, recollection, including source memory decisions, is associated with a LPC originating in the parietal lobe (Luo, Hendriks, & Craik, 2007; Senkfor & Van Petten, 1998), while familiarity is associated with an earlier frontal and temporal component (the FN400, Addante et al., 2012; Meng, Ye, & Gonsalves, 2014). Thus, because recollection is associated primarily with a later component (Mecklinger, 2000; Rugg & Curran, 2007; Wilding & Ranganath, 2011; Woodruff, Hayama, & Rugg, 2006), it follows that recollection unfolds more slowly than familiarity.

The temporal difference between the two processes has also been supported by behavioral work demonstrating that speeding recognition responses leads to a greater reliance on familiarity (e.g., Yonelinas & Jacoby, 1994), or to responses based solely on familiarity (Besson, Ceccaldi, Didic, & Barbeau, 2012; Besson et al., 2015). Furthermore, there is evidence from a rich neuroimaging literature that has explored the time course of these processes and converged on the same point: that familiarity is a faster process than recollection (e.g., Brown & Aggleton, 2001; Staresina, Fell, Do Lam, Axmacher, & Henson, 2012). Response-deadline tasks are those that exploit this temporal distinction to determine the contribution of familiarity to memory (e.g., Sauvage, Beer, & Eichenbaum, 2010). Evidence has shown that when a recognition task is speeded, relative to when it is not, responses based on recollection are offset or eliminated, while familiarity remains intact (Benjamin & Craik, 2001; Boldini, Russo, & Avons, 2004; Bowles et al., 2007; Duzel, Yonelinas, Mangan, Heinze, & Tulving, 1997; Sauvage et al., 2010; Toth, 1996; Yonelinas & Jacoby, 1994; but see, Gardiner, Ramponi, & Richardson-Klavehn, 1999). With this, and the findings reported in the current work indicating that the benefit of drawing relies on recollection, we expect that the drawing effect will be substantially reduced or eliminated as a result of speeding retrieval responses, but will remain when a longer response window is allowed.

Method

Participants. Participants were 70 (35 per group) undergraduate students (51 female) at the University of Waterloo, who completed the experiment for course credit. Participants ranged in age from 17 to 33 ($M = 20.11$, $SD = 2.27$). Performance in two of the participants in the speeded group, however, was below chance, meaning that they had an equal or greater amount of false alarms than they did hits. Accordingly, these participants were replaced with additional participants to compose the full sample of 70.[5]

Materials. Materials were the same as in Experiment 1.

Procedure. For this experiment, we again employed an entirely computer-based paradigm, and participants were all tested individually. Stimulus presentation and response recording were controlled using Python and displayed via a Toshiba Portege M750 touchscreen laptop/tablet with 12.5-inch monitor and a stylus. The timing and display of prompts and words was identical to Experiment 1, and the manner in which the act of drawing and the filler task was implemented was identical to Experiment 3B.

For retrieval, participants completed an old/new recognition task, incorporating a response deadline procedure. Participants were evenly divided into two groups. In the speeded group, a response deadline of 800 ms was imposed for each trial on the recognition test, while participants in the control group had 2,400 ms to respond to each trial. In either case, they were told that if an item was old, meaning that they had previously studied it, that they

[4] The analogous multilevel logistic regression analysis again suggested a smaller, but nevertheless significant contribution of familiarity (see Appendix).

[5] It is important to note that the inclusion of these participants does not alter the pattern of data in any substantive way. All significance tests follow precisely the same pattern, whether they are included in the analysis or not.

were to press the number 1, while if it was new, they were to press the number 0. Participants in each group were informed of how much time they had to respond to each trial, and that they would hear a short, shrill tone if they missed the deadline in either condition. They were also told that this tone served to remind them to respond more quickly on the next trial.

Results

Hit rate. Hit rate was analyzed using a 2 × 2 mixed measures ANOVA with group (speeded, control) as a between-subjects factor and trial type (draw, write) as a within-subjects factor. There was a significant main effect of both group, $F(1, 68) = 35.66$, $MSE = .05$, $p < .001$, $\eta^2 = .34$, and trial type, $F(1, 68) = 59.85$, $MSE = .01$, $p < .001$, $\eta^2 = .47$, as well as a significant interaction, $F(1, 68) = 19.84$, $MSE = .01$, $p < .001$, $\eta^2 = .23$. In unpacking the interaction, results show that drawing led to significantly more hits than did writing in both groups, though the benefit was smaller in the speeded group, $t(34) = 2.47$, $SE = .02$, $d = .51$, $p < .025$ (.05/2; WSR $Z = 2.44$, $p < .05$, PSS $Z = 1.95$, $p = .052$), than in the control group, $t(34) = 7.66$, $SE = .03$, $d = 1.32$, $p < .001$ (WSR $Z = 4.98$, $p < .001$, PSS $Z = 4.97$, $p < .001$). There was a large significant decrease in both drawn hit rate, $t(68) = 8.62$, $SE = .04$, $d = 2.06$, $p < .001$ (WRS $Z = 6.74$, $p < .001$), and in write hit rate, $t(68) = 3.01$, $SE = .05$, $d = 0.72$, $p < .01$ (WRS $Z = 3.01$, $p < .01$), in the speeded, relative to the control group (Table 4).

Sensitivity (d'). Sensitivity was analyzed in the same manner as hit rate. There was a significant main effect of both group, $F(1, 68) = 193.59$, $MSE = 0.74$, $p < .001$, $\eta^2 = .74$, and trial type, $F(1, 68) = 81.75$, $MSE = 0.14$, $p < .001$, $\eta^2 = .55$, as well as a significant interaction, $F(1, 68) = 44.92$, $MSE = 0.14$, $p < .001$, $\eta^2 = .40$. The interaction was driven by a larger drawing effect in the control group, $t(34) = 8.94$, $SE = .11$, $d = 1.52$, $p < .001$ (WSR $Z = 4.93$, $p < .001$, PSS $Z = 4.97$, $p < .001$), than in the speeded group, $t(34) = 2.47$, $SE = .06$, $d = .42$, $p < .025$ (.05/2; WSR $Z = 2.24$, $p < .05$, PSS $Z = 1.95$, $p = .052$).

Response time. RTs for hits only were included in the analysis. RTs were analyzed in a 2 × 2 mixed measures ANOVA with group (speeded, control) as a between-subjects factor and trial type (draw, write) as a within-subjects factor. There was a main effect of group, $F(1, 68) = 139.34$, $MSE = 21099.38$, $p < .001$, $\eta^2 = .67$, as well as a significant interaction, $F(1, 68) = 13.95$, $MSE = 4187.58$, $p < .001$, $\eta^2 = .17$. The main effect of trial type was not significant, $F(1, 68) = 2.13$, $MSE = 4187.58$, $p = .15$, $\eta^2 = .03$. The interaction was driven by faster response times for drawn than written items in the control group, $t(34) = 2.68$, $SE = 21.18$, $d = 0.47$, $p < .025$ (0.05/2) (WSR $Z = 2.31$, $p < .05$, PSS $Z = 2.57$, $p < .05$), and slower response times for drawn than written items in the speeded group, $t(34) = 4.56$, $SE = 5.46$, $d = 0.77$, $p < .001$ (WSR $Z = 3.67$, $p < .001$, PSS $Z = 3.38$, $p < .01$). RTs for false alarms were analyzed using an independent samples t test. RTs were slower in the control than the speeded group, $t(68) = 13.43$, $SE = 24.66$, $d = 3.51$, $p < .001$ (WRS $Z = 7.19$, $p < .001$).

