Cognition 238 (2023) 105435

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Symbol superiority: Why $ is better remembered than ‘dollar’

Brady R.T. Roberts [*], Colin M. MacLeod, Myra A. Fernandes

Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada

ARTICLE INFO

Keywords:
Symbols
Picture superiority
Memory
Dual-coding
Familiarity
Neural network

ABSTRACT

Memory typically is better for information presented in picture format than in word format. Dual-coding theory (Paivio, 1969) proposes that this is because pictures are spontaneously labelled, leading to the creation of two representational codes—image and verbal—whereas words often lead to only a single (verbal) code. With this perspective as motivation, the present investigation asked whether common graphic symbols (e.g.,!@#$%&) are afforded primarily verbal coding, akin to words, or whether they also invoke visual imagery, as do pictures. Across four experiments, participants were presented at study with graphic symbols or words (e.g., $ or ‘dollar’). In Experiment 1, memory was assessed using free recall; in Experiment 2, memory was assessed using old-new recognition. In Experiment 3, the word set was restricted to a single category. In Experiment 4, memory for graphic symbols, pictures, and words was directly compared. All four experiments demonstrated a memory benefit for symbols relative to words. In a fifth experiment, machine learning estimations of inherent stimulus memorability were found to predict memory performance in the earlier experiments. This study is the first to present evidence that, like pictures, graphic symbols are better remembered than words, in line with dual-coding theory and with a distinctiveness account. We reason that symbols offer a visual referent for abstract concepts that are otherwise unlikely to be spontaneously imaged.

Since the times of ancient Greece, philosophers have contemplated the definition and use of symbols. In those times, it was commonly held that ‘signs’ referred to aspects of nature that conveyed meaning, whereas ‘symbols’ referred to cultural conventions that had meaning. To these thinkers, the most obvious examples of symbols were words, which Aristotle referred to as the symbols of the inner images (Modrak, 2001). The notion of words as symbols has persisted for thousands of years. Writing in 1690, John Locke considered the ‘doctrine of signs’ to be fundamental to the communication of science and knowledge to others (da Costa e Silva, 2019). In Shakespeare’s *Romeo and Juliet*, Juliet laments that she cannot marry Romeo due to his family name, claiming “that which we call a rose by any other name would smell as sweet” (Shakespeare, 2011, 2.2.890–891). In essence, Juliet is arguing that the words used to name things are entirely arbitrary, and if only Romeo was born under another family name, the two would be allowed to wed. While most philosophical writing about symbols has been focused on words, contemporary definitions of symbols have explored other ways in which humans ascribe significance to graphic symbols that they have created (e.g.,!@#$%). The purpose of the present study was to investigate how these graphic symbols are represented and remembered. In so doing, empirical research on human cognition was used to bridge the gap between philosophy and psychology.

1.1. Historical context

The philosophical field of semiotics is entirely devoted to the study of signs. American philosopher Charles Peirce is often considered the father of semiotics, so it is not surprising that Peirce’s definitions of signs and symbols have dominated modern discussion of these concepts. By 1867, Peirce had developed a triadic structure of elements that were all encompassed under the category of ‘signs’ (Atkin, 2013). These three elements consisted of the index (associations based in logic: e.g., ☠ to represent poisonous), the icon (associations based in physical resemblance: e.g., ✆ to represent a phone), and the symbol (associations derived from culture: e.g., $ to represent dollar).

Although other definitions of symbols certainly exist (e.g., Goodman, 1976; Huttenlocher & Higgins, 1978; de Saussure & Harris, 1998), Peirce’s are among the most prevalent in semiotics. Peirce’s triadic structure of signs provides a useful framework to conceptualize visual communication, but it is not without its faults. For instance, the boundaries between the three types of signs are not as clear as one might expect: Signs can certainly share attributes of multiple triadic elements.

---

[*] Corresponding author at: Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada.
E-mail address: bradyrtroberts@gmail.com (B.R.T. Roberts).

https://doi.org/10.1016/j.cognition.2023.105435
Received 3 August 2022; Received in revised form 2 March 2023; Accepted 5 March 2023

The sign for 'cut' (✂) could be considered an icon because it resembles scissors, but it also could be considered an index because it is used to indicate a cutting action in virtual space (in word processing software) rather than signifying the presence or use of real scissors. Although Peirce acknowledged this potential for sign overlap in his writings, trying to fit signs into neat categories can cause confusion as multiple interpretations are often correct. Further, because words can be considered Peircean symbols, there is understandable confusion between what a philosopher might call a symbol and what a lay person might think a symbol is. Even in the case of words, which many consider to be the original 'symbols,' research on sound symbolism has found that some words (e.g., 'buzz') are more iconic than others (Vinson et al., 2021), suggesting that language is not always arbitrarily or culturally derived (meaning language is not always purely symbolic).

In the present study, a mix of common graphic symbols (see Fig. 1) was used to investigate how symbolic content is represented in memory, allowing parallels to the more established literature surrounding memory for images. Whereas Peirce likely would have claimed that the present study involves a mix of three types of 'signs,' his strict definitions are avoided here to prevent a misunderstanding that the current research concerns memory for physical signs (e.g., traffic signs, billboards) rather than graphic symbols like those found on a keyboard.

1.2. What do we know about symbols?

Despite philosophers having been intrigued by symbols for thousands of years, the discipline of psychology has been remarkably ambivalent toward them. What little cognitive research does exist on symbols often does not make use of symbols that participants would recognize. For instance, Lupyan and Spivey (2008) found that low-level perceptual processing of symbols could be moderated by top-down feedback from conceptual understandings, but they used entirely novel symbols that participants had to learn during the study. Similarly, Coppens, Verkoeijen, and Rikers (2011) demonstrated that the testing effect (Bjork, 1975; Roediger & Karpicke, 2006) is generalizable to symbol-word pairs, but the symbols that they used were West African 'Adinkra' characters that were novel to participants. Aggravating this problem, because symbols are culturally specific, some very useful and relevant work has been limited in its generalizability. For instance, Prada, Rodrigues, Silva, and Garrido (2016) developed the Lisbon Symbols Database which contains normative ratings for over 600 symbols. Unfortunately, because many of these symbols are more common in Portuguese society, they are less familiar to participants in other parts of the world.

Ironically, perhaps the most common use of meaningful symbols in psychological research to date has been for the purpose of filler stimuli, as in fMRI experiments (e.g., using '#####' or other symbols as a visual mask). Some of the more thorough work in this area, however, reports findings of neural activation during presentation of these 'filler' stimuli, providing insight into the underlying neural representations of symbols. For instance, Reinke, Fernandes, Schwindt, O'Craven, and Grady (2008) found that words and symbols both activated the left visual-word form area (VWFA) in the brain, whereas numbers and letters did not. In line with this finding, Kronbichler et al. (2004) demonstrated that symbols are processed in areas associated with perception of abstract visual stimuli. Finally, Carreiras, Armstrong, Perea, and Frost (2014) concluded that processing of letters, numbers, and symbols, while sharing some overlap, largely involve the recruitment of distinct neural regions. For example, they reported that symbols activated unique brain areas—namely, the right superior parietal and the right middle and inferior temporal cortices—different from those activated by letters. These are regions thought to be involved in semantic processing (Binder & Desai, 2011; Binder, Desai, Graves, & Conant, 2009). In addition, neural activation in the right middle temporal gyrus was higher when symbols, as opposed to numbers or letters, were presented (Carreiras et al., 2014). Importantly, damage to this specific brain region has been implicated in agraphia and alexia for Japanese kanji (logographic characters akin to symbols; Sakurai, Mimura, & Mannen, 2008). The aforementioned fMRI and neurological patient studies suggest that symbols not only hold semantic information similar to words, as one might expect, but that they also evoke neural mechanisms distinct from semantically-void single-character stimuli like letters.

Related work suggests that cognitive processing for Arabic numerals (e.g., 1, 2, 3) shifts from frontal to parietal sites later in life, which could be indicative of greater automaticity in the mapping between number and magnitude (Ansari, Garcia, Lucas, Hamon, & Dhital, 2005). Might similar character-to-meaning mapping processes be at play for symbols? Perhaps as one becomes increasingly familiar with a symbol, the learned association between the perceptual symbol and the semantic referent becomes increasingly strengthened. We examine this possibility in the present study.

1.3. Are symbols more like images or words?

Despite symbols having received relatively little attention in psychological research, pictures have benefitted from a rich history of work

Fig. 1. Graphic symbols used in the experiments.
Note. In Experiments 1 and 4, only the first 50 of these 80 symbols were used.

going back to the late 1800s (e.g., Bergstrom, 1893). A significant body of work emerged in the 1960s to suggest that pictures are often better remembered than words. This phenomenon—the picture superiority effect (PSE)—has been extensively studied for almost six decades. Early on, Allan Paivio (1969, 1971) proposed his influential dual coding theory—that pictures are remembered better than words because pictures often are represented by two distinct codes—verbal and image—whereas words tend to evoke only the verbal code.

Because symbols are often used alongside text, is it possible that they are processed like words? This could be. But symbols are not made up of meaningless graphemes (i.e., letters) as words are. Instead, symbols are more holistic and unique in their formation, and carry a meaning on their own, more like images. Therefore, when conceptualizing symbols within the dual coding framework designed for pictures, one could argue that symbols may also elicit an image code in memory as a result of their holistic, non-verbal nature.

Paivio's dual coding theory is also often used to explain why concreteness effects occur in memory: Concrete words are more easily imageable than abstract words, making concrete words better remembered as they are more likely to elicit spontaneous imagery (Ding, Liu, & Yang, 2017; Fliessbach, Weis, Klaver, Elger, & Weber, 2006; Hamilton & Rajaram, 2001; Jessen et al., 2000; Khanna & Cortese, 2021; Paivio, Walsh, & Bons, 1994; Roberts & Wammes, 2021). Therefore, according to dual coding theory, any word that is imaged (either in the physical world or via imagination) necessarily becomes 'concrete' insofar as it affords the stimulus an image code in memory. For example, dual coding theory would predict that the word 'apple' will be better remembered than the word 'peace' but that a picture of an apple should be remembered just as well as an image representing peace, such as, for example, pictures of white doves or olive branches. Recent work from our laboratory has found evidence consistent with this prediction: Drawing a picture related to an abstract word brings memory performance for that word up to the level of written concrete words (Roberts & Wammes, 2021). In that study, when asked to draw certain abstract words (e.g., 'love'), many participants chose to draw common symbols (e.g., ♥), rather than more elaborate scenes, to depict the concepts.

This pattern of response was a direct motivator for the current study. Is it possible that drawing a symbol could serve as an effective way to represent an abstract concept without picturing something else that is already concrete? It seems plausible: Drawing '☮' should serve to concretize the word 'peace' as well as olive branches or white doves but should simultaneously avoid the confound of using other related concrete things to depict the abstract concept. That is, representing an abstract idea with a graphic symbol (e.g., ☮) provides the participant with a pictorial (and therefore concrete) representation of that abstract concept, likely eliciting a secondary image code in memory to accompany the verbal code of the word itself (e.g., 'peace'). This logic formed the basis of our principal prediction: As a result of their dual codes in memory, symbols and images should both be better remembered than words but memory for symbols and images should not differ.

