Psychonomic Bulletin & Review (2021) 28:259–267
https://doi.org/10.3758/s13423-020-01804-w

BRIEF REPORT

Drawing and memory: Using visual production to alleviate concreteness effects

Brady R.T. Roberts[1] · Jeffrey D. Wammes[2]

Accepted: 26 August 2020 / Published online: 16 September 2020

Abstract

Countless experiments have been devoted to understanding techniques through which memory might be improved. Many strategies uncovered in the literature are thought to act via the integration of contextual information from multiple distinct codes. However, the mnemonic benefits of these strategies often do not remain when there is no clear link between a word and its multisensory referent (e.g., in abstract words). To test the importance of this link, we asked participants to encode target words (ranging from concrete to abstract) either by drawing them, an encoding strategy recently proven to be reliable in improving memory, or writing them. Drawing provides a compelling test case because while other strategies (e.g., production, generation) shift focus to *existing* aspects of to-be-remembered information, drawing may forge a link with *novel* multisensory information, circumventing shortcomings of other memory techniques. Results indicated that while drawing's benefit was slightly larger for concrete stimuli, the effect was present across the spectrum from abstract to concrete. These findings demonstrate that even for highly abstract concepts without a clear link to a visual referent, memory is reliably improved through drawing. An exploratory analysis using a deep convolutional neural network also provided preliminary evidence that in abstract words, drawings that were most distinctive were more likely to be remembered, whereas concrete items benefited from prototypicality. Together, these results indicate that while the advantageous effects of drawing exist across all levels of concreteness, the memory benefit is larger when words are concrete, suggesting a tight coupling between the drawing benefit and visual code.

Keywords Drawing effect · Concreteness · Abstract · Memory · Neural network

Introduction

As we navigate the world before us, we are bombarded with incoming information. There is considerable variability in the vehicle of presentation (i.e., multiple sensory modes), the context in which we encounter the information, and the strategies we might use to encode information into memory. In investigating the circumstances that precipitate robust (or alternatively, weak) memories, intrinsic stimulus qualities and encoding strategies are both highly relevant factors. However, these factors are frequently studied in isolation from one another. Understanding the interactions between the two can enrich our underlying theories of fundamental memory processes and inform how we can better learn in our everyday lives. The current investigation sought to provide insight into how one particularly effective encoding strategy – drawing – interacts with an intrinsic stimulus quality known to improve memory: concreteness. This particular pairing of encoding strategy and stimulus is especially compelling because concreteness effects are often conceptualized as being driven by inherent links between a verbal concept and its real-world visual referent (Paivio, Rogers, & Smythe, 1968), while drawing is thought to boost memory performance via the manufacturing or fortifying of a link between a word and its visual referent (Wammes, Meade, & Fernandes, 2016).

Recent work exploring how encoding strategies bear on later memory performance has demonstrated improved

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13423-020-01804-w) contains supplementary material, which is available to authorized users.

✉ Brady R.T. Roberts
brady.roberts@uwaterloo.ca

Jeffrey D. Wammes
j.wammes@yale.edu

[1] Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada

[2] Department of Psychology, Yale University, New Haven, Connecticut, USA

memory for information that was drawn during learning (e.g., Paivio & Csapo, 1973; Van Meter & Garner, 2005; Wammes et al., 2016). This "drawing effect" is robust to a wide array of perturbations to experimental parameters, including changes in encoding time, list length, comparison task (Wammes et al., 2016), test format, stimulus type (Blake & Castel, 2019; Fernandes, Wammes, & Meade, 2018; Wammes, Meade, & Fernandes, 2017), and population (Meade, Wammes, & Fernandes, 2018). The potency of the drawing benefit has naturally led to testing the efficacy of real-world applications, including learning definitions (Wammes et al., 2017; see Van Meter & Garner, 2005, for a review), improving conceptual knowledge (Anderson, Ellis, & Jones, 2014; Edens & Potter, 2010), and even promoting better understanding of scientific texts (Schmeck, Mayer, Opfermann, Pfeiffer, & Leutner, 2014; Van Meter, 2001; Van Meter, Aleksic, Schwartz, & Garner, 2006). When considering the variety of uses and applications for drawing in research, it is evident that this tool provides a window into various forms of cognition for which a simple button press or verbal report may be insufficient.

Recently, research has suggested that drawing improves memory via the encoding of information from multiple distinct codes – motor, generative, and visual – which together produce a rich multi-sensory memory (Wammes, Jonker, & Fernandes, 2019). Punctuating the importance of the generative code, it has been demonstrated that simply preparing to draw *without* actually performing the drawing still enhances memory relative to writing (Wammes, Roberts, & Fernandes, 2018). Most critically for the current work, this generative process likely acts by forging new connections with internal imagery and conceptual knowledge held in semantic memory, or, alternatively, capitalizing on the connections that already exist.