Table 4
Mean (Standard Error in Parentheses) Hit Rate, False Alarm Rate, Sensitivity, and Response Time (RT) for Each Item Type in Experiment 4

Group and measure	Draw	Write	New
Speeded			
Hits and false alarms	.63 (.03)	.57 (.03)	.24 (.02)
Sensitivity (d')	1.09 (.12)	.95 (.11)	
Correct RT	612.83 (16.98)	587.93 (20.83)	576.89 (8.93)
Control			
Hits and false alarms	.94 (.03)	.72 (.03)	.03 (.01)
Sensitivity (d')	3.53 (.12)	2.54 (.11)	
Correct RT	861.80 (16.98)	918.60 (20.83)	908.14 (22.99)

Discussion

As predicted, when the contribution of recollection to memory was reduced through speeding memory responses (e.g., Besson et al., 2012; Yonelinas & Jacoby, 1994), the drawing effect was also substantially reduced, though it did remain statistically significant. While performance on both draw and write trial types showed decreased memory performance in the speeded, relative to the normal condition, the data indicate that the draw trial type suffered more from our imposed response-deadline. This suggests that the benefit that drawing affords memory relies largely on recollection of specific episodic details from the initial encoding experience, which likely require more time to unfold before exerting their influence on memory decisions (e.g., Besson et al., 2012; Mandler, 1980). Interestingly, there remained a relatively small benefit of drawing lingering in the speeded condition, suggesting one of two possibilities. The first is that familiarity also contributes in part to the benefit to memory that participants enjoy as a result of drawing, while the second is that even in the speeded condition, there was some early influence of recollection. In support of the latter possibility, analyses showed that responses were given to drawn items more slowly than to written items in the speeded condition, suggesting that in some cases, participants might have delayed their response just long enough for the beginnings of a recollective experience to occur. However, in light of the partial evidence for increased familiarity in drawing in Experiments 3A and 3B, a modest contribution of familiarity to this effect seems more plausible. Critically, the overall findings from Experiment 4 are consistent with the foregoing three experiments, suggesting that the benefit that drawing provides to recognition memory performance is driven primarily by the creation of a recollection-based memory trace, though there may be more minor contribution of familiarity as well.

General Discussion

Our results across these four experiments consistently indicate that drawing pictures of to-be-remembered words or images, relative to writing them out, leads to a memory trace that is dependent primarily on recollection. The general finding across all of the experiments was that drawing led to not only an increase in overall memory performance (see Figure 1), but also an increase in recollection. In Experiment 1, we showed that "the drawing effect" occurred, such that memory for words that were drawn was superior to those that were written, and that this benefit was associated with increased memory sensitivity for the correct source of the words. In Experiment 2, we showed that when the to-be-remembered stimuli were images instead of words, raw memory

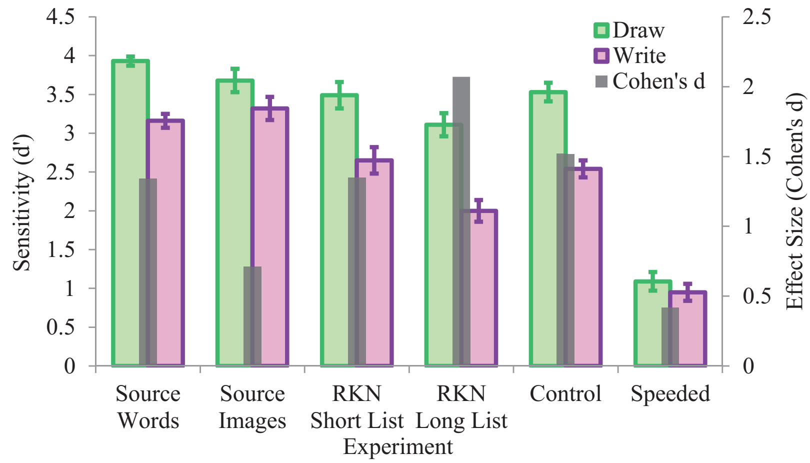

Figure 1. Sensitivity (d') in memory performance as a function of experiment and paradigm. The left-hand axis depicts mean d' for the draw (green/black bars) and write (purple/white bars) trial type in Experiments 1 (Source words), 2 (source images), 3a (remember-know/new; RKN short list), 3B (RKN long list), and 4 (control group and speeded group). The right-hand axis depicts effect size (Cohen's *d*; gray bars). Error bars represent standard error of the mean. See the online article for the color version of this figure.

performance and source memory remained better for drawn items. In addition, drawing was associated with more overall correct source identifications than writing in both words (0.96 vs. 0.77 correct) and images (0.93 vs. 0.82 correct). In Experiments 3A and 3B, we sampled recollection more comprehensively using a remember-know paradigm, and showed "the drawing effect" again, as well as demonstrating that the influence of recollection was of greater importance than familiarity in the drawing-related memory benefit. Lastly, in Experiment 4, we demonstrated that when responses were speeded in order to prevent recollection-based retrieval, there was a much smaller benefit of drawing, relative to writing. Moreover, drawn items were impacted more severely by the speeding manipulation than were written items. Taken together, these findings illustrate that not only does drawing improve memory performance in many recognition task variants, but more importantly, the findings converge on the conclusion that drawing promotes a strong, contextual memory that can be indexed by recollection measures.

In our previous work (Wammes et al., 2016) we first demonstrated a drawing effect, wherein drawn items were better remembered than written items on a later free recall test. We had predicted this on the basis of other encoding orientations which require participants to actively engage with the material at study, and also improve memory performance (production, MacLeod Gopie, Hourihan, Neary, & Ozubko, 2010; generation, Slamecka & Graf, 1978; and enactment, Engelkamp & Zimmer, 1997). In line with our predictions, and with previous work (Paivio & Csapo, 1973), we showed that drawing was a reliable encoding orientation, and in fact that it was superior to other strategies which invoked elaborative encoding, visual imagery, and pictorial encoding, respectively. Also, unlike other encoding orientations (McDaniel & Bugg, 2008), we discovered a large effect in a between-subjects design, suggesting that drawing led to a benefit that was driven by more than just being distinct relative to written items. We put forward an integrated trace view of the benefit that drawing affords memory, such that the benefit occurs because the process of drawing elicits elaborative, motoric, and pictorial encoding of the to-be-remembered information, all bound within one seamless experience (Wammes et al., 2016). In short, we were essentially proposing that because of this integrated encoding experience, there is more specific multisensory information that participants can recollect at retrieval. Previous work has suggested that better-integrated items, which perhaps includes items that were drawn at encoding, should be more accessible, because they can be recalled based on only some of their features (Graf & Mandler, 1984). It follows from this reasoning that if drawing produces a better-integrated memory, that many of the memory responses should be based upon recollection of some of the features (e.g., motoric information) from the encoding experience.