Of course, other theories have been put forth to explain picture superiority, most rooted in some variant of a distinctiveness account (Hunt & McDaniel, 1993). Thus, a conceptual distinctiveness account contends that pictures are conceptually more distinctive than words because they elicit greater elaboration (Hamilton & Geraci, 2006; Nelson, Reed, & McEvoy, 1977). In contrast, a physical distinctiveness account argues that pictures vary more in visual appearance than words (since words, in the English language at least, are made up of the same 26 recycled letters; Ensor, Surprenant, & Neath, 2019; Mintzer & Snodgrass, 1999).

When applied in the present case, these physical and conceptual distinctiveness accounts should also predict better memory for symbols than for words. Relative to words, symbols are more physically distinct: They consistent of varying lines, shapes, orientations, and degrees of symmetry. Furthermore, whereas words in the English language use the same 26 letters repeatedly, there are certainly more than 26 unique symbols, many of which share few visual attributes. In addition, symbols are also conceptually distinct insofar as they have hardly any semantic neighbours. For example, while the word 'play' (in the context of watching television) has many related words, such as 'begin,' 'start,' 'commence,' etc., it is further crowded in semantic space due to its nature as a homonym (e.g., 'play' can also refer to games or to theatrical performances). The symbol for 'play' (▶), on the other hand, stands in relative semantic isolation from other symbols; no other symbol that conveys a similar meaning comes to mind. Therefore, the symbol for play would be considered to bear high conceptual distinctiveness (Hamilton & Geraci, 2006), whereas its word counterpart would not. Importantly, however, more recent work (Ensor, Surprenant, & Neath, 2019) has suggested that distinctiveness and dual coding may each contribute to picture superiority, depending on the type of retrieval test that is employed, a key idea upon which we expand in the General Discussion.

1.4. What other factors might contribute to symbol processing?

A graphic symbol—like any other sign—must be known to the interpreter before it can convey meaning. Therefore, it is possible that memory for a symbol could be moderated by one's familiarity with it. That is, if one did not know that '%' means 'percentage,' dual coding theory would suggest that the symbol would fail to elicit a verbal code in memory. Rather, the symbol would simply appear as a meaningless shape and therefore would be less likely to be retrieved on a memory test. It is possible, however, that one could generate a verbal label or descriptor for the unknown symbol (e.g., % = two circles with a diagonal line between them), and this could boost memory for it, but any benefit of doing so would be lost in a context in which the meaning of the symbol must be known to permit interpretation of the information being conveyed (e.g., in sentences). Therefore, it stands to reason that, if symbols are indeed dual coded and consequently lead to better memory than words, this performance discrepancy ought to be attenuated when one's familiarity with a symbol is so low that it cannot be labelled.

The predictions made thus far for symbol superiority in memory have been rooted in Paivio's dual coding theory. But what if the physical form of the symbol is the true driver of differential retention, as proposed by the physical distinctiveness account of picture superiority? One way to investigate this question would be to turn to the emerging area of intrinsic stimulus memorability. Memorability—the likelihood that something will later be remembered—has recently been explored as a perceptual attribute of a stimulus, a quality that is independent of stimulus type, aesthetics, emotionality, priming, and attention (Bainbridge, 2019, 2020, 2021, 2022; Bainbridge, Dilks, & Oliva, 2017; Bainbridge, Isola, & Oliva, 2013; Bainbridge & Rissman, 2018; Brady & Bainbridge, 2022; Goetschalckx, Moors, & Wagemans, 2018; Isola, Jianxiong, Parikh, Torralba, & Oliva, 2014; Isola, Xiao, Torralba, & Oliva, 2011; Xie, Bainbridge, Inati, Baker, & Zaghloul, 2020). It could therefore be the case that, relative to words, symbols simply are unique in their physical and/or conceptual forms, and it is this property that underlies improved memory performance. These possibilities and others are examined in the present study.

1.5. The present study

Five experiments were conducted, all aimed at addressing how symbols are represented in and retrieved from memory. It could be that symbols, being unitary characters not comprised of graphemes (unlike words formed from letters), are processed holistically like pictures. Moreover, because symbols are most often used to represent abstract concepts, they could serve to 'concretize' those ideas by providing quick and easy-to-recognize visual referents. Such a visual referent could aid memory by providing an image record (along the lines of Paivio's dual coding theory) or by offering a physically distinct visual form that stands out in memory. In either case, memory for symbols should be superior to memory for words but should not differ from memory for pictures.

In Experiment 1, participants were presented with symbols or their word counterparts (i.e., ☮ or 'peace') and were later tested on their memory for the studied items. Experiment 2 conceptually replicated the first study with a different experimental design. Experiment 3 addressed the potential confound of differing set sizes between words and symbols by reducing words to a single category. Experiment 4 compared memory for symbols, pictures, and words directly in a single investigation. It is also noteworthy that in Experiments 1 and 4 testing was via recall, whereas in Experiments 2 and 3 testing was via recognition. This also permitted us to consider the relevance of type of retrieval test to the underlying mechanism.

In Experiment 5, familiarity and frequency ratings were collected for the set of symbols used in the first four experiments, the goal being to determine whether general familiarity with a symbol would predict later memory for it. The ResMem neural network (Needell & Bainbridge, 2022) was also employed to assess whether symbols are better remembered than words simply as a result of their inherent memorability. This set of five studies began with a straightforward initial experiment examining whether memory differs for symbols and their word counterparts.

2. Experiment 1

Recall that the major prediction was that symbols are effectively processed as pictures insofar as they serve to concretize abstract concepts by providing a unique visual referent. In the first experiment, it was predicted that if symbols are in fact 'mini-pictures,' and therefore are likely remembered like other images, then they should be better remembered than words as a result of dual coding and/or enhanced distinctiveness. To test this basic prediction, a simple two-block study was designed using nonoverlapping stimulus subsets in the two blocks. Participants were presented with symbols (e.g., '$') in one block and with words (e.g., 'dollar') in the other block. After a short delay, they were asked to recall the studied items. For the test of the word study block, participants were told to write down the words that they remembered; for the test of the symbol study block, they were asked to draw the symbols that they remembered. This two-block design was chosen to preserve the inherent encoding features of words and symbols and to minimize any 'conversion' of studied symbols into verbal labels (or vice versa) during study or during the memory test. Memory was tested by recall to ensure that when participants remembered a symbol but did not know its verbal label, they were still able to provide evidence of remembering it. Finally, to measure familiarity with the current set of symbols, a matching task was administered following recall in which participants had to decide whether symbols and words had corresponding meanings. This is the first experiment to have tested memory for common everyday graphic symbols and their word counterparts.

2.1. Method

2.1.1. Participants

An a priori power analysis was not conducted because there had been no previously published work documenting memory for symbols. Instead, data were collected from a minimum of 50 participants, continuing until the stopping rule date (the end of the academic term). A total of 155 University of Waterloo undergraduate students took part in exchange for course credit. Participants had self-selected to participate in the study and had self-reported normal or corrected-to-normal vision as well as having learned English before the age of 9.

From this initial sample, participants' data were filtered out if they were found not to have followed the experiment instructions correctly (e.g., if they wrote down items during encoding or wrote out labels of symbols during recall) or had corrupt data files ($n = 24$). Then, *R* (v. 4.1.1; R Core Team, 2020) statistical software was used to exclude participants whose performance was ±3 *SD*s from the mean for recall of symbols ($n = 0$), recall of words ($n = 1$), or accuracy on the symbol matching task ($n = 2$).[1] The final sample of 128 participants used in the statistical analyses was 71.88% female, with ages ranging from 17 to 31 ($M = 19.7$, $SD = 2.1$).

2.1.2. Materials

Two sources provided a set of common everyday symbols: (1) the Wikipedia entry for "Miscellaneous Symbols" (https://en.wikipedia.org/wiki/Miscellaneous_Symbols), and (2) aggregate response data from several different pages on Sporcle, an online quiz-sharing website (https://www.sporcle.com/). Sporcle hosts many user-generated quizzes where, for instance, participants need to match symbols to their definitions. Conveniently, the aggregate response data for all of the attempts made on any Sporcle quiz are freely available to use; four different Sporcle quizzes were found that assessed knowledge of common symbols. The aggregate data from these quizzes were from up to 216,000 plays. Based on the results of these quizzes, symbols that were poorly known were discarded. A list of 101 symbols was assembled. For Experiment 1, the best 50 symbols were selected based on the criteria that (1) they were physically and conceptually distinct from one another, and (2) they were the most common in everyday life for the target population of Western society undergraduate students. All symbol stimuli were created by copying the item from the web using the Segoe UI Symbol font, then editing them to be 110 × 116 pixels in greyscale on a white background (see Fig. 1).

All stimuli in the words block were single-word versions of the symbols on the master stimuli list. All word stimuli were presented at 5% of the total screen height, in black lowercase Calibri font on a white background.

2.1.3. Procedure

The experiment was administered to groups of six participants at a time, each using a separate Windows computer, separated by wall panel dividers. Following informed consent and demographic data collection, each participant was randomly assigned to one of four counterbalanced conditions (the product of the two stimulus subsets crossed with the two block orders)[2] before being seated at a testing desk. Each experiment computer was running PsychoPy (v. 3.2.4; Peirce et al., 2019) experiment builder software, outputting to a 24-in. monitor with 1920 x 1080p resolution (60 Hz). Participants sat roughly 24″ to 30″ from their screen.

Prior to the study phase, participants were told that they would see either words or symbols presented one at a time on the screen and were instructed to try to remember as many as they could for a later memory test. Participants were then presented either with 25 words or with 25 symbols sequentially in the center of the screen. Each stimulus was shown for 2 s, followed by a blank screen for 250 ms, a fixation dot for 500 ms, and finally another blank screen for 250 ms. Encoding stimuli were presented in a random order.

After all of the 25 stimuli for the block had been presented, participants completed a filled delay task. They were instructed to listen to a tone and then to respond by pressing '1', '2', or '3' if the pitch of the tone was 'low,' 'medium,' or 'high,' respectively. Examples of each pitch were provided in the task instructions. Tones were played for 500 ms, with a new tone played every 2 s or when a response was made, whichever came first. This task persisted for 2 min and was included to guard against potential ceiling effects by eliminating recency and by minimizing post-list rehearsal. Following the filled delay, participants were given two minutes to complete the free recall memory test for the items seen during the study phase. Participants were instructed to recall items

---

[1] Retaining data from these participants in the statistical analyses made no difference to the pattern of results.

[2] Due to a programming error, stimuli were not properly counterbalanced across participants. In Experiment 1 only, the items in the symbols list were always presented as symbols and the items in the words list were always presented as words.

'as presented': Word-based items were to be written whereas symbols were to be drawn.

Finally, following the two study-test cycles, a 'symbols matching task' was administered to assess the participants' familiarity with the symbols. In this task, participants saw 25 symbol-word pairs that matched (e.g., '\$ - dollar') and 25 that mismatched (e.g., '& - asterisk') in random order. The matching of symbols to words was counterbalanced across participants such that any particular symbol matched the presented word for half of participants. Participants were instructed to respond by pressing the '1' key if the symbol and word matched or the '2' key if they did not match. After each response, a new pair of stimuli immediately appeared on the screen (i.e., this task was self-paced).

Following the matching task, a paper version of the Mill Hill Vocabulary Scale (Set A; Raven, 1958) was administered to assess participants' English language competency. Finally, a feedback letter was provided that detailed the purpose of the study. The procedures and materials for this study were approved by the Office of Research Ethics at the University of Waterloo (project #41594). All data, analysis code, experiment programs, and other materials are listed on the Open Science Framework (OSF; https://osf.io/e53z4/).