Even without actively engaging with conceptual knowledge through an explicit task that requires it (e.g., drawing, generation, or deep level of processing), the importance of the link between word and conceptual knowledge is clear. Specifically, it is well established that concrete words are better remembered than abstract words (e.g., Begg & Paivio, 1969; Fliessbach, Weis, Klaver, Elger, & Weber, 2006; Moeser, 1974). Two prevailing theories regarding the origin of such concreteness effects are dual-coding theory (Paivio et al., 1968) and context-availability theory (Schwanenflugel & Shoben, 1983). Dual-coding theory posits that, in the context of concreteness, a memory boost is observed for concrete words due to the relative ease with which they elicit mental imagery, ultimately providing two forms of memory coding (verbal and imagery; Clark & Paivio, 1991; Mayer & Anderson, 1991; Paivio, Walsh, & Bons, 1994). Context-availability theory on the other hand suggests that more associated verbal contextual information is available for concrete words, allowing for enhanced retrieval (Bransford & McCarrell, 1974; Kieras, 1978; Schwanenflugel, Akin, & Luh, 1992; Schwanenflugel & Stowe, 1989). So, while dual-coding theory suggests that retrieval of concrete words benefits from multiple memory codes provided at encoding, context-availability theory posits instead that concrete words simply boast an inherently larger network of related information to call upon during retrieval.

While these two prevailing theories do not bear directly on questions regarding drawing's interaction with memory, we can reasonably extend them to this area, as the benefit of drawing is likely similar in its reliance on tight links between underlying conceptual knowledge and visual representations. In line with the dual-coding theory then, the drawing effect should easily manifest in concrete words due to the manufacturing or strengthening of the link between a word and its visual referent. This drawing benefit may, however, also persist for abstract words because although they do not have a clear visual referent, comparable imagery can be artificially induced through drawing. In contrast, context-availability theory might predict a heavily attenuated drawing effect for abstract words, as drawing would have no bearing on the amount of contextual information that is *intrinsically* associated with a word (i.e., the semantic information inherently linked to it). That is, while drawing may add contextual information (e.g., hand movement, visual feedback) that was not otherwise present, it does not fundamentally alter the set of existing semantic associations that a particular concept has.

In the current work, we aimed to determine just how critical the link between verbal label and underlying concept is. Research in this arena has primarily found concreteness effects in words, images, and definitions, raising an interesting question: Will the benefit of drawing persist when a word is abstract, and therefore its link to associated conceptual knowledge or multisensory referent is unclear? We used drawing as a tool to determine if a clear link between concept and visual imagery must be inherent for it to improve memory, or if this link can be flexibly added to a similar effect via drawing. The answer to this question will serve to inform our understanding of imagery, and the ways in which verbal information interacts with multimodal processing. We predicted that, consistent with the dual-coding theory, drawing the referent of an abstract concept would yield memory performance comparable to that in undrawn (i.e., written) concrete words, by providing an otherwise absent imagery trace. Thus, we expected the drawing effect to manifest in both concrete and abstract words.

Assuming that some abstract concepts have relatively prototypical associates (e.g., a ballot for "democracy" or a smiley face for "happiness"), then perhaps drawing these more common associates would confer a further enhanced memorial benefit. That is, at the level of visual imagery, some abstract items may seem qualitatively more "concrete" than those without a common referent. Abstract items with these more prototypical referents, when drawn, may benefit similarly to their concrete counterparts from the building or reinforcing of links with further visual information. Therefore, we were interested in exploring this relation between the prototypicality of drawing content, and later memory for those drawings.

In the current study, we empirically tested the prediction that drawing would provide a consistent benefit to memory regardless of word concreteness. We also employed a deep neural network to conduct an exploratory analysis testing the prediction that drawings of abstract words that are prototypical would be more likely to be remembered.

Method

Participants

Of the recent studies that explored drawing with word-recognition retrieval tests, the smallest reported drawing effect size (memory for draw > write) was $d = 0.67$ (Wammes et al., 2016, Experiment 5). We therefore performed an a priori power analysis (matched-pairs, two-tailed t-tests with a $d = 0.67$) using G*Power software (version 3.1; Faul, Erdfelder, Lang, & Buchner, 2007), which indicated a required sample size of 31 participants to achieve 95% statistical power to find a main effect of drawing. Accordingly, we aimed to include at least 31 participants in this experiment. However, we note that this power analysis had no bearing on our ability to measure an interaction with concreteness, and so it is possible that a required sample size was larger than this. Nonetheless, we originally collected data from 34 participants. In the initial experiment, however, we had no intention of performing our exploratory analysis of prototypicality, and as such, there were only two drawings of some words, for example, and ten of others. We subsequently collected data from more subjects to equalize the number of drawings per word across the sample. As a result, our eventual sample was composed of 48 University of Waterloo undergraduates (38 female, one undefined), ranging in ages from 18 to 29 years ($M = 19.70$, $SD = 2.22$), who were recruited to participate for course credit. The findings reported herein are from the entire sample of 48 participants, though we report the analysis for the initial sample of 34 in the Online Supplementary Materials. One participant declined to provide their age and gender. All participants had self-declared normal or corrected-to-normal vision and had learned to speak English before the age of 9 years to ensure language competency.