Thus, while further work will be needed to test the integration hypothesis, a necessary step in advancing understanding of this effect was determining whether drawing leads participants to develop a recollection-based retrieval of features from the encoding experience at all. In fact, our proposition of an integrated trace driving the benefit is predicated on participants' ability to better retrieve specific information about an item when it was drawn at encoding. In the context of our previous work, there was preliminary evidence to suggest that the benefit of drawing might be driven by recollection. Specifically, dual-process theorists have suggested that free recall necessarily draws primarily upon recollection (e.g., Gelbard-Sagiv, Mukamel, Harel, Malach, & Fried, 2008), and we showed a large benefit in free recall. It follows then, that drawing might improve recollection, and our four experiments in this study demonstrated this directly.

In the current work, by most measures of recollection, drawing was superior to writing. Decisions about the source of an item are thought to rely wholly on recollection, while misattributions of source are often thought to be indications of a general familiarity with the item (e.g., Migo et al., 2012; Yonelinas, 2002; Wais et al., 2010). In Experiments 1 and 2, while some individual analyses were inconclusive, the findings converge on the conclusion that drawing leads to better identification of source information than writing. Using the remember-know paradigm in Experiments 3A and 3B, not only did we show that there was a greater proportion of R responses to items that were drawn relative to written, it appeared that K responses were not diagnostic whatsoever of oldness within drawn items in either experiment (accuracy of −0.02). Furthermore, when looking at recollection-based responses as a proportion of items identified as old more broadly (R/[R + K]), draw led to proportionally more R responses in both experiments (0.94 vs. 0.75). Similarly, despite our calculations being based on only those participants who did not respond "remember" to *all* drawn items, the independence remember-know method indicated a recollection advantage for drawn items, while a substantially smaller difference existed between trial types with respect to the contribution of familiarity. Finally, when the opportunity for recollection was reduced (e.g., Bowles et al., 2007) or taken away from the participant in Experiment 4, using the response-deadline procedure, the drawing effect was relatively small in comparison. These results each suggest that recollection is critical to the benefit drawing affords memory. While each of the foregoing methods has its own shortcomings in terms of characterizing the contribution of recollection to recognition memory, together, they point compellingly to the importance of recollection in dictating the strength of this emerging effect. This is not to say that drawing is driven exclusively by recollection. Indeed, the data seem to suggest a modest contribution of familiarity as well. However, the data do indicate that recollection is the more important influence of the two.

While we interpret our data within a dual-process framework, many researchers are critical of this view. It has been proposed that rather than recollection and familiarity, recognition memory is driven by a single process, based on degrees of memory strength (Donaldson, 1996; Dunn, 2004; Hirshman & Master, 1997; Wixted & Stretch, 2004, but see Staresina, Fell, Dunn, Axmacher, & Henson, 2013). Within this model, "R" and "K" responses are simply two separate decision criteria on a single continuum or dimension: strength of evidence (which may be derived from a combination of recollection and familiarity, see Wixted & Squire, 2004). Recent results have suggested that a single-process view can account for remember–know data (Brezis, Bronfman, Yovel, & Goshen-Gottstein, 2017; Dunn, 2008; Malmberg, 2008). It is possible that in the current work, the act of drawing created a stronger memory than did writing. This, however, does not undermine the conclusion that drawing leads to a memory trace that is associated with more contextual details, as a memory that is rich in context would surely be encoded more strongly than one with sparse contextual information. However, recent work has suggested that hippocampal activation during retrieval is not associated with memory strength, but rather the amount of contextual detail that is retrieved (Rugg et al., 2012), so there is preliminary evidence that strength and amount of context are separable constructs. In any case, it is clear from our data that drawing leads to a memory trace that is associated with specific contextual information.

The drawn trial type led to such strong recognition performance in some experiments that hit rate was often near ceiling, which is especially problematic in Experiments 1 and 2 where the false alarm rate was correspondingly low, and thus influenced strongly by small shifts of false alarms across source responses. As we explore this effect in future studies, it will be important to use longer encoding lists to alleviate this issue. We have done this in Experiment 3B by doubling the number of items; however, for our other experiments which were completed prior to Experiment 3B, it was our goal to remain as similar as possible to our previous work which used free recall. While these issues do not take away from our general conclusion that the benefit of drawing over writing is driven by an increase in recollection, they do complicate any interpretations of relative effect size between Experiments 1 and 2. We recognize this issue, but maintain that the consistency between the two experiments (i.e., overall memory and source memory was at least numerically better for drawn items in both) provides a compelling case for the conclusion that drawing improves memory through the provision of more contextual information that can be used during retrieval. To clarify, if anything, the high level of performance for drawn items would mean that the difference in performance on drawn relative to written items could conceivably have been even larger had we used longer lists.

The finding that drawing improves memory through augmenting recollection, or at least through an increase in the amount of contextual information that can be used for retrieval, is consistent with our rationale and motivation for this study. First, it is concordant with the previous literature showing that encoding orientations that are related, and possibly contribute in part, to the drawing effect are driven primarily by recollection. To clarify, we argued that drawing involved elaborative, motoric, and pictorial components, which map loosely onto deep LoP, enactment, and picture superiority encoding manipulations, respectively. Akin to our findings here using drawing, elaborative strategies such as LoP manipulations and generation tend to lead to an increase in recollection rather than familiarity (Gardiner et al., 1996; Mulligan & Hirshman, 1995; Rajaram, 1993). Likewise, the enactment effect, which involves creating an item-specific motion at encoding, leads to memory retrieval that is more reliant on recollection than familiarity (Engelkamp & Dehn, 1997; Lövden et al., 2002; Manzi & Nigro, 2008). While the interaction between pictorial encoding and dual-process models is murkier than the foregoing effects, most recent research suggests that picture superiority relies more on recollection-based memory (e.g., Boldini et al., 2007; Curran & Doyle, 2011). Because we expect that the memory trace produced by drawing relies on components (i.e., elaborative, enactive, and pictorial information) that are roughly analogous to each of these three effects, it follows that drawing too should lead to a memory trace that is rich with contextual information, and thus primarily recollection-based. Our data here are consistent with the literature on these encoding orientations.