2.2. Results

2.2.1. Memory performance

The first author scored recall of symbols by determining whether the symbols drawn during recall had or had not appeared on the study list. Symbols written out with words were considered incorrect in this experiment because (1) it was desirable to avoid 'conversion' from symbol format to word format, and (2) writing symbols in word format was contrary to the experiment instructions and was therefore rare among participants (seven participants recalled all symbols by writing out their respective labels and were therefore not included in subsequent analyses). Word recall was scored using both lenient (synonyms acceptable) and strict (exact matches only) scoring methods before being converted into proportions by dividing by 25. To ensure the most conservative test of the predictions, recall of symbols was compared to recall of words using the lenient scoring criteria for the latter condition throughout the results reported here.[3]

Critically, the results of a paired-samples *t*-test demonstrated that the proportion of symbols recalled was greater than the proportion of words recalled (see Fig. 2),[4] $t(127) = 8.75$, $p < .001$, $d = -0.77$, $\mathrm{CI}_{95}[-0.97, -0.58]$, $BF_{10} = 6.10\mathrm{e}{+}11$, with extreme Bayesian evidence in support of the alternative model.[5]

2.2.2. Exploratory correlations

To investigate the possibility that one's knowledge of symbols, or of the English language, might enhance recall performance for symbols or words, a series of exploratory Pearson correlations was conducted. Mill Hill Vocabulary Scale scores were weakly correlated with the proportion of words recalled, $r(126) = .17$, $p = .062$, but not with proportion of symbols recalled, $r(126) = .08$, $p = .363$. Performance on the symbols matching task was not significantly correlated with recall performance in either condition ($rs \leq |.14|$, $ps \geq .126$).

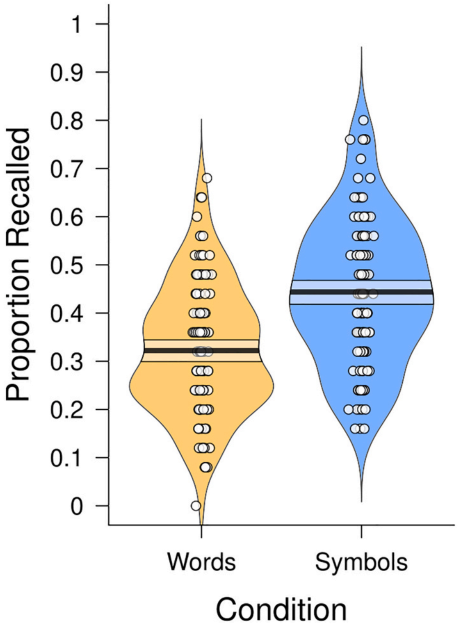

Fig. 2. Experiment 1: recall of words and of symbols.
Note. This figure depicts lenient scoring for word recall. Errors bars = 95% confidence intervals.

2.3. Discussion

In Experiment 1, a basic hypothesis was tested—that symbols would be better remembered than their word counterparts. This hypothesis was based on the dual coding account of how pictures are represented in memory (Paivio, 1969, 1971). This prediction clearly was borne out: Symbols were indeed better remembered than their word-based counterparts. However, there were two limitations to this experiment. First, as noted in Footnote 1, due to a programming error, items from the master word list were not properly counterbalanced into the symbols and words groups. Second, and more critically, because participants were asked to *draw* symbols and *write* words during the recall test, it is possible that the superior memory for symbols observed here was due to some drawing-related boost, along the lines of memory benefits known to operate following drawing at encoding (i.e., the drawing effect; see Wammes, Meade, & Fernandes, 2016). The goals of Experiment 2 were to address these two limitations and to expand the generalizability of the findings by switching the setting, study design, and type of retrieval test.

3. Experiment 2

Experiment 1 provided evidence that memory is indeed better for symbols than for words. There were, however, limitations to the experiment. To address these, Experiment 2 switched to a between-subjects design and used a recognition test rather than a free recall test to assess memory. The switch to recognition testing was intended to ensure that any difference in memory between symbols and words was not due to having to draw (the symbols) at retrieval. The switch to a between-subjects design was intended to ensure that participants did not have better memory for symbols simply because they saw these as more unusual or important compared to words when the two types of stimuli appeared in the same list.

In Experiment 2, participants were randomly assigned to one of four

---

[3] Using a strict scoring criterion for word recall yielded no change in the pattern of results: The difference between words and symbols simply became larger, $t(127) = 9.36$, $p < .001$, $d = -0.83$, $\mathrm{CI}_{95}[-1.03, -0.63]$, $BF_{10} = 1.61\mathrm{e}{+}13$.

[4] This analysis was also conducted with the pre-registered target sample size of 50 participants by randomly selecting data from the full data set. The pattern of results was identical.

[5] Throughout this article, Bayes factors were calculated using the *BayesFactor* (Morey et al., 2011) package for *R*, enlisting a default Jeffreys-Zellner-Siow (JZS) prior with a Cauchy distribution (center = 0, $r = 0.707$). This package compares the fit of various linear models. In the present case, Bayes factors for the alternative (BF_{10}) are in comparison to intercept-only null models. Bayes factor interpretations follow the conventions of Lee and Wagenmakers (2013). Bayes factors in favor of the alternative (BF_{10}) or null (BF_{01}) models are presented in accordance with each preceding report of NHST analyses (i.e., based on a $p < .05$ criterion) such that $BF > 1$.

groups in a 2 × 2 design, the factors being stimulus type at encoding and stimulus type at test. As a result, stimuli seen on the recognition test would either by congruent with those seen at encoding (e.g., $ ➔ $; dollar ➔ dollar), or they would be incongruent (e.g., $ ➔ dollar; dollar ➔ $). The prediction was that a pure symbols-symbols (encoding-retrieval) condition would still lead to better memory than a pure words-words condition because the hypothesized memory difference occurs due to better encoding of symbols than of words.

In addition, we predicted that, based on the assumption that dual codes are elicited for symbols at encoding (like images), participants who studied symbols would experience superior memory even when the recognition test consisted entirely of words. Of course, some reasonable attenuation in performance was expected as a result of the incongruency of stimulus types between encoding and retrieval phases. In essence, all four groups were expected to be significantly different from each other: Both congruent groups should exhibit memory performance superior to that of the two incongruent groups due to transfer appropriate processing (Morris, Bransford, & Franks, 1977), or encoding specificity in the former case (Tulving & Thomson, 1973). Nonetheless, groups that encoded symbols should display better memory performance regardless of the format in which the items were presented on the recognition test.

3.1. Method

3.1.1. Participants

An a priori power analysis was conducted using G*Power software (v. 3.1.9.7; Faul, Erdfelder, Lang, & Buchner, 2007), targeting a medium-sized between-subjects contrast ($d = 0.50$, $\alpha = .05$, two-tailed). This indicated required sample sizes of 64 participants in each of the four groups to achieve 80% statistical power. Accordingly, a goal was set to collect a minimum of 64 participants per group ($N = 256$ total), with the ideal target sample size set higher at 70 participants per group ($N = 280$ total). These sample sizes should also have provided 98% and 99% power to detect medium sized ($f = .25$, $\alpha = .05$, two-tailed) main effects and interactions as well, respectively. In the end, data were collected from 363 University of Waterloo undergraduate students who took part in a single session in exchange for course credit. Participants had self-selected to participate in the study and had self-reported normal or corrected-to-normal vision as well as having learned English before the age of 9.

From this initial sample, participants' data were filtered out in sequential steps if they (1) had corrupted or incomplete data files, or were duplicate attempts ($n = 34$), (2) took <5 min to complete the study ($n = 1$), (3) took >30 min to complete the study ($n = 13$), or (4) were ± 3 *SDs* away from the mean of remaining participants for study duration ($n = 0$). Therefore 315 valid data files entered statistical analyses. Then, *R* statistical software was used to exclude participants in successive steps who were missing more than ten recognition responses (due to rushing through the study causing technical errors; $n = 28$), as well as those whose performance was ±3 *SDs* away from the mean in their group on any memory performance metric (hits, false alarms, or accuracy; $n = 4$) or on the symbol matching task ($n = 10$).[6] The final sample of 273 participants used in the statistical analyses was 78.39% female, with ages ranging from 18 to 48 ($M = 20.1$, $SD = 2.8$; one participant declined to provide their age).

3.1.2. Materials

The same materials as used in Experiment 1 were used here, except that the master stimulus list was expanded to include 80 symbols and their word counterparts (see Fig. 1). Five of the newly added symbols required two words to be used in the words condition (e.g., � = 'fast-forward'). Also, word-based encoding stimuli were now presented in Times New Roman size 24 lowercase black font on a white background.

3.1.3. Procedure

The procedure generally followed that of Experiment 1, except that now—due to the COVID-19 pandemic—the study took place online. The experiment was built using Qualtrics software (https://www.qualtrics.com/) and was administered through Prolific (an online data collection platform; https://www.prolific.co/). As a result, the study was administered on participants' personal computers. This change in setting led to four minor procedural updates: (1) demographic data were collected at the end of the study (rather than at the beginning), (2) the tone classification filler task now involved using as mouse to click on a response of 'low,' 'medium,' or 'high' for each tone instead of responding using keypresses, (3) participants now responded with the 'n' and 'm' keys on the symbol matching task, and (4) an electronic version of the Mill Hill Vocabulary Scale was used.

In addition, to test the robustness of the findings from Experiment 1, Experiment 2 switched to a between-subjects design with an old/new recognition memory test. Because of the change to a 2 × 2 between-subjects design, participants were randomly sorted into one of four groups. In the Symbols-Symbols group participants studied and then were tested using symbol stimuli only ($ ➔ $), in the Symbols-Words group the test stimuli switched to words ($ ➔ dollar), in the Words-Symbols group participants studied words and then were tested using symbols (dollar ➔ $), and in the Words-Words group both the study and test stimuli were words (dollar ➔ dollar).

In each group, 40 items were randomly selected from the 80-item master stimulus list to be presented at encoding. All 80 items were included on the recognition test (40 targets plus the remaining 40 items that served as lures). Participants responded with the 'n' key if the item was 'new' (i.e., not studied previously), or the 'm' key if the item was 'old' (i.e., they remembered it from the study phase). The procedures and materials for this study were approved by the Office of Research Ethics at the University of Waterloo (project #41594). All data, pre-registrations, analysis code, experiment programs, and other materials are listed on the Open Science Framework (OSF; https://osf.io/e53z4/).

3.2. Results

3.2.1. Memory performance

To examine memory performance across groups, a one-way independent-groups analysis of variance (ANOVA) was conducted using the rstatix (v. 0.7.0; Kassambara, 2021) package for *R*, with Group as the independent variable with four levels (Symbols-Symbols, Symbols-Words, Words-Symbols, and Words-Words) and memory accuracy (hit rate minus false alarm rate) as the dependent measure. This analysis revealed a significant main effect of Group, $F(3, 269) = 56.31$, $p < .001$, $\eta_p^2 = .39$, $BF_{10} = 1.21\text{e}{+}25$. Welch-adjusted[7] pairwise comparisons with Holm corrections revealed that all between-group contrasts were statistically significant ($ps \leq .011$, $ds \geq 0.44$; see Table 1).[8]

To examine whether there were differences in encoding or retrieval of symbols and words, two separate Welch-adjusted independent-samples *t*-tests were conducted, each with Holm corrections. First, the effects of the two stimulus sets at encoding were tested by contrasting participants who studied symbols (Symbols-Symbols combined with Symbols-Words) versus those who studied words (Words-Words combined with Words-Symbols). This analysis demonstrated that studying symbols on

---

[6] Retaining data from these participants in the statistical analyses made no difference to the pattern of results.