Materials

Wordlists were created using 320 words taken from a larger list of normed English word lemmas (Brysbaert, Warriner, & Kuperman, 2014). Wordlists were of overall average length (characters, $M = 7.20$, $SD = 1.29$), frequency (SUBTLEX, $M = 161.16$, $SD = 105.97$), and concreteness ($M = 3.20$, $SD = 0.93$), and were always presented in Arial font size 72 pt. For each participant, 160 words were pseudorandomly chosen from the master list to act as targets for the encoding task (80 drawn, 80 written), such that word concreteness was reasonably spread and evenly distributed within both trial-types. An additional 160 words were randomly chosen to serve as lures on the subsequent recognition test. Critically, all words used to form the master list were selected based on pre-existing concreteness ratings (Brysbaert et al., 2014), such that they ranged from highly abstract (minimum score = 1.60) to highly concrete (maximum score = 4.79). Participants viewed stimuli and provided responses using a convertible laptop/tablet (Acer One 10), managed by custom Python scripts.

Procedure

Participants were first told that they would be drawing or writing a set of words, one at a time. They were instructed to keep drawing or writing until the computer proceeded to the next trial. Participants were unaware that their memory would be tested.

The tablet was laid flat on the table in landscape orientation in front of the participant. Participants first saw a task prompt (indicating to “draw” or “write”), followed by a word that they were to perform the task with. On each trial, the task prompt appeared for 1,050 ms, followed by a fixation cross for 350 ms, and then finally by a target word for 1,050 ms (each target word was only ever presented once per participant). Participants were then given 10 s to perform the task before the next prompt and word appeared. Instructions were given to encourage participants to make use of the entire time allotted to each trial.[1] The position of their stylus was documented throughout the trial (for examples of typical drawings in the experiment, see Fig. 1). The entire duration of the encoding task was approximately 30 min. Words were pseudorandomly assigned to the “draw” and “write” trial types, which were mixed and presented in random order.

Following 160 trials of encoding (80 drawn, 80 written), the researcher flipped the computer screen 180° back to laptop mode. Instructions were then provided for the tone-classification filler task. During this task, the participant was instructed to respond by pressing 1, 2, or 3 on the keyboard, if a presented tone was low, medium, or high in pitch, respectively, for 2 min. The purpose of this task was to guard against any potential ceiling effects by preventing rehearsal during a test delay, as young adult samples typically display impeccable memory (Light & Singh, 1987). Following 2 min of this

[1] To ensure that participants were consistently on-task throughout the study phase, we analyzed how long participants spent with the stylus pressed to the touchscreen, as well as how late into each trial a participant committed their last stylus press. The results of this analysis can be found in the Online Supplemental Materials.

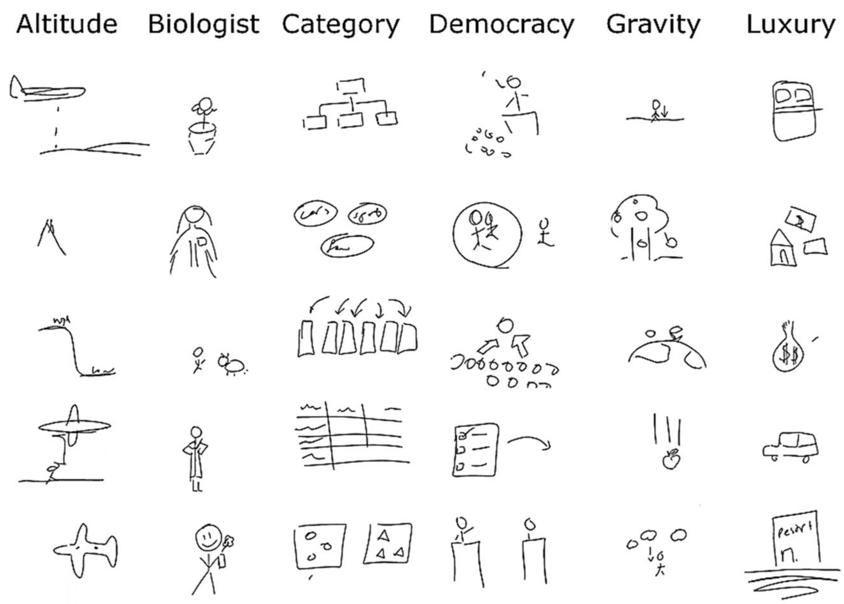

Fig. 1 Examples of drawings created at encoding

tone-classification task, an instruction screen again prompted the participant to wait for the researcher to explain the next phase of the experiment. Next, with the computer still in laptop mode, the researcher then explained the subsequent task: a Remember-Know-New (RKN) recognition test (Tulving, 1985). This test used all 160 target words (80 drawn, 80 written) from the study phase, as well as 160 new randomly selected words that would serve as lures. All words were presented one at a time in the center of the screen in random order. Participants had 3 s to respond on each trial, at which point the next word was presented. If they did not respond in 3 s, a short high-pitched tone was played to indicate an "error," and that they should respond faster on the next trial. If a timing error was made and the previously described tone was played, the screen advanced to the next trial and no response was recorded. Participants were told to press 1, 2, or 3 on the keyboard, to indicate their response of "remember," "know," or "new," respectively. For this task, "remember" responses were verbally defined to the participant as being a conscious recollection of specific contextual information about their initial encounter with the word during the study phase. A "know" response was defined as only having a feeling that they had seen the word previously, but could not remember specific details of the event. "New" responses were of course explained as being for words that had not been seen during the study phase. An RKN style test was used to allow for a more thorough investigation of varying memorial strength for words, rather than recording whether a participant simply had a recollective experience or not (as is the case in free-recall tests; for a review see Yonelinas, 2002). Finally, a chance for questions was provided before the recognition test began. After a maximum period of 18 min (360 trials, 3 s each), the recognition test ended. Participants were given a detailed feedback letter about the experiment, and if applicable, the researcher answered participants' questions.