Interestingly, similar to previous work on the production effect (Fawcett & Ozubko, 2016; Ozubko, Gopie, & MacLeod, 2012), we demonstrated that the within-subjects drawing effect is driven by both recollection and in part, familiarity. However, in this previous work, both processes contributed to the benefit to a similar extent, a finding which was additionally supported by estimates from a meta-analytic approach (Fawcett & Ozubko, 2016). This suggests that there may be some common underlying mechanisms in driving these effects. However, in the case of drawing, it appears that recollection was a much stronger determinant of later memory performance than was familiarity (most prominently in Experiments 3A, 3B, and 4). Accordingly, it seems that the drawing effect arises because it provides one with many different multisensory forms of contextual information that they can draw on in retrieval.

A secondary finding of this work provides further support for the consistency and reliability of the drawing effect. That is, while we did demonstrate the effect in free recall of words (Wammes et al., 2016) and definitions (Wammes, Meade, & Fernandes, 2017), we had not yet delineated whether the drawing effect also benefits recognition memory. The current work suggests a resounding "yes" (see Figure 1) across four different task variants (old/new, remember-know, and source judgments). Moreover, the effect was less pronounced when we predicted it would be, due to *both* trial types benefiting from a pictorial trace in Experiment 2, and due to setting a stringent response-deadline that was prohibitive of recollection in Experiment 4. Thus, the drawing effect is robust to changes in methodology from recall to recognition, and also seems to be relatively consistent within recognition, regardless of subtle variations in response options.

Despite the compelling case in the data for drawing being largely dependent on recollection, our results suggest that familiarity also plays a role in improving memory following this encoding strategy. That is, the difference in familiarity estimates based on the independence remember-know method between drawn and written items was numerically larger in traditional analyses in Experiments 3A and 3B, and the corresponding multilevel logistic analyses (see Appendix) suggested familiarity was playing a small but significant role. Furthermore, there remained some sign of an effect of drawing in the speeded condition in Experiment 4. Though these differences indicate that familiarity plays a role, one must bear in mind that the differences are both quite small. Thus, even though familiarity does play a role in driving the drawing effect, it is a much smaller and less influential one than that of recollection.

Future work will be needed not only to further characterize how the act of drawing allows for such stable benefits to memory, but also to determine the potential implications of the finding in terms of its utility beyond the laboratory. Perhaps most importantly, there is still work to be done to determine the mechanistic underpinnings of the effect: Is drawing leading to integration of multisensory processing components within a cohesive memory trace, or is it simply increasing overall memory strength? How important are the individual components of the drawing process, and how much time spent drawing is enough to gain a benefit? Is drawing simply a visual instantiation of the production effect, and if so, does it behave in the same way as production in various settings? These are all important questions that can be, and in some cases already are targets for future work. There are also practical implications to explore. Specifically, it will be important to understand how long lasting and durable the benefits of drawing are, in order for drawing to have any veridical impact on learning. We have already begun to explore whether creating pictorial representations can lead to better memory for textbook definitions (Wammes et al., 2017), but it will also be important to understand whether students can fruitfully incorporate drawing into their note-taking and/or studying behaviors.

Conclusions

Across four experiments utilizing variants of recognition tasks we showed that drawing is a strategy that leads to a contextually rich recollection-based memory trace relative to writing. This benefit to recollection contributes to an overall benefit to memory for drawn relative to written items, which is remarkably consistent across paradigms (see Figure 1). Specifically, we showed that drawing leads to better overall memory, and increased recollection of the encoding source of words (Experiment 1), which was reduced when the encoding stimuli were pictures (Experiment 2). Drawing also leads to better overall memory, and more reported recollection in a remember-know paradigm testing word recognition (Experiment 3A and 3B). Finally, drawing leads to an overall memory benefit in old/new recognition, but this benefit is substantially reduced, if not eliminated when time pressure offsets recollection in a response-deadline paradigm (Experiment 4). Thus, these findings converge on the point that not only is drawing a beneficial strategy for later recognition memory, but also that drawing better enables a recollection-based memory trace of the encoding experience, which contains specific information about what task was performed with the to-be-remembered information at encoding, as well as specific details about the encoding experience.

References

Addante, R. J., Ranganath, C., & Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: Evidence of accurate source recognition without recollection. *NeuroImage, 62,* 439–450. http://dx.doi.org/10.1016/j.neuroimage.2012.04.031

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583–612). New York, NY: Academic Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (in press). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*.

Benjamin, A. S., & Craik, F. I. (2001). Parallel effects of aging and time pressure on memory for source: Evidence from the spacing effect. *Memory & Cognition, 29,* 691–697. http://dx.doi.org/10.3758/BF03200471

Besson, G., Ceccaldi, M., Didic, M., & Barbeau, E. J. (2012). The speed of visual recognition memory. *Visual Cognition, 20,* 1131–1152. http://dx.doi.org/10.1080/13506285.2012.724034

Besson, G., Ceccaldi, M., Tramoni, E., Felician, O., Didic, M., & Barbeau, E. J. (2015). Fast, but not slow, familiarity is preserved in patients with amnestic mild cognitive impairment. *Cortex, 65,* 36–49. http://dx.doi.org/10.1016/j.cortex.2014.10.020

Boldini, A., Russo, R., & Avons, S. E. (2004). One process is not enough! A speed-accuracy tradeoff study of recognition memory. *Psychonomic Bulletin & Review, 11,* 353–361. http://dx.doi.org/10.3758/BF03196582

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 & Cognition, 35,* 113–123. http://dx.doi.org/10.3758/BF03195948

Bowles, B., Crupi, C., Mirsattari, S. M., Pigott, S. E., Parrent, A. G., Pruessner, J. C., . . . Köhler, S. (2007). Impaired familiarity with preserved recollection after anterior temporal-lobe resection that spares the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 16382–16387. http://dx.doi.org/10.1073/pnas.0705273104

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,* 14325–14329. http://dx.doi.org/10.1073/pnas.0803390105

Brezis, N., Bronfman, Z. Z., Yovel, G., & Goshen-Gottstein, Y. (2017). The electrophysiological signature of remember–know is confounded with memory strength and cannot be interpreted as evidence for dual-process theory of recognition. *Journal of Cognitive Neuroscience, 29,* 322–336. http://dx.doi.org/10.1162/jocn_a_01053

Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience, 2,* 51–61. http://dx.doi.org/10.1038/35049064

Curran, T., & Doyle, J. (2011). Picture superiority doubly dissociates the ERP correlates of recollection and familiarity. *Journal of Cognitive Neuroscience, 23,* 1247–1262. http://dx.doi.org/10.1162/jocn.2010.21464

Defeyter, M. A., Russo, R., & McPartlin, P. L. (2009). The picture superiority effect in recognition memory: A developmental study using the response signal procedure. *Cognitive Development, 24,* 265–273. http://dx.doi.org/10.1016/j.cogdev.2009.05.002

Dewhurst, S. A., & Conway, M. A. (1994). Pictures, images, and recollective experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1088–1098. http://dx.doi.org/10.1037/0278-7393.20.5.1088

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59,* 447–456. http://dx.doi.org/10.1016/j.jml.2007.11.004

Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: Multiple prefrontal processes subserve source memory. *Neuron, 35,* 989–996. http://dx.doi.org/10.1016/S0896-6273(02)00858-9

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523–533. http://dx.doi.org/10.3758/BF03200940

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111,* 524–542. http://dx.doi.org/10.1037/0033-295X.111.2.524

Dunn, J. C. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review, 115,* 426–446. http://dx.doi.org/10.1037/0033-295X.115.2.426

Duzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H. J., & Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences of the United States of America, 94,* 5973–5978. http://dx.doi.org/10.1073/pnas.94.11.5973

Engelkamp, J., & Dehn, D. M. (1997). Strategy and consciousness in remembering subject-performed actions. *Sprache & Kognition, 16,* 94–109.