[7] Welch-corrected *t*-tests are still valid and recommended as a default method for between-subjects comparisons (Delacre, Lakens, & Leys, 2017; Delacre, Leys, Mora, & Lakens, 2020; Ruxton, 2006), even when homogeniety of variance is maintained.

[8] This analysis was also conducted with the pre-registered target sample size of 64 participants per group by randomly selecting data from the full data set. The pattern of results was identical.

Table 1

Memory performance metrics as a function of experiment and condition.

Experiment	Condition	*n*	% Female	Age		Proportion Recalled		Accuracy		Hit Rate		False Alarm Rate	
				M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	Words	128	71.88%	19.70	2.10	0.32	0.13						
	Symbols					0.44	0.15						
2	Words-Words	68	76.47%	19.87	1.65			0.56	0.22	0.72	0.13	0.17	0.15
	Words-Symbols	69	78.26%	20.16	1.95			0.24	0.16	0.47	0.18	0.23	0.15
	Symbols-Words	67	76.11%	19.54	2.15			0.36	0.20	0.64	0.15	0.28	0.13
	Symbols-Symbols	69	82.61%	20.65	4.36			0.66	0.24	0.84	0.13	0.18	0.16
3	Words	92	77.17%	20.77	4.53			0.59	0.21	0.83	0.12	0.24	0.15
	Symbols	97	76.29%	20.35	3.02			0.68	0.22	0.82	0.14	0.14	0.13
4	Words	218	50.46%	33.40	11.80	0.41	0.20						
	Symbols					0.47	0.17						
	Pictures					0.46	0.17						

Note. The demographic values in Experiments 1 and 4 represent single groups of participants that saw each condition.

average led to superior memory than did studying words (see Fig. 3), $t(269.01) = 3.63, p < .001, d = 0.44$, $\mathrm{CI}_{95}[0.20, 0.68]$, $BF_{10} = 62.07$, with very strong Bayesian evidence for the alternative model. Next, the effects of the two stimulus sets at retrieval were evaluated by contrasting those tested with symbols (Symbols-Symbols combined with Words-Symbols) versus those tested with words (Words-Words combined with Symbols-Words). This analysis showed no significant difference in test stimuli format between symbols and words, $t(259.82) = -0.31$, $p = .761$, $d = -0.04$, $\mathrm{CI}_{95}[-0.27, 0.20]$, $BF_{01} = 7.20$, with moderate Bayesian evidence in support of the null model.

3.2.2. Exploratory correlations

Once again, a series of exploratory Pearson correlations was conducted to determine whether English language proficiency or symbol knowledge correlated with memory performance. The results of these analyses revealed that performance on the Mill Hill Vocabulary Scale was significantly correlated with memory accuracy in each group ($rs \geq .29$, $ps \leq .014$), except for the Symbols-Words group ($p = .134$). This effect seemed to be driven by a reduction in false alarms rather than by an increase in hits, as the former measure correlated negatively with Mill Hill score in each group ($rs \leq -.26$, $ps \leq .031$), except for the Symbols-Symbols group where the effect was smaller ($r(67) = .22$, $p = .071$). Hit rate, on the other hand, did not correlate with Mill Hill score ($rs \leq |.14|$, $ps \geq .255$), except for the Symbols-Symbols group ($r(67) = .29$, $p = .012$).

Further analyses demonstrated that performance on the symbols matching task correlated significantly with memory accuracy in in the symbols encoding groups (Symbols-Symbols, $r(67) = .39$, $p < .001$, and Symbols-Words, $r(65) = .35$, $p = .004$), but not in the words encoding groups (Words-Words, $r(66) = .17$, $p = .159$, and Words-Symbols, $r(67) = .22$, $p = .066$). Like the correlations conducted with Mill Hill performance, this effect seemed to be driven by a reduction in false alarms rather than by an increase in hits, as the former measure correlated negatively with symbols matching task performance in each group ($rs \leq -.25$, $ps \leq 0.038$). On the other hand, it only correlated with hit rate in the Symbols-Symbols group, $r(67) = .34$, $p = .004$ (remaining $rs \leq |.19|$, $ps \geq .115$).

3.3. Discussion

Experiment 2 had three primary goals. The first was to determine whether the results of Experiment 1 (i.e., symbols > words) could be conceptually replicated. The second was to ascertain whether any observed effect would be driven by stimulus format at encoding rather than at retrieval. The third was to explore whether there would be a transfer-appropriate processing benefit when stimulus format between encoding and retrieval was congruent as opposed to incongruent. To address these goals, four groups were created that each saw a unique combination of symbols and words presented at encoding and retrieval.

A priori predictions were supported by the data. The findings from Experiment 1 were replicated, plus there was an overall benefit to memory when stimuli were congruent relative to incongruent between encoding and retrieval. That is, a pure symbols-symbols condition led to superior memory performance compared to a pure words-words condition. In addition, the memory benefit for symbols over words clearly appeared to be driven by processes at encoding rather than at retrieval. That is, grouping together types of materials based on what was presented during encoding (words or symbols) showed that, overall, studying symbols led to better memory than studying words, regardless of the format of the later recognition test. In contrast, memory was unaffected by stimulus format on the recognition test. The results of this experiment therefore indicate that the effect of 'symbol superiority' is not only generalizable to between-subjects designs and to recognition testing, but that it also occurs primarily due to differences at encoding rather than at retrieval.

From the outset, a cost to memory performance was predicted to result from switching stimulus format between encoding and retrieval, based on the well-known concept of transfer appropriate processing (Morris et al., 1977). This predicted cost manifested as expected: Engaging in imagery-based (or word-based) processing at encoding benefitted memory on tests that also used imagery (or words) again at retrieval. More pointedly, this cost to memory also aligns with earlier picture superiority work (Mintzer & Snodgrass, 1999; Stenberg, Radeborg, & Hedman, 1995) which demonstrated significant costs when switching between words and pictures.

In this experiment, positive correlations were observed between English language proficiency (as measured by the Mill Hill Vocabulary Scale) and memory performance in each group. It is possible that familiarity with the English language tracks with immersion in North American culture. Therefore, insofar as the symbols used here are prominent in North American societies, it makes sense that higher familiarity with the dominant North American language would also indicate higher familiarity with the current set of symbols and words, which could serve to enhance memory. It is worth noting, however, that the significant correlations between Mill Hill score and memory performance observed in Experiment 2 were non-significant in Experiment 1, forcing one to question the reliability of these particular observations. Nonetheless, the idea that knowledge of a symbol's meaning could in part be driving superior memory performance was also supported in the second set of correlations showing that, whereas performance on the symbol matching task correlated positively with performance for groups that studied symbols, it had no relation to performance for groups that studied words. Most pertinent, hit rate in the 'pure' Symbols-Symbols group was positively related to performance on the symbol matching

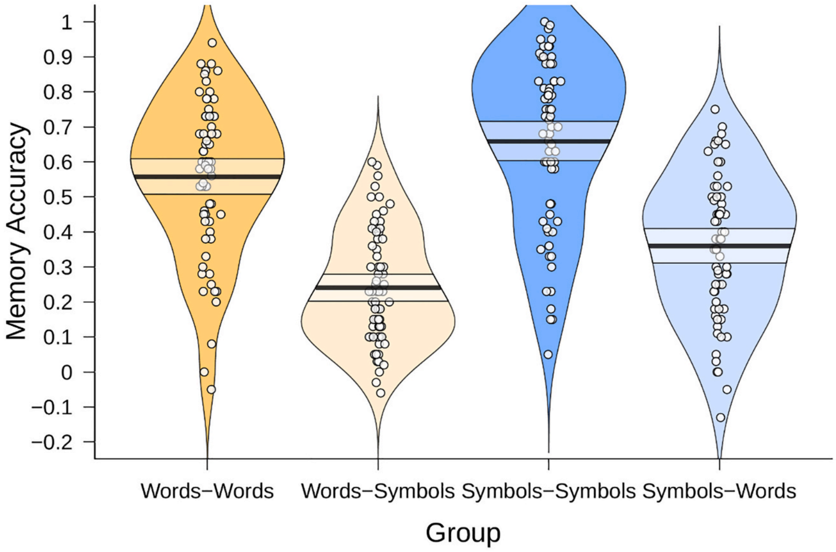

Fig. 3. Experiment 2: recognition in each of the four study-test conditions.
Note. Errors bars = 95% confidence intervals.

task whereas false alarm rate was negatively correlated with the same task. Thus, it is possible that knowledge of a symbol's meaning could predict later memory performance for these items because of an increase in memory for old items as well as better protection from false memories (i.e., from false alarms).

At this point, the basic phenomenon—superior memory for symbols than for words—appeared to be on a solid foundation. A correlated factor could, however, be at play: Perhaps symbols are better remembered because they constitute a smaller overall set size than words. That is, perhaps there are fewer symbols in memory for people to search through relative to words, thus easing retrieval and improving memory for symbols. Because this same argument had been made to explain picture superiority in memory (Nelson, Bajo, & Casanueva, 1985; Nelson, Cañas, Casanueva, & Castaño, 1985), this potential 'set size' confound was addressed in a third experiment.

4. Experiment 3

Following two successful experiments showing the basic effect of superior memory for symbols relative to words, a third experiment investigated potential confounds inherent in the previous two studies. The most prominent alternative explanation was that of set size, which had been used in the past to try to explain superior memory for 'closed' (items from a single category) relative to 'open' (unique items only) sets of words (Coltheart, 1993; Hulme et al., 1997; Poirier & Saint-Aubin, 1995; Roodenrys & Quinlan, 2000). A similar 'set size' explanation was initially put forth to explain picture superiority in memory: Pictures may belong to a smaller set size than words, making pictures easier to search through and select from during a memory test (Nelson, Bajo, & Casanueva, 1985; Nelson, Cañas, et al., 1985; Nelson & McEvoy, 1979). By way of extension to the current study, symbols could be better remembered not because of their preferential encoding but because of eased retrieval processes due to reduced interference or search time resulting from being part of a smaller set than words. To illustrate, consider that any motivated individual likely could state thousands upon thousands of words that they know. In contrast, the same individual likely could come up with fewer than two hundred symbols. The actual values here are arbitrary, but the point is clear: Symbols may be easier to retrieve because there simply are fewer of them.

In Experiment 3, this possibility was tested by reducing the set size of the words stimuli to roughly equate it to that for symbols. Because there were 80 symbols in the stimuli set of the previous study, a closed set of words was needed that contained roughly 80 items that most people would be able to recognize. The taxonomic category 'common kitchen vegetables' satisfied this requirement. In Experiment 3, memory for the set of symbols from the previous experiment was compared to memory for a list of words that are part of the vegetable category.