All procedures and materials were approved by the Office of Research Ethics at the University of Waterloo (ORE #41253). Data for the current experiment are available on the Open Science Framework (OSF) at https://osf.io/6gcpt/.

Results

Before formal statistical analyses were performed, four participants were removed from the data for having false-alarm rates above 95%, making their memory sensitivity scores (*d'*)[2] 2.80 or more standard deviations below the mean. Therefore, we proceeded with formal analyses using a remaining sample of 44 participants.

Concreteness as a continuous predictor of recognition accuracy

Experimental items were selected such that they spanned the spectrum of possible concreteness ratings. As such, we analyzed the data using raw concreteness values (i.e., instead of dichotomizing) as a predictor. To do this while also

[2] Throughout this article, all *d'* and *c'* values have been corrected using the log-linear rule applied to raw hits and false alarms (Hautus, 1995). This method reduces extreme values in a consistent, less biased, and more conservative way than other common corrections.

incorporating the effects of trial type required using a multi-level logistic regression approach.[3] We used this approach to predict the likelihood of an "old" response (collapsed across 'R' and 'K').[4] The primary model of interest incorporated Concreteness as a predictor, as well as dummy-coded Trial Type (draw and write), and their interactions (the Full model). We compared the Full model to models that did not include the interactions (the Concreteness model), included only trial type (the Trial Type model), and included only whether an item was actually old or not (the Null model). Akaike weights (AICw; Wagenmakers & Farrell, 2004) were computed based on the likelihood estimates for each model, and used to determine the best fitting model among the candidate models. We also confirmed the outcome of the AICw comparison using direct model comparisons.

The Full model produced the best fit to the data among the candidate models, $AICw > 0.999$, and significantly outperformed the next-best Concreteness model, $\chi^2(2) = 100.44$, $p < .001$. The estimate for the intercept was -0.85 (95% confidence interval (CI_{95}) [-1.39, -0.31]), indicating that, not surprisingly, lures are unlikely to be called "old." The estimates for draw (1.92; CI_{95} [1.51, 2.33]) and write (1.28; CI_{95} [0.93, 1.63]) were reliable, indicating that both trial types led to above-chance memory performance. The estimate for Concreteness (-0.17; CI_{95} [-0.23, -0.09]) was also reliable, indicating that concreteness had an impact on recognition performance. These main effects are qualified by significant interactions between concreteness and the draw (0.63; CI_{95} [0.49, 0.76]) and write (0.36; CI_{95} [0.25, 0.46]) trial types. Together, these effects indicate that drawing improved memory more than writing, and that concreteness led to a larger improvement in drawn items than in written (see Fig. 2 for an item-based depiction of this general pattern).

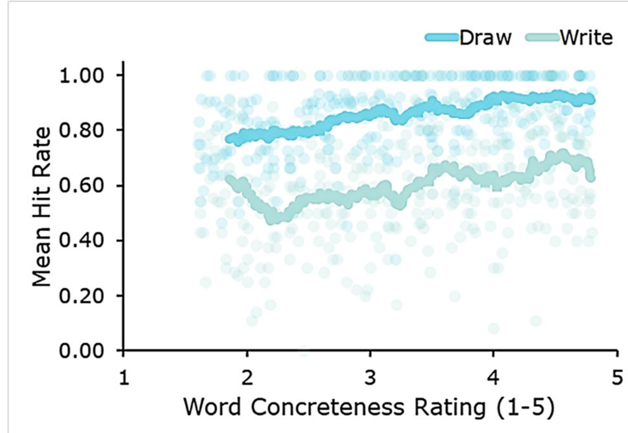

Fig. 2 Moving averages (hit rate) for drawn and written items, as a function of word concreteness rating (1–5; higher is more concrete). Outlier data are not included here

Exploratory visual similarity analysis

Next, we conducted an exploratory analysis (see the Online Supplementary Materials for methods) as a preliminary foray into testing whether the content of the actual drawings produced by participants could be used to predict memory performance. To provide a rich, descriptive measure of the images' content and their similarity to one another, we extracted features from a deep convolutional neural network, or dCNN (the Visual Geometry Group model; VGG19; Simonyan & Zisserman, 2015). While this network was not explicitly trained on drawings, there is evidence that the abstract features that drive categorization of real-world images can also be used to categorize drawings (Fan, Yamins, & Turk-Browne, 2018).

Similarity between a given pair of images was estimated by extracting the features of each image from a given layer of the dCNN and computing the Pearson correlation between them. For each image (e.g., subject 1's drawing of "democracy"), this similarity estimate was computed with all other drawings of the same item (i.e., all other subjects' drawings of "democracy") and averaged into one visual similarity score. A high visual similarity score, therefore, would mean that the drawing is highly prototypical (i.e., similar to most other drawings of the item), while a low score would indicate that it was distinctive among other drawings of that same item. Importantly, the features corresponding to drawings of the same item were more highly correlated to one another than to drawings of other items, *ps* < .01 (for a thorough treatment of this visual similarity metric, see the Online Supplementary Materials).