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica, 96,* 43–60. http://dx.doi.org/10.1016/S0001-6918(97)00005-X

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology, 70,* 99–115. http://dx.doi.org/10.1037/cep0000089

Forrin, N. D., Groot, B., & MacLeod, C. M. (2016). The d-Prime directive: Assessing costs and benefits in recognition by dissociating mixed-list false alarm rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1090–1111. http://dx.doi.org/10.1037/xlm0000214

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313. http://dx.doi.org/10.3758/BF03197041

Gardiner, J. M. (2001). Episodic memory and autonoetic consciousness: A first-person approach. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences, 356,* 1351–1361. http://dx.doi.org/10.1098/rstb.2001.0955

Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology, 50,* 114–122. http://dx.doi.org/10.1037/1196-1961.50.1.114

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition, 18,* 579–583. http://dx.doi.org/10.3758/BF03197100

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1999). Response deadline and subjective awareness in recognition memory. *Consciousness and Cognition, 8,* 484–496. http://dx.doi.org/10.1006/ccog.1999.0409

Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. *Science, 322,* 96–101. http://dx.doi.org/10.1126/science.1164685

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23,* 553–568. http://dx.doi.org/10.1016/S0022-5371(84)90346-3

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, 27,* 46–51. http://dx.doi.org/10.3758/BF03203619

Hayama, H. R., Vilberg, K. L., & Rugg, M. D. (2012). Overlap between the neural correlates of cued recall and source memory: Evidence for a generic recollection network? *Journal of Cognitive Neuroscience, 24,* 1127–1137. http://dx.doi.org/10.1162/jocn_a_00202

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition, 25,* 345–351. http://dx.doi.org/10.3758/BF03211290

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340. http://dx.doi.org/10.1037/0096-3445.110.3.306

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13–47). Mahwah, NJ: Erlbaum, Inc.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007

Jones, G. V. (1987). Independence and exclusivity among psychological processes: Implications for the structure of recall. *Psychological Review, 94,* 229–235. http://dx.doi.org/10.1037/0033-295X.94.2.229

Joordens, S., & Merikle, P. M. (1993). Independence or redundancy? Two models of conscious and unconscious influences. *Journal of Experimental Psychology: General, 122,* 462–467. http://dx.doi.org/10.1037/0096-3445.122.4.462

Juola, J. F., Fischler, I., Wood, C. T., & Atkinson, R. C. (1971). Recognition time for information stored in long-term memory. *Perception & Psychophysics, 10,* 8–14. http://dx.doi.org/10.3758/BF03205757

Kelley, C. M., Jacoby, L. L., & Hollingshead, A. (1989). Direct versus indirect tests of memory for source: Judgments of modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1101–1108. http://dx.doi.org/10.1037/0278-7393.15.6.1101

Knowlton, B. J., & Squire, L. R. (1995). Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 699–710. http://dx.doi.org/10.1037/0278-7393.21.3.699

Koen, J. D., & Yonelinas, A. P. (2010). Memory variability is due to the contribution of recollection and familiarity, not to encoding variability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1536–1542. http://dx.doi.org/10.1037/a0020448

Leiker, E. K., & Johnson, J. D. (2015). Pattern reactivation co-varies with activity in the core recollection network during source memory. *Neuro-*

psychologia, 75, 88–98. http://dx.doi.org/10.1016/j.neuropsychologia.2015.05.021

Lövdén, M., Rönnlund, M., & Nilsson, L. G. (2002). Remembering and knowing in adulthood: Effects of enacted encoding and relations to processing speed. *Neuropsychology, Development, and Cognition Section B, Aging, Neuropsychology and Cognition, 9,* 184–200. http://dx.doi.org/10.1076/anec.9.3.184.9612

Luo, L., Hendriks, T., & Craik, F. I. (2007). Age differences in recollection: Three patterns of enhanced encoding. *Psychology and Aging, 22,* 269–280. http://dx.doi.org/10.1037/0882-7974.22.2.269

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

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57,* 335–384. http://dx.doi.org/10.1016/j.cogpsych.2008.02.004

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271. http://dx.doi.org/10.1037/0033-295X.87.3.252

Manzi, A., & Nigro, G. (2008). Long-term memory for performed and observed actions: Retrieval awareness and source monitoring. *Memory, 16,* 595–603. http://dx.doi.org/10.1080/09658210802070749

Mazzoni, D., & Dannenberg, R. (2000). Audacity [software]. Pittsburg, PA: Carnegie Mellon University.

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

Mecklinger, A. (2000). Interfacing mind and brain: A neurocognitive model of recognition memory. *Psychophysiology, 37,* 565–582. http://dx.doi.org/10.1111/1469-8986.3750565

Meng, Y., Ye, X., & Gonsalves, B. D. (2014). Neural processing of recollection, familiarity and priming at encoding: Evidence from a forced-choice recognition paradigm. *Brain Research, 1585,* 72–82. http://dx.doi.org/10.1016/j.brainres.2014.08.024

Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember-know procedure. *Consciousness and Cognition, 21,* 1435–1455. http://dx.doi.org/10.1016/j.concog.2012.04.014

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125. http://dx.doi.org/10.1037/1082-989X.7.1.105

Mulligan, N., & Hirshman, E. (1995). Speed-accuracy trade-offs and the dual process model of recognition memory. *Journal of Memory and Language, 34,* 1–18. http://dx.doi.org/10.1006/jmla.1995.1001

Nelson, D. L., Schreiber, T. A., & McEvoy, C. L. (1992). Processing implicit and explicit representations. *Psychological Review, 99,* 322–348. http://dx.doi.org/10.1037/0033-295X.99.2.322

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition, 40,* 326–338. http://dx.doi.org/10.3758/s13421-011-0165-1

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

Perfect, T. J., Mayes, A. R., Downes, J. J., & Van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *The Quarterly Journal of Experimental Psychology, 49,* 797–813. http://dx.doi.org/10.1080/713755644

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21,* 89–102. http://dx.doi.org/10.3758/BF03211168

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 365–377. http://dx.doi.org/10.1037/0278-7393.22.2.365

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia, 42,* 2–13. http://dx.doi.org/10.1016/j.neuropsychologia.2003.07.006

Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember-know paradigm. *Psychonomic Bulletin & Review, 15,* 825–832. http://dx.doi.org/10.3758/PBR.15.4.825

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11,* 251–257. http://dx.doi.org/10.1016/j.tics.2007.04.004

Rugg, M. D., Vilberg, K. L., Mattson, J. T., Yu, S. S., Johnson, J. D., & Suzuki, M. (2012). Item memory, context memory and the hippocampus: FMRI evidence. *Neuropsychologia, 50,* 3070–3079. http://dx.doi.org/10.1016/j.neuropsychologia.2012.06.004

Sauvage, M. M., Beer, Z., & Eichenbaum, H. (2010). Recognition memory: Adding a response deadline eliminates recollection but spares familiarity. *Learning & Memory, 17,* 104–108. http://dx.doi.org/10.1101/lm.1647710

Senkfor, A. J., & Van Petten, C. (1998). Who said what? An event-related potential investigation of source and item memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1005–1025. http://dx.doi.org/10.1037/0278-7393.24.4.1005

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

Staresina, B. P., Fell, J., Do Lam, A. T., Axmacher, N., & Henson, R. N. (2012). Memory signals are temporally dissociated in and across human hippocampus and perirhinal cortex. *Nature Neuroscience, 15,* 1167–1173. http://dx.doi.org/10.1038/nn.3154

Staresina, B. P., Fell, J., Dunn, J. C., Axmacher, N., & Henson, R. N. (2013). Using state-trace analysis to dissociate the functions of the human hippocampus and perirhinal cortex in recognition memory. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 3119–3124. http://dx.doi.org/10.1073/pnas.1215710110

Starns, J. J., & Ratcliff, R. (2008). Two dimensions are not better than one: STREAK and the univariate signal detection model of remember-know performance. *Journal of Memory and Language, 59,* 169–182. http://dx.doi.org/10.1016/j.jml.2008.04.003

Toth, J. P. (1996). Conceptual automaticity in recognition memory: Levels-of-processing effects on familiarity. *Canadian Journal of Experimental Psychology, 50,* 123–138. http://dx.doi.org/10.1037/1196-1961.50.1.123

Tulving, E. (1983). *Elements of episodic memory*. Oxford, UK: Clarendon.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12. http://dx.doi.org/10.1037/h0080017

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11,* 192–196. http://dx.doi.org/10.3758/BF03206482

Wagner, A. D., Gabrieli, J. D., & Verfaellie, M. (1997). Dissociations between familiarity processes in explicit recognition and implicit perceptual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 305–323. http://dx.doi.org/10.1037/0278-7393.23.2.305

Wais, P. E., Mickes, L., & Wixted, J. T. (2008). Remember-know judgments probe degrees of recollection. *Journal of Cognitive Neuroscience, 20,* 400–405. http://dx.doi.org/10.1162/jocn.2008.20041

Wais, P. E., Squire, L. R., & Wixted, J. T. (2010). In search of recollection and familiarity signals in the hippocampus. *Journal of Cognitive Neuroscience, 22,* 109–123. http://dx.doi.org/10.1162/jocn.2009.21190

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: Human Experimental*

Psychology, 69, 1752–1776. http://dx.doi.org/10.1080/17470218.2015.1094494

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

Wilding, E. L., & Ranganath, C. (2011). Electrophysiological correlates of episodic memory processes. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of ERP components* (pp. 373–396). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780195374148.013.0187

Wixted, J. T., & Squire, L. R. (2004). Recall and recognition are equally impaired in patients with selective hippocampal damage. *Cognitive, Affective & Behavioral Neuroscience, 4,* 58–66. http://dx.doi.org/10.3758/CABN.4.1.58

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember-know judgments. *Psychonomic Bulletin & Review, 11,* 616–641. http://dx.doi.org/10.3758/BF03196616

Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Research, 1100,* 125–135. http://dx.doi.org/10.1016/j.brainres.2006.05.019

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354. http://dx.doi.org/10.1037/0278-7393.20.6.1341

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1415–1434. http://dx.doi.org/10.1037/0278-7393.25.6.1415

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517. http://dx.doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology, 48,* 516–535. http://dx.doi.org/10.1037/1196-1961.48.4.516

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34,* 622–643. http://dx.doi.org/10.1006/jmla.1995.1028

Yonelinas, A. P., & Jacoby, L. L. (1996). Noncriterial recollection: Familiarity as automatic, irrelevant recollection. *Consciousness and Cognition, 5,* 131–141. http://dx.doi.org/10.1006/ccog.1996.0008

Yonelinas, A. P., Kroll, N. E., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12,* 323–339. http://dx.doi.org/10.1037/0894-4105.12.3.323

Appendix

Multilevel logistic regression analyses

Experiment 1

Statistical approach. Here we present an alternate analysis approach that is well suited to our dataset. Simulations have suggested that for proportional data, or indeed any data with a binary outcome, multilevel logistic regression may be a superior and less biased technique for analysis than standard analyses of variance (Dixon, 2008; Jaeger, 2008). Specifically, complications arise because proportional data, especially when there are relatively few trials, are likely to be non-normal, constrained in range and thus, noncontinuous and heteroscedastic. Logistic regression remedies these issues, and has already been implemented in studies of the production effect in memory (Fawcett & Ozubko, 2016). In essence, multilevel logistic regression analyses use the raw outcome of each trial as the dependent variable, and model the contributions of specified fixed and random factors. Because there are multiple observations (trials) per participant, subject is typically incorporated as a random effect to account for this dependency.

Data were analyzed using the lme4 package (Bates, Mächler, Bolker, & Walker, in press) in the R environment. Across all experiments, we modeled the likelihood of a particular dichotomous outcome (i.e. items called old vs. not, items identified as drawn vs. not) as a function of whether each item was, in actuality drawn and written, using lures as the relevant intercept. To accomplish this, our conditions were dummy coded as 0 or 1 within their own separate variables. For instance, in the "drawn" variable, drawn items are coded as 1, while all other items are coded as 0. Similarly, in the "written" variable, written items are coded as 1 while all other items are coded as 0. As a byproduct of this structure, lures are coded as 0 within both variables and thus, act as the baseline or intercept for the model. The beta coefficients produced by such a logistic regression deal with the model prediction of the odds (in logit space) of a factor relative to the intercept. For example, in these analyses, the beta coefficient for the "drawn" variable captures the magnitude by which the relative logit odds of an item being called old ($ln\frac{p}{(p-1)}$, where p is the probability of an item being called old) increases as a function of an item being drawn.