Because semantics were held constant between stimulus formats in the earlier experiments (i.e., $ vs. dollar), it seemed unlikely that differences in category set size were driving the previously reported memory benefit for symbols. Given this, the prediction for the current experiment was that, due to enhanced encoding for symbols relative to words, the symbol superiority benefit seen previously would persist even when set sizes were equated. Based on well-known concreteness effects in memory, however, the magnitude of the benefit for symbols was predicted to decrease as a result of higher concreteness for the words in the vegetable category, relative to the abstract words used in previous experiments. Experiment 3 also examined whether average familiarity with the current set of symbols would track with memory performance, given that this was the case in Experiment 2.

4.1. Method

4.1.1. Participants

An a priori power analysis was conducted using G*Power software (v. 3.1.9.7; Faul et al., 2007), targeting the smallest effect of interest in this study: a medium effect size of memory accuracy between the words and symbols groups ($d = 0.5$, $\alpha = .05$, two-tailed). This analysis indicated a minimum sample size of $N = 64$ per group or an ideal target sample size of $N = 86$ per group to achieve statistical power of 80% or 90%, respectively. Participants self-selected to participate in the study and had self-reported normal or corrected-to-normal vision, as well as having learned English before the age of 9. A total of 249 University of Waterloo undergraduate students took part in a single session in exchange for course credit.

From this initial sample, participants' data were filtered out in sequential steps if they (1) had made duplicate responses ($n = 20$), (2) had corrupted or incomplete data files ($n = 16$), (2) took <5 min to complete the study ($n = 0$), (3) took >40 min to complete the study ($n = 6$), or (4) were ± 3 *SDs* away from the mean of remaining participants for study duration ($n = 10$). After these exclusions, the sample consisted of 197 participants. Then, *R* statistical software was used to exclude participants whose performance was ±3 *SDs* away from the mean in their group on any memory performance metric (hits, false alarms, or accuracy; $n = 6$), or on the symbol familiarity task ($n = 2$).[9] The final sample in the statistical analyses consisted of 189 participants, 76.72% female, with ages ranging from 17 to 49 ($M = 20.6$, $SD = 3.8$).

4.1.2. Materials

Materials matched those in Experiment 2, except now the word stimuli were no longer word versions of the symbols. Instead, the word stimuli were 80 common kitchen vegetables. Like the word-based stimuli in Experiments 1 and 2 (e.g., dollar), most of the items used here were one word (e.g., 'carrot'); 24 required two words (e.g., 'sweet potato').

4.1.3. Procedure

The procedure was similar to that for the pure words and symbols groups of Experiment 2. Participants were randomly sorted into one of two groups—encode and test symbols or encode and test words. The only difference, other than the changed set of words, was the addition of a 'symbols familiarity rating task', which was completed immediately following the recognition memory test. In this final task, all 80 symbols from the master stimulus list were presented one at a time in the center of the screen. Participants were instructed to indicate their familiarity with each symbol by clicking on one of seven response options along a Likert-style scale from '1: Very Unfamiliar' to '7: Very Familiar', with '4: Not Sure' as a middle response option. In the instructions for this task, a 'very familiar' symbol was defined as "you know what it means, personally use it, and/or see it used frequently." A definition of a 'very unfamiliar' symbol was also provided with the opposite description. The procedures and materials for this study were approved by the Office of Research Ethics at the University of Waterloo (project #41594). All data, pre-registrations, analysis code, experiment programs, and other materials are listed on the Open Science Framework (OSF; https://osf.io/e53z4/).

4.2. Results

4.2.1. Memory Performance

To compare memory performance between words and symbols, a Welch-adjusted independent-samples *t*-test was conducted with memory accuracy (hits minus false alarms) as the dependent measure. Once again, the results showed that memory accuracy was higher for symbols than for words (see Fig. 4),[10] $t(187) = 3.03$, $p = .003$, $d = 0.44$, $CI_{95}[0.15, 0.73]$, $BF_{10} = 10.43$, with strong Bayesian evidence in support of the alternative model.

4.2.2. Exploratory correlations

Correlations were used again to examine whether English language proficiency correlated with memory performance for symbols or words. Differently from Experiments 1 and 2, these analyses failed to yield any significant correlations between Mill Hill scores and measures of memory performance in either condition (hits, false alarms, or accuracy; $rs \leq |.15|$, $ps \geq .154$). Next, we examined whether higher familiarity with the current symbols set would predict greater memory accuracy for symbols used in the memory study. Average symbols familiarity rating was, however, not correlated significantly with measures of memory performance in either condition (hits, false alarms, or accuracy; $rs \leq |.09|$, $ps \geq .391$).

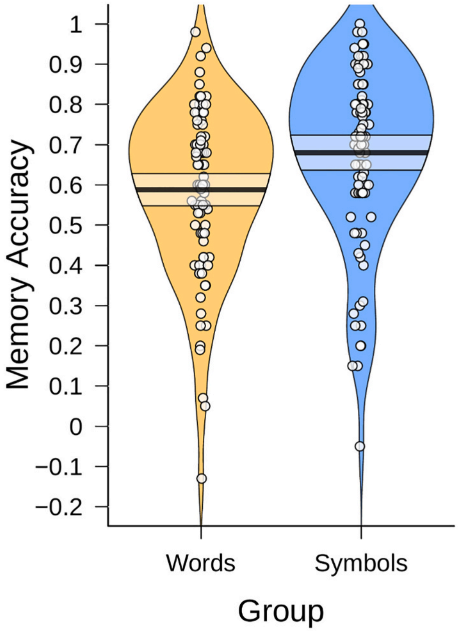

Fig. 4. Experiment 3: memory performance for symbols and words.
Note. Errors bars = 95% confidence intervals.

4.3. Discussion

In this experiment, a potential confound of set size was addressed in accounting for the prior results showing that symbols are better remembered than words. To do so, the set sizes of word and symbol stimuli were roughly equated by constraining the studied words to those in the vegetable category. The results here echoed those of the previous two experiments: Symbols were still better remembered than words. Contrary to our predictions, however, higher familiarity with symbols did not correlate with better memory for the same symbols.

The findings of this experiment align well conceptually with those of Experiment 2 which showed that the effect is likely encoding-based, and therefore any easing of retrieval processes due to a smaller set size for symbols would likely be inconsequential. Thus far, it has been consistently shown that symbols have led to superior memory performance both relative to their word counterparts and now relative to words from a closed set of equivalent size.

Here, we eliminated the concreteness effect in memory by representing abstract concepts using symbols rather than words. That concrete words from the vegetable category were less well remembered than graphic symbols representing abstract concepts, suggests that encoding of symbols more reliably leads to image codes in memory relative to the spontaneous imaging thought to occur with concrete words (Paivio & Csapo, 1973). This finding would be predicted by those who subscribe to Paivio's dual-coding theory: If the current set of symbols is indeed a set of 'mini-pictures,' then they should serve to 'concretize' their associated abstract concepts. That is, if concrete words lead to better memory because they are more easily imageable, as Paivio argued, then pictures must be inherently concrete. Therefore, rendering abstract concepts in symbol format may serve to make the to-be-remembered content imageable, thereby bringing about dual representations in memory despite the underlying association still being abstract. If symbols are indeed afforded dual codes in memory, akin to pictures, then memory for symbols should be similar to memory for images of even concrete objects. This prediction was directly tested in the following experiment.

---

[9] Retaining data from these participants in the statistical analyses made no difference to the pattern of results.

[10] This analysis was also conducted with the pre-registered target sample size of 86 participants per group by randomly selecting data from the full data set. The pattern of results was identical.

5. Experiment 4

Recall that Paivio's classic dual-coding theory postulates that pictures are better remembered because they routinely evoke two codes in memory: verbal and image. If one were to assume based on the results of Experiments 1 through 3 that symbols are also reliably dual-coded, then symbols should lead to memory superior to that for words (as has now been demonstrated) but equivalent to that for pictures. Critically, all pictures are inherently concrete insofar as they depict something visually and therefore immediately provide an image code in memory (Paivio, 1969). For images of easily recognizable content (e.g., a picture of an apple), spontaneous verbal labelling should also occur (Paivio, 1969). Importantly, however, there should be no extra benefit to memory for images that depict physical things relative to those that represent abstract concepts (such as symbols). That is, the concreteness effect should be eliminated for familiar visual stimuli: When pictures are compared to other recognizable graphics (e.g., symbols), both should be highly likely to elicit dual codes in memory, regardless of the concreteness of their underlying concepts.

It is important, however, to disentangle the related ideas of 'abstract pictures' and 'pictures representing abstract concepts.' Whereas the former may take the form of meaningless geometric shapes or patterns, the latter must be identifiable as related to their underlying abstract concepts. Studies have previously found that images of concrete things are better remembered than are abstract pictures (Bellhouse-King & Standing, 2007; Smith, Park, Cherry, & Berkovsky, 1990; Vogt & Magnussen, 2005). This makes sense from a dual-coding perspective because the abstract material is less readily labelled and therefore may lack a verbal code in memory. Symbols, on the other hand, may represent an intermediate category of 'pictures representing abstract concepts,' such that the image could be labelled but the underlying association is still abstract. Nevertheless, if dual coding of such stimuli is important for memory benefits, then memory for symbols should be equivalent to that for pictures of concrete stimuli such as objects because both should evoke imaginal and verbal representations. Here, this hypothesis was tested directly.

5.1. Method

5.1.1. Participants

An a priori power analysis was conducted using G*Power software (v. 3.1.9.7; Faul et al., 2007), targeting the smallest effect of interest in this study: a small- to medium-sized within-subject effect for the Symbols vs. Pictures pairwise comparison ($d = 0.2$, $\alpha = .05$, power = 80%, two-tailed) as measured by free recall performance. This analysis indicated a minimum target sample size of $N = 199$. This target sample size would also provide adequate power to detect a small overall main effect of condition as well (target $N = 163$; $f = .10$, $\alpha = .05$, power = 80%, two-tailed). A small effect of Pictures vs. Symbols was targeted because there is no literature on the subject to date, and the hypothesized effect could be small if variation in the quality or fidelity of the image in memory is the only difference between these conditions.

Once again, participants were recruited using the Prolific data collection website. Filters were applied to allow participants to sign-up for the study only if they (1) were between the ages of 18 and 64, (2) were living in Canada or the USA, (3) were fluent in English, and (4) had normal or corrected-to-normal vision. Eligible participants self-selected to join the study. A built-in balancing service was used to ensure equivalent numbers of male and female participants in this particular experiment. A total of 241 participants took part in a single 25-min session in exchange for $4.12 USD.

From this initial sample, participants' data were filtered out in sequential steps if they (1) duplicated responses ($n = 0$), (2) had self-reported non-ideal conditions (distractions) while completing the experiment ($n = 10$), (3) took <5 min to complete the study ($n = 0$), (4) took >40 min to complete the study ($n = 8$), or (5) were ± 3 *SDs* away from the mean of the remaining participants for study duration ($n = 5$). These exclusions resulted in a sample of 218 participants. *R* statistical software was used to exclude participants whose performance was ±3 *SDs* away from the mean in their group on any memory performance metric (hits, false alarms, or accuracy; $n = 0$). The final sample of 218 participants used in the statistical analyses was 50.46% female, with ages ranging from 18 to 64 ($M = 33.4$, $SD = 11.8$).