Visual similarity scores were then entered into a logistic regression, along with word concreteness, and these were used as predictors of a participant's memory for studied words on the recognition test. We discovered that exclusively at the first pooling layer (i.e., the earliest layer) there was a significant interaction between visual similarity score and word concreteness in predicting later memory for a drawn word. That is, the Full model containing the interaction term fit the data the best, $AICw = 0.578$, outperforming a model with both main effects, $\chi^2(1) = 4.82$, $p = .028$, models with just one of the main effects, and the Null model. The second-best model, $AICw = 0.281$, contained only a main effect of Concreteness, which is unsurprising given that concreteness is a known factor predicting memory, and our own earlier results highlight its preferential impact on drawn items.

[3] We also employed a more traditional analysis of variance (ANOVA) approach, treating concreteness as a median-split dichotomous factor (concrete vs. abstract). That analysis, found in the Online Supplemental Materials, replicated the present findings.

[4] An RKN recognition test was employed at retrieval to allow for a more thorough investigation of varying memorial strength for encoded words. For a full analysis of R and K responses separately, see the Online Supplemental Materials.

In the Full model, the estimate for the intercept was 1.32 (CI_{95} [0.80, 1.84]). The main effect of Visual Similarity (-1.10; CI_{95} [-2.03, -0.17]) was reliable, but the effect of Concreteness (-0.12; CI_{95} [-0.27, 0.03]) was not. These main effects were qualified by a reliable interaction (0.31; CI_{95} [0.03, 0.59]) between the two. So, while in general prototypical drawings were more likely to be forgotten, this effect was dependent on concreteness. The interaction indicated that for concepts that are more abstract, distinctive drawings are better remembered, while for concepts that are more concrete, prototypical drawings are better remembered (see Fig. 3).

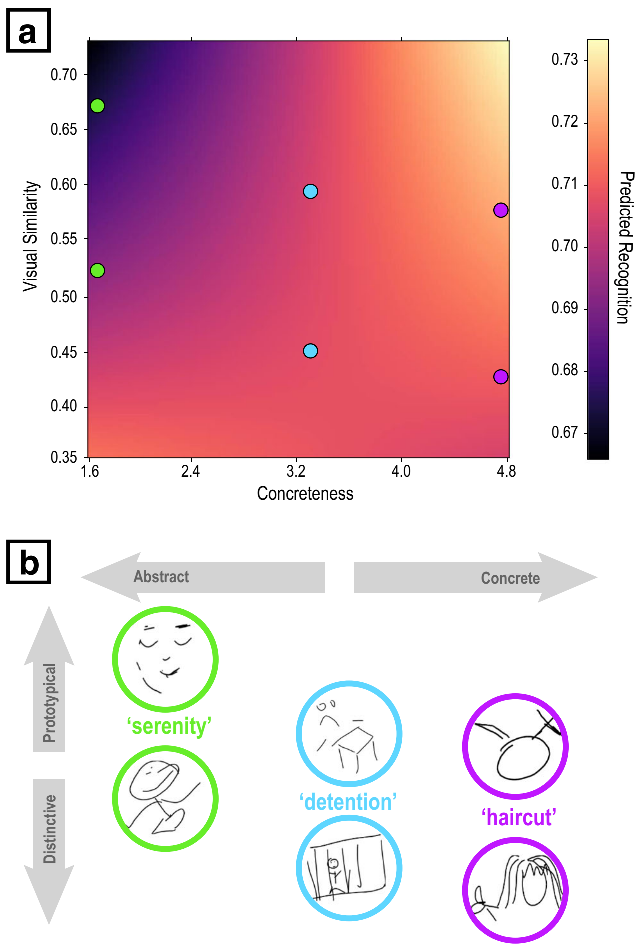

Fig. 3 In both panels, the x-axis or horizontal plane represents Concreteness, ranging from Abstract (left) to Concrete (right), and the y-axis or vertical plane represents Visual Similarity ranging from Distinctive (bottom) to Prototypical (top). **Panel A** shows Predicted Recognition memory (colorbar on right) as a function of Visual Similarity and Concreteness. Three pairs of points at the same level of Concreteness as one another are highlighted in green (left), blue (middle), and purple (right) in Panel A. These are the coordinates of the example images depicted on the left, middle, and right in Panel B, respectively. **Panel B** shows sample drawings for the words "serenity" (left, green; very abstract), "detention" (middle, blue; moderate) and "haircut" (right, blue; very concrete), including an example of a very distinctive example on top, and a very prototypical example on the bottomof each pair

Discussion

Taken together, our findings here highlight that establishing a link between underlying verbal semantic codes and rich multimodal perceptual details can be a potent driver of improved memory. The present experiment demonstrates that even in abstract words, where there is not a strong link between the verbal label and a real imageable referent, memory performance is improved following drawing relative to repeatedly writing. The current findings replicate the drawing effect as delineated in prior work (e.g., Wammes et al., 2016). Trial-wise analyses also revealed that the drawing effect is present across all levels of word concreteness, but we observed a significant interaction in both dichotomized and continuous data, demonstrating that the magnitude of the drawing effect is larger in the most concrete words. It has been proposed – both for drawing and concreteness effects in memory – that a benefit arises because of strong or strengthened associations with the multisensory and semantic information about a word's referent. The observed interaction suggests that a concrete word benefits from the existence of these associations, but that drawing can still provide an *additional* benefit, perhaps by strengthening or emphasizing this link. While we observed better memory in words with high concreteness, the pattern of data here also indicated that drawing can boost memory performance for abstract words up to the same level as written concrete words. This provides compelling preliminary evidence that drawing can dramatically attenuate concreteness effects, perhaps in part through the addition of an otherwise absent visual referent.