Old responses. For our first model, responses of "drawn" and "written" were collapsed into "old" responses. The outcome variable was whether or not an item was called old, and multilevel logistic regression with trial type (lure, draw, write) as a fixed effect was used to estimate three fixed effect coefficients. The intercept coefficient is the logit odds that a lure would be identified as old, and the slope coefficients for drawn and written are the contrasts between these trial types and lures. The estimate for the intercept was -4.14 (CI_{95} [-4.61, -3.70]), indicating not surprisingly, that lures are very unlikely to be characterized as old. The slope for drawn items was 8.36 (CI_{95} [7.68, 9.12]), while the slope for written items was 5.88 (CI_{95} [5.43, 6.38]). As is apparent by the nonoverlapping confidence intervals, the estimate for drawn items is significantly larger than for written, $Z = 5.63$, $p < .001$. Furthermore, this model was compared with a null model, which assumed no effect of trial type (i.e. items were only coded as old or new). For each model, Akaike weights (*AICw*; Wagenmakers & Farrell, 2004) were computed based on the likelihood estimates. These weights can be interpreted as the probability that a particular model fits the data better than all other models. The model including trial type ($AICw > 0.999$) fit the data better than the null model, $\chi^2(1) = 108.16$, $p < .001$, reinforcing that there was a main effect of trial type.

Source decisions. Next we analyzed the likelihood that an item would be identified as "drawn." To reflect the manner in which source memory data are typically analyzed (i.e. the proportion of correct old responses that specified the correct source), only trials wherein an item was identified as old were included in the model. Thus, in this analysis, the intercept for the model represents the logistic odds that a lure that was given an old response (i.e. "drawn" or "written") was identified as drawn. The coefficient for written items estimates the likelihood, in relation to that intercept, that a written item was identified as drawn. Similarly, the coefficient for drawn items estimates the relative likelihood that a drawn item was identified as drawn. The estimate for the intercept was -0.75 (CI_{95} [-2.06, 0.46]), the slope for drawn items was 5.05 (CI_{95} [3.79, 6.45]), and the slope for written items was -2.52 (CI_{95} [-3.79, -1.19]). Because a comparable analysis predicting the likelihood of an item being identified as written would necessarily be the inverse of this model, the absolute values of the draw and write coefficients can be directly compared. The contrast between drawn and written items indicated that drawn items led to more successful source memory, $Z = 2.70$, $p < .01$. This model was also compared with a null model, which assumed no effect of trial type (i.e. items were only coded as old or new), as well as models including only the fixed effect for draw, or only the fixed effect for write. The model including both trial types ($AICw = 0.993$) fit the data best, and the fit was significantly better than the next best model, $\chi^2(1) = 62.29$, $p < .001$.

(*Appendix continues*)

These data can also be analyzed including *all* items, regardless of whether they received an old or new response. Such an analysis tests whether overall, drawn items were more likely than written items to be identified as their correct source (as opposed to the incorrect source or a new response). However, because all items are included, an analysis predicting whether an item was called drawn is no longer the inverse of a similar analysis predicting "written" responses, so both models were necessary. When the outcome variable was "drawn" responses, the intercept was −6.46 (CI_{95} [−7.58, −5.47]), the slope for drawn items was 9.91 (CI_{95} [8.89, 11.15]), and the slope for written items was 3.02 (CI_{95} [2.13, 4.13]). When analyzing for the likelihood of being called written, the intercept was −5.04 (CI_{95} [−5.73, −4.45]), the slope for drawn items was 0.86 (CI_{95} [0.08, 1.66]), and the slope for written items was 6.41 (CI_{95} [5.83, 7.09]). The likelihood of a drawn item being called drawn was significantly greater than the likelihood of a written item being called written, $Z = 5.33$, $p < .001$. The results of these analyses were largely upheld in standard GLM analyses, though the difference between draw and write in the proportion of hits that were source-correct was only marginal.

Experiment 2

Old responses. As in Experiment 1, responses of drawn and written were collapsed into "old" responses. The estimate for the intercept was −3.38 (CI_{95} [−3.71, −3.07]), and the slope for drawn items (6.33, CI_{95} [5.88, 6.82]), was significantly greater than the slope for written items (5.43, CI_{95} [5.04, 5.84]), $Z = 2.88$, $p < .01$. This model ($AICw > 0.999$) was also superior to the null model, $\chi^2(1) = 19.85$, $p < .001$.

Source decisions. The estimate for the intercept was 0.57 (CI_{95} [−0.02, 1.18]), the slope for drawn items was 3.24 (CI_{95} [2.45, 4.06]), and the slope for written items was −3.15 (CI_{95} [−3.85, −2.50]). The contrast between drawn and written items however, was not significant, $z = 0.16$, $p = .87$. Importantly, this model was also compared with a null model, which assumed no effect of trial type (i.e. items were only coded as old or new), as well as models including only the fixed effect for draw, or only the fixed effect for write. The model including both trial types ($AICw > 0.999$) fit the data best, and the fit was significantly better than the next best model, $\chi^2(1) = 55.61$, $p < .001$. It is important to clarify why there was no significant difference between the draw and write trial types in this instance. Because only items which received an old response are included in the model, the baseline or intercept is the relative odds that a false alarm was called drawn. In this case, 64% of false alarms were given a drawn response, meaning that the proportion of source correct responses to drawn items is compared with this baseline, while written items are compared with the inverse baseline (36%). When there are small numbers of false alarms, as is the case in these experiments, differences can be amplified. To clarify, if just five false alarms *total* across 35 participants were shifted to "write" responses, the estimates would be significantly different from one another.

When analyzing *all* items for the likelihood of being called drawn, the intercept was −3.79 (CI_{95} [−4.18, −3.45]), the slope for drawn items was 6.35 (CI_{95} [5.90, 6.83]), and the slope for written items was 1.07 (CI_{95} [0.60, 1.56]). When analyzing for the likelihood of being called written, the intercept was −4.43 (CI_{95} [−4.97, −3.96]), the slope for drawn items was 0.51 (CI_{95} [−0.22, 1.22]), and the slope for written items was 5.98 (CI_{95} [5.48, 6.55]). The coefficients for each item type were not significantly different from one another $z = 1.00$, $p = .32$.

These analyses are inconsistent with the findings from paired-samples *t*-tests, and nonparametric comparisons, which indicated that in hit rate, proportion of hits that were source-correct, proportion of studied items that were source-correct, and d' computed in two different ways, source decisions were more accurate for drawn items relative to written items. This is because each of the coefficients is in contrast to the lure distribution, which was biased toward drawn items. This suggests that perhaps, when participants believed that they remembered an item, but could not ascertain the source, they were more likely to guess that the item was drawn. However, on average, a participant only committed 1.38 false alarms, 0.88 of which were drawn responses. Because of the small number of false alarms, the small shift was much more pronounced, and becomes difficult to interpret.

Experiment 3A

Old responses. Responses of "remember" and "know" were collapsed into "old" responses. The estimate for the intercept was −2.25 (CI_{95} [−2.46, −2.04]), and the slope for drawn items (5.87, CI_{95} [5.40, 6.40]), was significantly greater than the slope for written items was (3.66, CI_{95} [3.40, 3.93]), $Z = 7.68$, $p < .001$. This model ($AICw > 0.999$) was also superior to the null model, $\chi^2(1) = 117.01$, $p < .001$.