5.1.2. Materials

Picture stimuli were 50 separate line drawings in black ink, representing small, everyday objects, presented on white backgrounds; all were sourced from the International Picture Naming Project (IPNP) database (Szekely et al., 2004). To select images of common, easily identifiable objects from the IPNP, the following sequential sorting procedure was implemented: (1) only images from the 'small artefacts' category were selected, then (2) images were sorted from the least to the most number of alternative object names (etype), then (3) images were sorted from the most to the least percentage name agreement (elex1), then (4) images were sorted by highest to lowest CELEX frequency (efreq), and finally (5) any 'outdated'(e.g., an old radio) or vague images (including ones that were hard to discern or that contained multiple objects; e.g., a table set with a fork, knife, napkin, and plate) were removed. After this sorting procedure, the top 50 images were retained for use in this study. Pictures were then re-sized to match the dimensions of the symbols stimuli (110 × 116 pixels). The word-based encoding stimuli were the word counterparts of the images sourced from the IPNP. The symbols stimuli were the same set of 50 items as had been used in Experiment 1.

5.1.3. Procedure

The procedure generally followed that of Experiment 3, except for the following changes: (1) a third condition was added that included pictures of objects, (2) word stimuli were now labels of the selected pictured objects, (2) the symbols familiarity rating task and the Mill Hill Vocabulary Scale were no longer administered, and (3) there was now a final attention check question that asked participants whether they completed the experiment in 'ideal' conditions (they were not distracted, etc.). Finally, this experiment used a within-subject design.

Participants studied and were subsequently tested on three types of stimuli: pictures, words, and symbols. Stimuli were studied and tested in blocks, with block order randomized for each participant. During the picture and the word study phases, each participant was randomly assigned one of two 25-item stimulus sets. These two sets were counterbalanced such that no participant was shown the picture *and* word versions of the same item from the IPNP. To match the pictures and words blocks, symbols stimuli were randomly sorted into two lists of 25 items (only one of which was randomly assigned to each participant). During each encoding block, each participant was presented with 20 items, selected randomly from these master lists.

Following the tone classification filler task (as in Experiments 2 and 3), participants completed a free recall memory test for two minutes. Here, they were told to type out as many words, picture labels, or symbol labels as they could remember from the preceding study phase. Participants were encouraged to type out a physical descriptor of the remembered picture or symbol if they were unaware of its label. The procedures and materials for this study were approved by the Office of Research Ethics at the University of Waterloo (project #41594). All data, pre-registrations, analysis code, experiment programs, and other materials are listed on the Open Science Framework (OSF; https://osf.io/e53z4/).

5.2. Results

5.2.1. Memory performance

Two naïve research assistants scored all recall data, determining whether each response was an intrusion (code = 0), a close response (code = 1), or an exact response (code = 2). As a reminder, in this experiment, participants typed their recall so labels and physical descriptions of symbols and pictures were considered exact responses whereas synonyms were considered close responses. Then, three related metrics were tabulated based on these scores. First, a set of 'lenient' scores was calculated whereby all codes of 1 and 2 were counted the same as correctly recalled items. Next, a set of 'strict' scores was formed whereby only codes of 2 were counted as correctly recalled items. Finally, a weighted scheme was created whereby close and exact codes maintained their values (1 and 2, respectively), but to accommodate this the denominator was changed to 40 (rather than 20). To ensure the most conservative test of the current hypotheses, the lenient scoring criteria was used throughout the results reported here, but additional analyses confirmed that the other weighting schemes made no difference to the overall pattern of effects.

To examine memory performance across conditions, a one-way repeated-measures ANOVA was conducted using the rstatix package for *R*, with Condition as the independent variable with three levels (Symbols, Words, and Pictures) and proportion of items recalled as the dependent measure. This analysis revealed a significant main effect of Group (see Fig. 5), $F(2, 651) = 6.38$, $p = .002$, $\eta_p^2 = .019$, $BF_{10} = 7.37$, with strong Bayesian support for the alternative model.[11] Paired-samples *t*-tests showed that recall performance was higher in the Symbols and Pictures conditions than in the Words condition ($t(217) = 4.04$, $p < .001$, $d = 0.27$, CI_{95} [0.14, 0.41], $BF_{10} = 1.77e+02$, and $t(217) = 3.90$, $p < .001$, $d = 0.26$, CI_{95} [0.13, 0.40], $BF_{10} = 1.06e+02$, respectively), but that the memory in the Symbols and Pictures conditions did not differ, $t(217) = 0.72$, $p = .475$, $d = 0.05$, CI_{95} [−0.08, 0.18], $BF_{01} = 10.25$, with strong Bayesian evidence for the null model in this final comparison.

5.3. Discussion

The purpose of Experiment 4 was to examine whether memory is comparable for symbols and pictures. The prediction was that if symbols and pictures share similar underlying representation formats then their memory performance should be equivalent, with both being superior to words. The results supported this prediction: Symbols and pictures were both better remembered than words and recall in the former two conditions was practically identical. That symbols were remembered better than words, but equally well to pictures of concrete objects, is suggestive of an all-or-none boost for images in memory. That is, there was no further enhancement to memory as a result of higher concreteness for pictures of objects relative to symbols. This implies that symbols were already serving the function of concretizing their underlying abstract concepts.

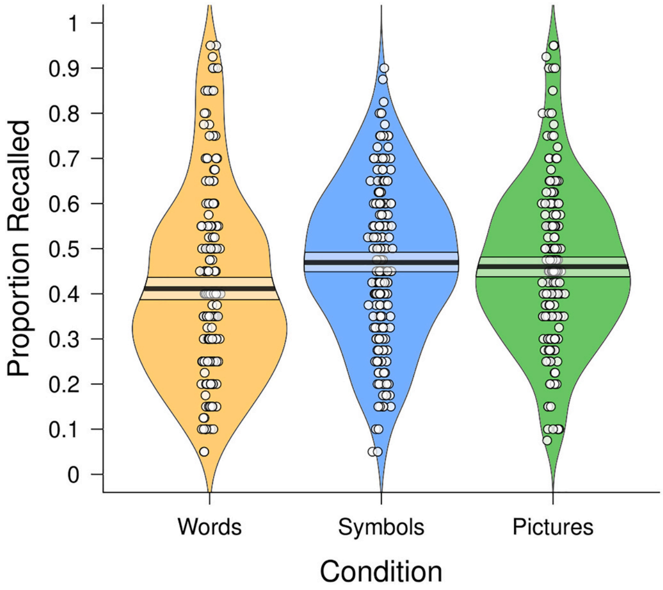

Fig. 5. Experiment 4: recall of words, symbols, and pictures.
Note. Errors bars = 95% confidence intervals.

One potential limitation unique to this experiment is that of possible item-selection effects. In a typical picture superiority effect study, pictures (e.g., a picture of a dog) are compared with corresponding words (e.g., the word 'dog'). By definition, however, there are no pictures that are semantically identical to symbols because the former depict physical things whereas the latter depict abstract concepts. This makes counterbalancing symbols and pictures impossible. It is conceivable, therefore, that because we could not counterbalance content between symbols and pictures, we may have inadvertently chosen a set of images that lead to particularly poor memory or, conversely, we may have selected a set of symbols that are especially memorable. To combat this limitation, we opted to use concrete nouns and their exact associated images to give pictures and words the best chance of being remembered. Yet, we still found the predicted pattern of results such that memory performance for symbols and pictures was practically identical and both were superior to words. In prior work that compared words to sounds, Ensor, Bancroft, and Hockley (2019) ran into a similar item-selection issue. Following a series of experiments, they concluded that item-selection effects were not likely biasing their results. Given the strong arguments against item-selection effects presented by Ensor, Bancroft, and Hockley (2019) when in a similar predicament, and that we chose stimuli that should bolster picture and word memory, we do not believe item selection effects were the main determinant of symbols superiority in memory found here.

That pictures of abstract content (e.g., random shapes and/or patterns) are often remembered worse than pictures of concrete content (e.g., scenes or objects; Smith et al., 1990) speaks to the importance of being able to identify and label the content within a picture. That said, the results of this study, in general, are also consistent with both conceptual (Hamilton & Geraci, 2006; Nelson et al., 1977) and physical (Ensor, Surprenant, & Neath, 2019; Mintzer & Snodgrass, 1999) distinctiveness accounts of picture superiority in memory, accounts that will be discussed further in the General Discussion.

6. Experiment 5

Thus far, it has been argued—largely based on dual coding theory—that it is possible that symbols elicit image codes in memory which effectively serve to 'concretize' abstract concepts by providing them with visual referents. In other words, it has been posited that symbols allow for imagery-based processing of concepts that would otherwise only invoke verbal processing. To provide a further test of a dual coding account for symbol (and picture) superiority in memory, Experiment 5 focused on the verbal memory code.

Symbols are only effective communicators of information when a person already knows, or can figure out, what they are intended to convey. For practical reasons, then, it is preferable for symbols to be highly standardized and better yet for them to have intuitive interpretations. But what if someone has never seen a symbol before or cannot figure out what it means? In that case, the symbol would fail to elicit a verbal code in memory because the viewer would not have a verbal label ready to apply. Therefore, familiarity with a symbol is likely key to adding the second code.

In the case of pictures, on the other hand, researchers often use images of objects and therefore need not be concerned with how familiar a

---

[11] This analysis was also conducted with the pre-registered target sample size of 199 participants by randomly selecting data from the full data set. The pattern of results was identical.

person is with the imaged content because the pictured objects typically are all highly familiar and easily labelable. The prediction is that memory for symbols should improve with greater familiarity. To access a verbal code, it helps to be able to readily retrieve the meaning and corresponding verbal label for a symbol.

6.1. The influence of familiarity

In the fifth and final experiment, familiarity ratings were gathered from naïve participants for the same set of symbols that had been used in the four preceding experiments, the goal being to assess whether average familiarity with a symbol would correlate with its average memorability in the current experiments. Research from the field of psycholinguistics has measured 'familiarity' in at least two distinct forms: 'meaning familiarity' and 'frequency of occurrence.' For instance, Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004); Balota, Pilotti, and Cortese (2001) demonstrated that traditional measures of familiarity are strongly associated with meaning, whereas subjective estimates of frequency are more related to actual frequency of occurrence. Keeping with current psycholinguistic conceptualizations of familiarity, two distinct but related conceptual components were studied: meaning-based familiarity (Juhasz, Lai, & Woodcock, 2015) and subjective frequency of occurrence (Balota et al., 2001). For the former measure, participants were asked to rate how well they knew the symbol; for the latter measure, they were asked how often they had encountered it.

Based once again on dual coding theory, it was reasoned that meaning-based familiarity with a symbol would increase the likelihood of it being dually encoded. The prediction was that high familiarity with a symbol would make labelling it easier and would increase the likelihood of eliciting an additional verbal code in memory. Higher rates of dual coding were expected and therefore a positive correlation was predicted between symbol familiarity and memory performance for symbols. As a result, this analysis included data from all participants who encoded symbols in the four preceding experiments.

Although the main predictions are centered in meaning-based familiarity, it is also possible that subjective estimates of frequency could provide a window into how reliably a symbol can be identified or imaged. For instance, the '#' symbol can be found on the top row of most English QWERTY keyboards and is encountered frequently on social media sites, but its meaning could be relatively ambiguous to participants, especially when taken out of context. It is in cases such as these that one would expect subjective frequency estimates to correlate positively with memory performance because, while the true meaning of the symbol is somewhat variable, it is still frequently encountered and consequently a participant—without knowing the meaning of the symbol—could very well be able to label it (e.g., 'hashtag'). By determining whether familiarity and/or frequency correlate with memory performance, one can obtain a glimpse into the underlying mechanisms driving encoding of symbols insofar as verbal labelling or understanding are required for dual coding.