While our results clearly bear on mechanistic explanations for drawing and concreteness effects, they also illuminate basic principles of how memory operates. Current theories (i.e., dual coding and context-availability) differ in whether a memory for a more concrete or imageable concept is encoded so strongly because of newly forged links with multimodal and conceptual knowledge, or due to longstanding associative interconnections, respectively. Our work provides novel evidence, adding to an existing literature, to support that the active introduction of multiple memory codes may provide the scaffold for strongly encoded and/or more retrievable memories. Based on arguments presented in the *Introduction*, our hypothesis was that if drawing aided memory performance for abstract words, this could be taken as evidence that concreteness is a phenomenon that occurs primarily at encoding, supporting a dual-coding account. In contrast, context-availability theory would have predicted *no* increase in memory performance for abstract words after being drawn, as the semantic contextual network in which they reside had been unaltered. Since our work showed a clear attenuation of the concreteness effect by implementing the creation of visual imagery at encoding via drawing, it stands to reason that our results support the dual-coding theory.

We also directly explored the actual contents of participants' drawings in an attempt to quantify the relation between the distinctiveness or prototypicality of one's drawing, and the likelihood that they would later remember the word. Specifically, we found that in highly concrete words, increased prototypicality predicted better later memory. In contrast, for highly abstract words, increased distinctiveness instead led to superior memory. These results ran contrary to our predictions that the prototypicality of a drawing would play a supporting role for highly abstract words that do not inherently have an imageable referent. That is, we predicted that drawing an abstract concept (e.g., "love") using its prototypical referent (e.g., a heart) would be more akin to drawing a concrete word since a referent already exists. This analysis was exploratory, and our predictions were primarily speculation, but we believe that it still provides promising evidence that the content of drawings can be used to predict later memory performance. While our findings for drawing similarity of abstract words were not in line with our predictions, they were in line with literature on distinctiveness effects, and the improvement to memory that these effects entail (von Restorff, 1933). More broadly, these exploratory results may provide a reason to re-evaluate any one-size-fits-all theories of imagery and encoding, instead suggesting that there may be interactions between what one is trying to remember (e.g., the concreteness effect), how they are encoding it (e.g., the drawing effect), and its relation to other content (e.g., prototypicality).

It is worth noting that the VGG19 dCNN employed in the current study had been pre-trained using millions of real-world images, and had not "seen" any drawings before this study. Accordingly, this dCNN encodes more generic visual perceptual features, as opposed to features highly tuned to black and white line drawings. While this could be viewed as a shortcoming, evidence presented both here (see Online Supplementary Materials) and in other work (Fan et al., 2018) has demonstrated that features extracted from dCNNs trained on real-world images carry information relevant to recognizing and classifying drawings. Importantly, while these neural networks can only provide an approximation of visual similarity, feature correspondences in similarly trained networks have been found to highly correlate with human raters' ability to judge image similarity based on high-level perceptual features such as shape (Kubilius, Bracci, & Op de Beeck, 2016).

A classic view of memory encoding is that it can be improved by a deeper level of processing, implying a greater degree to which one is cognitively engaged with semantic information about the material (LoP; Craik, 2002; Craik & Lockhart, 1972; Craik & Tulving, 1975). Arguably, almost any engaging encoding task will command a greater depth of processing than the traditional "shallow" controls (e.g., determining whether a word is printed in capital letters). While drawing surely engages deep semantic processing, we argue that it is a special form of encoding that promotes the integration of not only this semantic processing, but also visuomotor processing of what the item looks like and the movements required to depict it. Drawing outperforms writing, even when greater depth of processing is emphasized in the writing condition by instructing participants to add visual details to their writing, or when semantic features of concepts were directly highlighted by asking participants to list them at encoding (Wammes et al., 2016). Within the current work, items that were written were also likely processed deeply due to the fact that a concreteness effect was found within written items (deep processing is thought to be a prerequisite for this effect to occur, see West & Holcomb, 2000). Moreover, analyses of time spent on-task at encoding found significantly higher engagement while writing relative to drawing (task engagement time is often used as a proxy for deeper processing; Craik & Lockhart, 1972; Lupyan, 2008; see the "time-on-task" analyses in Wammes et al., 2018, and in the current Online Supplemental Materials).