Remember responses. The estimate for the intercept was −3.76 (CI_{95} [−4.12, −3.42]), and the slope for drawn items (6.09, CI_{95} [5.66, 6.54]), was significantly larger than that for written items (4.17, CI_{95} [3.80, 4.57]), $Z = 6.49$, $p < .001$. This model ($AICw > 0.999$) was also superior to the null model, $\chi^2(1) = 194.15$, $p < .001$.

Know responses. When analyzing the likelihood of giving a K response to items, the estimate for the intercept was −2.54 (CI_{95} [−2.78, −2.32]), and the slope for drawn items was −0.23 (CI_{95} [−0.61, 0.13]), while the slope for written items was 1.11 (CI_{95} [0.84, 1.39]). Interestingly, these coefficients indicated that participants were *more* likely to give a K response to written than drawn items, $z = 5.78$, $p < .01$. This model ($AICw > 0.999$) was also superior to the null model, $\chi^2(1) = 61.41$, $p < .001$.

(*Appendix continues*)

In line with the independence remember-know method (e.g. Yonelinas et al., 1998), we also tested a model wherein only responses that were *not* classified as remember were included. In this analysis, the estimate for the intercept was −2.54 (CI_{95} [−2.81, −2.30]), the slope for drawn items was 3.40 (CI_{95} [2.82, 4.02]), while the slope for written items was 2.51 (CI_{95} [2.19, 2.83]). The estimate for drawn items is significantly larger than for written, $Z = 2.59$, $p < .01$, and this model ($AICw > 0.969$) was superior to the null model, $\chi^2(1) = 8.88$, $p < .01$.

Experiment 3B

Old responses. The estimate for the intercept was −2.15 (CI_{95} [−2.45, −1.86]), and the slope for drawn items (5.63, CI_{95} [5.32, 5.95]) was significantly larger than the slope for written items (3.26, CI_{95} [3.07, 3.46]), $z = 12.50$ $p < .001$. The significance of trial type as a factor was reinforced, as model comparisons indicated that this model ($AICw > 0.999$) fit the data better than the null model, $\chi^2(1) = 359.71$, $p < .001$.

Remember responses. The estimate for the intercept was −3.25 (CI_{95} [−3.58, −2.93]), and the slope for drawn items (5.66, CI_{95} [5.38, 5.96]), was significantly larger than the slope for written items (3.48, CI_{95} [3.24, 3.72]), z = 11.41, $p < .001$. In addition, this model ($AICw > 0.999$) fit the data significantly better than the null model, $\chi^2(1) = 518.03$, $p < .001$.

Know responses. When modeling the likelihood of giving a K response in *all* items, the estimate for the intercept was −2.81 (CI_{95} [−3.16, −2.47]), and the slope for drawn items (−0.25, CI_{95} [−0.52, 0.02]), was significantly *smaller* than the slope for written items (1.02, CI_{95} [1.89, 2.38]), $z = 11.41$, $p < .001$. In addition, this model ($AICw > 0.999$) fit the data significantly better than the null model, $\chi^2(1) = 102.66$, $p < .001$.

However, the raw model likely underestimates the contribution of familiarity. When only including non-R responses, the estimate for the intercept was −2.71 (CI_{95} [−3.09, −2.36)], the slope for drawn items was 2.88 (CI_{95} [2.46, 3.31]), while the slope for written items was 2.13 (CI_{95} [1.89, 2.38]). The estimate for drawn items is significantly larger than for written, $z = 2.98$, $p < .01$, and the model ($AICw = 0.993$) fit the data significantly better than the null model, $\chi^2(1) = 11.83$, $p < .001$.

Experiment 4

Old responses. In this analysis, drawn or written items were dummy coded as in previous experiments. The speeded condition was coded as −0.5 and the normal condition as 0.5. In doing so, the conditions were mean centered. Accordingly, the intercept represents a pooled false alarm rate, while each of the estimated slopes represents a main effect of the fixed factor it applies to. Interaction terms were also included, which represent the magnitude by which the "trial type" slopes are incremented as a result of the change in conditions. In this model, the estimate for the intercept was −3.17 (CI_{95} [−3.48, −2.87]), while the slope for condition was 2.32 (CI_{95} [1.72, 2.93]). The slopes for drawn and written items were 5.04 (CI_{95} [4.74, 5.36]) and 3.83 (CI_{95} [3.58, 4.09]), respectively. In essence, these coefficients indicate a main effect of condition, such that performance was worse in the fast condition, and a main effect of trial type, such that drawn items were better identified than write, $Z = 5.90$, $p < .001$. This finding was corroborated as this model ($AICw > 0.999$) fit the data significantly better than a null model which was agnostic to condition, $\chi^2(3) = 395.90$, $p < .001$, as well as a null model that was agnostic to trial type, $\chi^2(2) = 156.71$, $p < .001$.

However, condition also interacted with both drawn (4.90, CI_{95} [4.31, 5.53]) and written items (3.05, CI_{95} [2.56, 3.58]), such that both trial types were negatively impacted by speeding. The interaction with drawn items was greater, $z = 4.45$, $p < .001$, indicating that drawn items are impacted more by the speeded condition than are written items. Moreover, a direct contrast of models including either of the interactions indicated that the model including the interaction with drawn items ($AICw > 0.999$) fit the data best. Because the models do not differ in the number of parameters (i.e. $df = 0$), the typical statistical test for model contrasts could not be performed, however the raw difference in log likelihood was 196.80, and the weighted AIC shows clearly that the model including the drawn interaction is most likely to fit the data. To make this clearer, we modeled the likelihood of making an old response within each condition separately.

Analysis by condition. For the speeded condition, the intercept was −2.00 (CI_{95} [−2.32, −1.69]), the slope for drawn items was 2.58 (CI_{95} [2.35, 2.81]), and the slope for written items was 2.29 (CI_{95} [2.07, 2.52]). Draw and write were not significantly different from one another, $Z = 1.75$, $p = .08$. While this model *was* significantly better than the null model, $\chi^2(1) = 6.40$, $p < .05$ the weighted AIC was only 0.900, meaning that there is still a 10% probability that the null model is the best one. These analyses suggest that there remains a small lingering drawing effect in the speeded condition.

In the normal condition, the intercept was −4.35 (CI_{95} [−4.90, −3.84]), the slope for drawn items was 7.54 (CI_{95} [6.98, 8.14]), and the slope for written items was 5.38 (CI_{95} [4.94, 5.87]). Draw was significantly superior to write, $z = 5.64$, $p < .001$, which was also reflected in the significant difference between this model ($AICw > 0.999$) and the null model, $\chi^2(1) = 151.03$, $p < .001$. The traditional GLM analyses reported in the main article were entirely consistent with these findings (see Table 4 for mean signal detection measures).

Received August 2, 2016
Revision received April 27, 2017
Accepted May 21, 2017 ■