6.2. The influence of inherent memorability

Recent work has demonstrated that all stimuli have some degree of inherent memorability that can be separated from the cognitive and neural signatures of attention, priming, and low-level perception (Bainbridge et al., 2017, 2013). To investigate whether the symbols in the set of stimuli used across the previous experiments were better remembered simply because of their intrinsic memorability properties, the freely available ResMem neural network (Needell & Bainbridge, 2022) was enlisted to assign memorability scores to each of the symbols. ResMem is a deep residual neural network that was built upon two existing models—ResNet-152 (Khosla, Raju, Torralba, & Oliva, 2015) and AlexNet (Krizhevsky, Sutskever, & Hinton, 2012)—before being retrained for the purpose of optimizing predictions of memorability. ResNet-152 was originally trained for the purpose of classifying images into semantic categories using the ImageNet (Deng et al., 2009) dataset of over 14 million pictures. AlexNet, on the other hand, was trained using 1.2 million images to extract features of low-level perceptual attributes in order to classify pictures (Krizhevsky et al., 2012). As a result of the combination of these two networks, the new ResMem network that was employed here makes use of both semantic and perceptual features of the image to generate memorability predictions (Needell & Bainbridge, 2022).

Although ResMem has not been explicitly trained on or validated with symbols, it has been shown to work well with highly similar visual stimuli from the same semantic category that make heavy use of black and white shading (the Food Folio dataset; Lloyd et al., 2020). In brief, ResMem takes into account perceptual features as well as conceptual features of the image (such as category) to provide a single value, ranging from 0 to 1, that indicates the item's estimated likelihood of being remembered by a person. The creators of this neural network have suggested that within the ResMem network, semantic features may contribute more variance to memorability scores than do perceptual features (Needell & Bainbridge, 2022).

The use of ResMem in exploratory analyses was motivated by the question of whether symbols have intrinsic properties that make them particularly memorable, and whether these properties are separable from participants' ratings of familiarity and frequency of occurrence. From the previous experiments, it was expected that the everyday symbols used here would be classified as highly memorable, and that—based on previous work on memorability (e.g., Bainbridge et al., 2017)—the inherent memorability of these symbols would contribute unique variance to their memorability, apart from that contributed by familiarity and frequency.

6.3. Method

6.3.1. Participants

For this experiment, no a priori power analysis was conducted as there was no targeted effect size of interest. Instead, the target sample size was set to match the largest sample size that had been collected in the previous experiments (target $N = 315$). This target sample size also substantially exceeds the 30–100 participant ratings per item that are often found in psycholinguistic norming research with rating scales (e.g., Balota et al., 2001).

Once again, participants were recruited from Prolific. Several restrictions were set to match the sample characteristics of the previous experiments. Participants could sign up for the study if they (1) were between the ages of 18 and 26, (2) lived in Canada or the USA, (3) were fluent in English, (4) had normal or corrected-to-normal vision, and (5) were currently undergraduate students. Finally, an attempt was made to roughly match the average percent of female participants across the preceding experiments (75.89%). Eligible participants self-selected to join the study. A total of 350 participants took part in this single 9-min session in exchange for $2.13 USD.

From this initial sample, participants' data were filtered out in sequential steps if they (1) were duplicate responses ($n = 0$), (2) took <3 min to complete the study ($n = 0$), (3) took >40 min to complete the study ($n = 0$), (4) were ± 3 *SDs* away from the mean of remaining participants for study duration ($n = 8$), or (5) had self-reported non-ideal conditions for their participation in the experiment ($n = 5$). These exclusions resulted in a final sample of 337 participants who were entered into the statistical analyses, consisting of 76.26% females, with ages ranging from 18 to 26 ($M = 20.9$, $SD = 1.8$).

6.3.2. Materials

The full set of 80 symbols used in Experiments 2 and 3 (see Fig. 1) was used here. As in the preceding studies, symbols were sized at 110 × 116 pixels using Segoe UI Symbol black font on a white background. Participants rated each symbol using two different conceptualizations of familiarity with an item—one for meaning-based familiarity and one for frequency of occurrence—each with a different 7-point scale.

The measure of meaning-based familiarity that was used, including the response options and instructions, was adapted from Juhasz et al. (2015). Response options ranged from (1) Very Unfamiliar to (7) Very Familiar. Intermediate response options were unlabelled. The instructions for this scale were as follows:

> “Please provide a rating between 1 (very unfamiliar) to 7 (very familiar). If you feel you know the meaning of the symbol and use it frequently, then give it a high rating on this scale. For example, Jim has known the symbol * (asterisk) since he was a child, uses the symbol frequently, and if asked could easily tell anyone what it is. He should give this symbol a very high rating. If the symbol is not familiar at all, you do not know its meaning, or you are not sure whether it is a symbol or not, then give it a low rating. For example, Jim has never encountered the symbol № (numero) and has no sense of whether it is a symbol or not, or for what it means. He should give this symbol a very low rating. If the item falls somewhere in the middle of these two extremes, where you have some familiarity with the symbol, then give it a rating in the middle of the scale.”

The measure of subjective frequency that was used, including the response options and instructions, was adapted from Balota et al. (2001). Responses once again were made on a 7-point scale. The instructions for this scale were as follows:

> “Symbols differ in how commonly or frequently they have been encountered. Some symbols are encountered very frequently, whereas other symbols are encountered infrequently. The purpose of this scale is to rate each symbol with respect to the frequency you encounter it. You should base your ratings according to the following 7-point scale: 1 = never, 2 = once a year, 3 = once a month, 4 = once a week, 5 = every two days, 6 = once a day, 7 = several times a day.”

6.3.3. Procedure

Following informed consent, participants were told that they would be rating a series of symbols, one at a time, on two different dimensions (familiarity and frequency). Participants were then provided with definitions of each scale prior to the rating task.

During the rating task, each symbol was presented in the center of the screen with each of the two scales below it, spanning horizontally across the screen. Participants responded by clicking on one of the radio buttons of each scale, at which point the experiment program automatically and immediately proceeded to the next rating trial. Following presentation of all 80 symbols, participants were asked whether their participation occurred under ‘ideal’ conditions (as in Experiment 4). They were then provided with a detailed feedback letter. The procedures and materials for this study were approved by the Office of Research Ethics at the University of Waterloo (project #41594). All data, preregistrations, analysis code, experiment programs, and other materials are listed on the Open Science Framework (OSF; https://osf.io/e53z4/).

6.4. Results

6.4.1. Correlations with average familiarity ratings

To determine whether memory performance for symbols correlated with average ratings of symbol familiarity and frequency, memory data from the four preceding experiments were aggregated. For each item, average proportions of free recall for symbols (in Experiments 1 and 4) and of old/new recognition hit rate for symbols (in Experiments 2 and 3) were collected.

In Experiments 2 and 3, half of the symbols were randomly presented at encoding. Because of this, hit rate for those experiments was calculated as number of hits divided by the number of times the symbol was presented at encoding. In the end, every symbol was rated by each of the 337 participants in the current experiment. Memory performance data were aggregated from 579 participants across the four preceding experiments.

A series of Pearson correlations was conducted to compare memory performance, average ratings of familiarity, and average ratings of frequency for each of the 80 symbols. As expected, familiarity ratings for the current set of symbols along a 7-point scale were quite high ($M = 5.74$, $SD = 1.42$); frequency ratings were more moderate ($M = 4.06$, $SD = 1.42$). Average memory performance for symbols did not correlate with average familiarity ratings of symbols, $r(78) = .03$, $p = .776$ (see Fig. 6A), or with frequency of encountering symbols, $r(78) = .01$, $p = .945$.

6.4.2. Predictors of memory for symbols

Submitting the full 80-item set of symbols to ResMem indicated that the stimuli that were used in the current experiments were highly memorable, with scores ranging from 0.83 to 0.98 ($M = 0.91$, $SD = 0.03$, on a scale from 0 to 1). It makes intuitive sense that scores would be rather high: Everyday symbols often are designed to be simple in terms of their low-level visual features, to be easily identifiable, and to be pervasive in visual communication media. Memorability estimates for symbols provided by ResMem, rather unsurprisingly, correlated positively with memory performance for symbols in Experiments 1–4, $r(78) = .23$, $p = .038$ (see Fig. 6B). When word versions of the same symbols set were input into ResMem, each item received a score of exactly 0.694, suggesting that the model could not differentiate memorability of words.

Finally, a hierarchical linear regression was conducted to determine whether the intrinsic memorability scores assigned to the symbols by ResMem could predict memory performance over and above contributions stemming from familiarity ratings or subjective frequency estimates. First, familiarity and frequency ratings for the set of symbols were entered into the model; second, scores from the ResMem network for the same symbols were added. Results indicated that the second model predicted significantly more variance than the first, $\Delta R^2 = .05$, $F(1, 76) = 4.17$, $p = .045$, suggesting that intrinsic memorability properties were still able to predict memory performance beyond familiarity ratings and frequency estimates. This analysis highlights that the inherent memorability of symbols, stemming from their perceptual qualities, predicts later memory over and above the combined contributions of one’s personal familiarity with the symbols and the frequency with which these symbols are encountered.

6.5. Discussion

This fifth and final experiment sought to assess whether familiarity, frequency, or inherent memorability influence memory for symbols. A dual coding explanation for picture superiority in memory rests entirely on the notion that images are more likely to be spontaneously labelled than words are to be imaged. This is also often touted as the reason that concreteness effects occur in memory: Concrete words are more easily imageable than abstract words. Based on this idea, the current experiment tested an extrapolation from dual coding theory—that as a person’s familiarity with a concept increases, so too should their likelihood of producing a mental label when presented with a symbol. Hence, higher ratings of familiarity were expected to correlate with higher rates of dual coding and therefore better memory performance.