When using task contrasts (e.g., draw vs. trace) to gauge the relative role of types of processing involved in drawing, recent work has established that while deep generative semantic processing contributed to the memory boost following drawing, motor and visual processing reliably did as well, with the motor aspect having the greatest contribution of the three (Wammes et al., 2019). Moreover, related literature has shown strong behavioral and functional connectivity evidence for the formation of links between visual and motor systems following repeated practice of handwriting (Longcamp, Tanskanen, & Hari, 2006; Vinci-Booher, James, & James, 2016; for a review, see Feder & Majnemer, 2007), and writing of letters and digits (Zemlock, Vinci-Booher, & James, 2018). Likewise, repeated drawing practice is associated with changes in how information is shared between visual and motor planning systems in the brain (Fan et al., 2020). It has been speculated that this visuomotor integration creates an action-perception link whereby the discriminability of neural representations of practiced items is enhanced (Fan et al., 2020; Longcamp et al., 2006). It is for these reasons that, here, we focus on the beneficial aspects of visuomotor integration induced by drawing that occur beyond any effect of deep semantic processing. In summary, while a depth of processing account could partially explain why engaging tasks like drawing improve memory, we believe that a sensorimotor integration mechanism more thoroughly explains the observed effects.

Planned future work will seek to build upon this foundation with experiments more explicitly designed to probe the *content* of drawings as it relates to subsequent memory for studied items. Indeed, there may be other characteristics of a drawing's content that contribute to its overall likelihood to be remembered. Beyond the prototypicality of an item's referent, these other factors could include, for instance: Inclusion of text or labels in the drawing, whether the drawing is contained within a larger scene context, if the depiction to be drawn is instead based on a phrase or passage rather than a single word, or even the animacy of a drawn item.

Another potential future avenue for research could involve further dissemination of the "generative" factor of drawing in relation to concreteness. As discussed in the *Introduction*, the generative stage of preparing to draw can lead to improved memory, even without the physical act of drawing itself (Wammes et al., 2018). It remains an open question, however, whether the enhancement of memory for abstract words in the current study was resultant from initial generative combinations of *internal* imagery and conceptual knowledge, or if the memory boost observed for abstract words was instead primarily due to clear *external* visual feedback provided once an item was drawn. Nevertheless, it is clear to us that the addition of imagery at some point during the drawing process explains the observed memory improvement for abstract words in the current study.

Our findings contribute novel evidence to both the emerging domain of drawing research as well as more broadly established conversations surrounding memory improvement techniques in the literature. It is clear that rich multimodal integration of an item with its underlying verbal code can prove to be a compelling method of enhancing memory. The current study has demonstrated that despite ostensibly being reliant on direct links to imageable content, the benefits of drawing are robust even for abstract targets. Further, drawing demonstrably attenuated concreteness effects. Exploratory analyses using a neural network provided preliminary evidence that prototypicality of a drawing can influence its likelihood of being remembered. These results suggest that not only can drawing be used in part to form an otherwise absent imagery trace for abstract concepts, but also that *what* one chooses to include in the drawing determines later memory. Overall, the current results highlight the importance of understanding possible interactions between stimulus properties and the ways with which they are encoded when investigating the determinants of successful memory.

Acknowledgements This work was supported by an NSERC postdoctoral fellowship and a Banting SSHRC postdoctoral fellowship awarded to J. D. W.

Open Practices Statement Data for the current experiment are available at: https://osf.io/6gcpt/. The experiment was not preregistered.

References

Anderson, J. L., Ellis, J. P., & Jones, A. M. (2014). Understanding early elementary children's conceptual knowledge of plant structure and function through drawings. *CBE Life Sciences Education*, *13*(3), 375–386. https://doi.org/10.1187/cbe.13-12-0230

Begg, I., & Paivio, A. (1969). Concreteness and imagery in sentence meaning. *Journal of Verbal Learning and Verbal Behavior*, *8*(6), 821–827. https://doi.org/10.1016/S0022-5371(69)80049-6

Blake, A. B., & Castel, A. D. (2019). Memory and availability-biased metacognitive illusions for flags of varying familiarity. *Memory and Cognition*, *47*(2), 365–382. https://doi.org/10.3758/s13421-018-0872-y

Bransford, J. D., & McCarrell, N. S. (1974). A sketch of a cognitive approach to comprehension. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes* (pp. 189–229). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. https://doi.org/10.1007/BF01320076

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Craik, F. I. M. (2002). Levels of processing: Past, present ... and future? *Memory*, *10*, 305–318. https://doi.org/10.1080/09658210244000135

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Deng, J., Li, L.-J., Li, K., Li, F. F., Dong, W., Socher, R., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Edens, K. M., & Potter, E. (2010). Using descriptive drawings as a conceptual change strategy in elementary science. *School Science and Mathematics*, *103*(3), 135–144. https://doi.org/10.1111/j.1949-8594.2003.tb18230.x

Fan, J. E., Wammes, J. D., Gunn, J. B., Yamins, D. L. K., Norman, K. A., & Turk-Browne, N. B. (2020). Relating visual production and recognition of objects in human visual cortex. *Journal of Neuroscience*, *40*(8), 1710–1721. https://doi.org/10.1523/JNEUROSCI.1843-19.2019

Fan, J. E., Yamins, D. L. K., & Turk-Browne, N. B. (2018). Common object representations for visual production and recognition. *Cognitive Science*, *42*(8), 2670–2698. https://doi.org/10.1111/cogs.12676

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Feder, K. P., & Majnemer, A. (2007). Handwriting development, competency, and intervention. *Developmental Medicine and Child Neurology*, *49*, 312–317. https://doi.org/10.1111/j.1469-8749.2007.00312.x

Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science*, *27*(5), 302–308. https://doi.org/10.1177/0963721418755385

Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *NeuroImage*, *32*(3), 1413–1421. https://doi.org/10.1016/j.neuroimage.2006.06.007

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Kieras, D. (1978). Beyond pictures and words: Alternative information-processing models for imagery effect in verbal memory. *Psychological Bulletin*, *85*(3), 532–554. https://doi.org/10.1037/0033-2909.85.3.532

Kubilius, J., Bracci, S., & Op de Beeck, H. P. (2016). Deep neural networks as a computational model for human shape sensitivity. *PLoS Computational Biology*, *12*(4). https://doi.org/10.1371/journal.pcbi.1004896

Light, L. L., & Singh, A. (1987). Implicit and explicit memory in young and older adults. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *13*(4), 531–541.

Long, B. L., Fan, J. E., Chai, Z., & Frank, M. C. (2019). Developmental changes in the ability to draw distinctive features of object categories. *Journal of Vision*, *19*(10), 59b. https://doi.org/10.1167/19.10.59b

Longcamp, M., Tanskanen, T., & Hari, R. (2006). The imprint of action: Motor cortex involvement in visual perception of handwritten letters. *NeuroImage*, *33*(2), 681–688. https://doi.org/10.1016/j.neuroimage.2006.06.042

Lupyan, G. (2008). From chair to "chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, *137*(2), 348–369. https://doi.org/10.1037/0096-3445.137.2.348

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, *83*(4), 484–490. https://doi.org/10.1037/0022-0663.83.4.484

Meade, M. E., Wammes, J. D., & Fernandes, M. A. (2018). Drawing as an encoding tool: Memorial benefits in younger and older adults. *Experimental Aging Research*, *44*(5), 369–396. https://doi.org/10.1080/0361073X.2018.1521432

Moeser, S. D. (1974). Memory for meaning and wording in concrete and abstract sentences. *Journal of Verbal Learning and Verbal Behavior*, *13*(6), 682–697. https://doi.org/10.1016/S0022-5371(74)80055-1

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*(4), 406–419. https://doi.org/10.1037/a0024377

Morey R. D., Rouder J. N. (2018): BayesFactor: Computation of Bayes factors for common designs [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=BayesFactor (R package version 0.9.12-4.2)

Morey, R. D., Wagenmakers, E.-J., & Rouder, J. N. (2016). Calibrated bayes factors should not be used: A reply to Hoijtink, van Kooten, and Hulsker. *Multivariate Behavioral Research*, *51*(1), 11–19. https://doi.org/10.1080/00273171.2015.1052710

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*(2), 176–206. https://doi.org/10.1016/0010-0285(73)90032-7

Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, *11*(4), 137–138. https://doi.org/10.3758/BF03331011

Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: When and why? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1196–1204. https://doi.org/10.1037/0278-7393.20.5.1196

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Schmeck, A., Mayer, R. E., Opfermann, M., Pfeiffer, V., & Leutner, D. (2014). Drawing pictures during learning from scientific text: Testing the generative drawing effect and the prognostic drawing effect. *Contemporary Educational Psychology*, *39*(4), 275–286. https://doi.org/10.1016/j.cedpsych.2014.07.003

Schwanenflugel, P. J., Akin, C., & Luh, W. M. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition*, *20*(1), 96–104. https://doi.org/10.3758/bf03208259

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 82–102. https://doi.org/10.1037/0278-7393.9.1.82

Schwanenflugel, P.J., & Stowe, R.W. (1989) Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly, 24*(1), 114–126. https://doi.org/10.2307/748013

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*. San Diego, CA.

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

Van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology*, *93*(1), 129–140. https://doi.org/10.1037/0022-0663.93.1.129

Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology*, *31*(2), 142–166. https://doi.org/10.1016/j.cedpsych.2005.04.001

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, *17*, 285–325. https://doi.org/10.1007/s10648-005-8136-3

Vinci-Booher, S., James, T. W., & James, K. H. (2016). Visual-motor functional connectivity in preschool children emerges after handwriting experience. *Trends in Neuroscience and Education*, *5*(3), 107–120. https://doi.org/10.1016/j.tine.2016.07.006

von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychologische Forschung*, *18*(1), 299–342. https://doi.org/10.1007/BF02409636

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

Wammes, J. D., Jonker, T. R., & Fernandes, M. A. (2019). Drawing improves memory: The importance of multimodal encoding context. *Cognition*, *191*, 103955. https://doi.org/10.1016/j.cognition.2019.04.024

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

Wammes, J. D., Roberts, B. R. T., & Fernandes, M. A. (2018). Task preparation as a mnemonic: The benefits of drawing (and not drawing). *Psychonomic Bulletin & Review*, *25*(6), 2365–2372. https://doi.org/10.3758/s13423-018-1477-y

West, W., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, *12*(6), 1024–1037. https://doi.org/10.1162/08989290051137558

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. https://doi.org/10.1006/jmla.2002.2864

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *European Conference on Computer Vision*, *8689 LNCS* (Part 1), 818–833. https://doi.org/10.1007/978-3-319-10590-1_53

Zemlock, D., Vinci-Booher, S., & James, K. H. (2018). Visual–motor symbol production facilitates letter recognition in young children. *Reading and Writing*, *31*(6), 1255–1271. https://doi.org/10.1007/s11145-018-9831-z