Our predictions, however, were not supported in this study: The correlation between familiarity and memory performance collapsed across the four preceding experiments was not statistically significant. Subjective estimate of frequency also failed to exhibit a significant relation with memory performance suggesting that the number of encounters with a given symbol is perhaps inconsequential to memory. These findings are at odds with extrapolations from dual coding theory as they indicate that knowing what is being imaged, so as to obtain a verbal memory code, is not a determinant of later memory for graphic

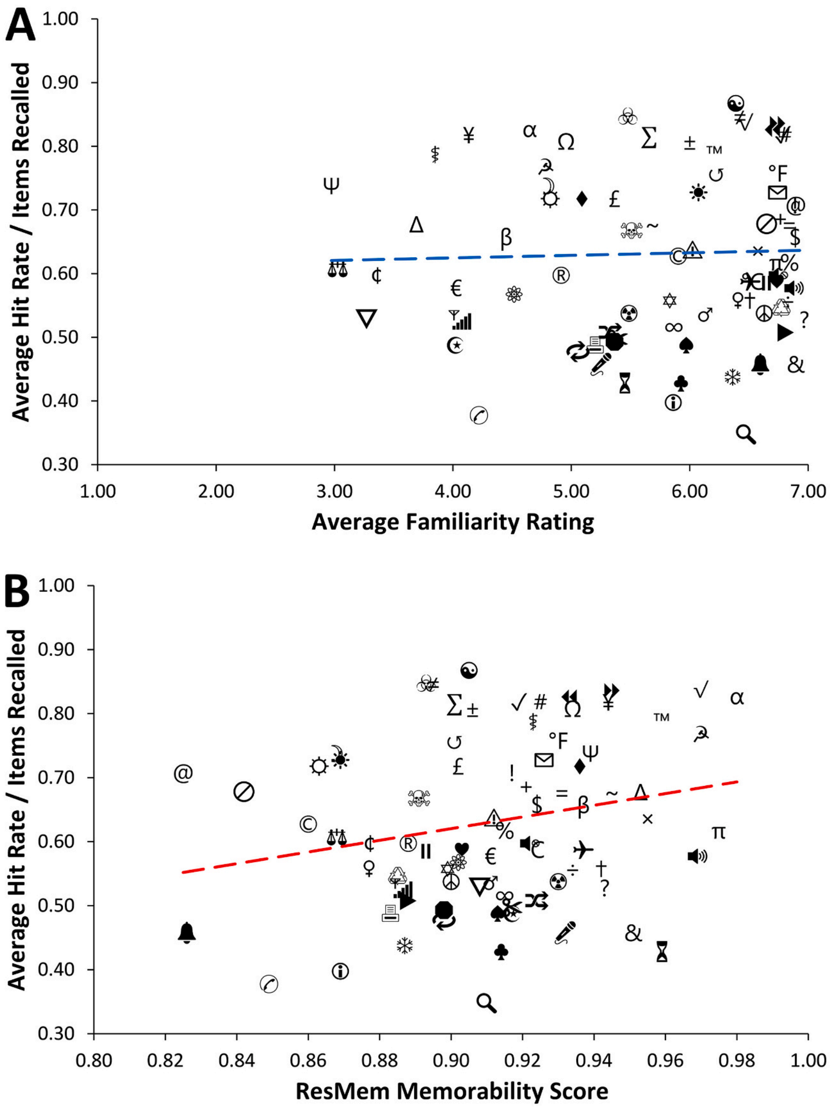

Fig. 6. Memory performance scores by familiarity rating and ResMem score.
Note. Memory performance on the y-axes of panels A and B, as well as the x-axis of panel B, were on scales from 0 to 1 but are truncated here for the sake of clarity. Linear trendlines are included in each panel.

symbols. One might predict that without this knowledge, trying to remember a meaningless shape would prove difficult, much like the studies of abstract pictures mentioned previously. This prediction, however, was not supported in the present case. Overall, our data suggest that dual coding theory is perhaps limited to the extent that it cannot explain memory for graphic stimuli as a function of familiarity.

The use of ResMem here was motivated by the question of whether symbols have intrinsic properties that make them highly memorable. Using this neural network allowed us, via computational means, to quantify how memorable the stimulus set was and to determine whether the intrinsic memorability of the stimuli was separable from participants' ratings of familiarity and frequency of those same stimuli. As one might expect, the common everyday symbols used here were determined to be quite memorable by ResMem. The results of regression analyses suggest that this high inherent memorability contributes to memory performance over and above participants' familiarity with the symbols and their frequency of occurrence. Despite the analyses based on ResMem scores having been purely exploratory, the findings are in line with past literature showing that memorability is a perceptually-based phenomenon that is largely separable from any previous memory or experience with the item (Bainbridge et al., 2017).

7. General discussion

This study investigated the cognitive mechanisms underlying representations of graphic symbols in memory. The major predictions—based on parallels drawn between symbols and pictures—were that symbols would be remembered better than words and that they would be remembered as well as pictures. These predictions, although rooted in Paivio's (1969, 1971) dual coding theory of memory, would have been quite similar if one were to begin instead with a physical distinctiveness account (Mintzer & Snodgrass, 1999). The core hypothesis is that graphic symbols behave like pictures in terms of their memorability: They serve as convenient ways to 'image' abstract words without relying on other concrete referents. Given this, memory for graphic symbols should match that of pictures, even if the pictures are of concrete (rather than abstract) objects. Hence, encoding of symbols should eliminate the often-observed performance decrement for abstract words that likely stems from their low imageability.

Across the first four experiments, it was consistently demonstrated that graphic symbols are indeed better remembered than words. Experiment 1, using a within-subject design, demonstrated the basic finding that recall of symbols was greater than recall of their word counterparts. Experiment 2 showed that the effect is generalizable to a between-subjects design and to recognition testing. Experiment 3 showed that the symbol superiority effect (better memory for symbols than words) is robust, regardless of concreteness and set size of the word comparators. Experiment 4 pitted symbols, words, and pictures directly against each other and found that, even when pictures and words were of concrete objects, memory was still better for symbols than for words and was on par with memory for pictures.

Experiment 5 combined memory data from all four preceding experiments with new human ratings of familiarity and memorability scores derived using the ResMem deep neural network (Needell & Bainbridge, 2022). This experiment demonstrated that symbol superiority in memory is likely determined in part by potent inherent memorability characteristics for symbols. Next, an account for the present findings based in dual coding theory is presented before discussing other possible factors contributing to superior memory for symbols.

7.1. Dual coding and distinctiveness accounts

From the outset, we reasoned that symbols may be processed akin to pictures. Our prediction was originally couched in a dual coding explanation whereby symbols serve to concretize abstract words by providing otherwise absent visual referents. Although our logic was originally derived from Paivio's (1971) dual coding theory, distinctiveness accounts (Hamilton & Geraci, 2006; Mintzer & Snodgrass, 1999) of picture superiority are equally well supported by our data. According to more recent work (e.g., Ensor, Surprenant, & Neath, 2019), however, it could also be the case that whether dual coding or distinctiveness mechanisms are at play depends on the type of retrieval test used. Next, we consider this possibility based on the present data.

Recall that Paivio's (1969, 1971) dual coding theory of memory posits that pictures are better remembered than words because pictures benefit from two codes in memory—verbal and image—whereas words rely only on a verbal code. Paivio's theory rests entirely on the notion that people are more likely to spontaneously label images than they are to spontaneously create mental images of the referents of words. Moreover, words that are more easily imageable will be more likely to be imagined automatically, thus increasing their probability of dual coding (and hence of better memory; Paivio et al., 1994). In this study, it was predicted that symbols would be better remembered than words because, like pictures, symbols would elicit dual codes in memory. That is, upon being viewed, a symbol would first provide an image and thereafter would likely be spontaneously labelled. As a result, memory should be better for symbols relative to words, but memory performance for symbols and pictures should be equivalent.

Across the first four experiments, consistent support for these predictions was found. Graphic symbols were better remembered relative to (1) their word counterparts, (2) concrete words from a constrained set, and (3) concrete words representing highly imageable objects. The fourth experiment also confirmed that memory performance for symbols representing abstract concepts was equivalent to that for pictures of concrete objects. These results align precisely with dual coding theory for three primary reasons. First, memory was superior for symbols than for their abstract word counterparts, suggesting that—with semantics held constant—symbols were providing something additional to improve memory. What could this additional factor be? In line with dual coding theory, symbols are provided an extra image record in memory, as is the case for pictures. That Experiment 2 pointed to the encoding phase as the locus of the symbol superiority effect in memory also suggests that improved encoding—rather than eased retrieval—drives the benefit of symbols. The notion of enhanced encoding of symbols aligns with the prediction that they elicit dual codes in memory upon being viewed during the study phase. Finally, because symbols were as well remembered as black and white line drawings of objects, the two types of stimuli likely are similar in terms of the cognitive processes used to remember them.

What is more, there is evidence that dual coding occurred in Experiments 1 and 4 in which a free recall test was used. Ensor, Surprenant, and Neath (2019) suggest that the picture superiority effect might manifest due to dual coding on tests of free recall, as verbal reporting of recalled images would require access to the verbal label of the studied item (the 'logogen pathway', as Paivio, 1991 would call it). To illustrate this point, consider that, in Experiments 1 and 4, similar sets of symbols were used but the words being compared differed. Experiment 1 used abstract words that were counterbalanced with the symbols set, whereas Experiment 4 used concrete words that were counterbalanced with the images used in that experiment. Comparing the effect size for symbol superiority in memory for each experiment, it was clearly larger in Experiment 1 ($d = 0.77$) than in Experiment 4 ($d = 0.27$). As shown in Table 1, this difference in memory outcomes was due to average word recall performance rising in Experiment 4 (from 0.32 to 0.41) while symbol recall performance was relatively unchanged in the two experiments (from 0.44 to 0.47). It stands to reason that the effect size in Experiment 4 was reduced as a result of encoded words now being concrete (rather than using abstract words, as was the case in Experiment 1). Paivio and Csapo (1969) would predict precisely this finding based on dual coding theory, as concrete words are more likely than abstract words to elicit spontaneous imagery, resulting in better memory (i.e., the concreteness effect). Thus, there is evidence indicating dual coding was at play in our experiments where free recall testing was used.

Recall that the two main types of distinctiveness accounts concern physical and conceptual aspects of an item. With respect to the conceptual version, symbols could be better remembered because, relative to words, there are fewer possible lures (i.e., distracting items) available for symbols. Intuitively, it is important that symbols be unique both in form and in meaning, lest their semantic or visual spaces become so overcrowded that their ability to represent general abstract concepts becomes diluted or ambiguous. For example, the poison symbol (☠) represents *all* poisons and in so doing fails to distinguish important information such as which type of poison is being referenced, its current physical state of matter, or how exactly it is harmful. Certain consumer product warning systems use combinations of symbols to circumvent this problem. For example, some countries use a household hazardous waste warning system that combines common warning symbols (e.g., the poison symbol; ☠) but nests them within upside-down triangles, diamonds, or octagons to denote increasing levels of danger. In this way, symbols can be combined with other symbols and/or shapes to enhance their specificity when required.

Nonetheless, symbols are most often used to convey information quickly and universally, and in so doing they are intentionally limited in their specificity and therefore in the number of semantic neighbours that they have. This absence of plausible lures could be what leads to conceptual distinctiveness, which in turn may drive the observed memory benefits for symbols. Even while conceptually broad, however, there is little overlap between symbols. Continuing with the example of the poison symbol (☠), the closest plausible lures would perhaps be the biohazard (☣) and radiation (☢) symbols, as all three refer to dangerous materials that should not be mishandled. However, two critical aspects are apparent: (1) The symbols are all highly visually distinct from each other, and (2) they still have distinct underlying meanings even if they are potentially confusable with each other. Therefore, if one knows the meaning of the poison symbol but not of the biohazard or radiation symbols, the universality of the former should imply that the latter two symbols are not thought to *also* represent poison. Even in cases of semantic neighbours, symbols are often conceptually distinct from each other.

Consider as well that most symbols have few related items while words often have many, further exacerbating differences in conceptual distinctiveness. Returning to an example from the beginning of this article, the word 'play' (in the context of watching television) has many related words, such as 'begin,' 'start,' 'commence,' etc., while the symbol for 'play' (▶) is semantically distinct from other symbols (i.e., no other symbol is used for the same purpose). Therefore, the symbol for play would be considered to possess high conceptual distinctiveness (Hamilton & Geraci, 2006).

The same symbols that lack semantic competitors could also be more distinct physically than their word counterparts. As noted previously, words in the English language are constructed from the same set of 26 letters, recycled repeatedly. So, while the physical form of words can vary depending on their underlying letters, length, or case, they are rather similar to each other in appearance. Symbols are instead an open-ended medium that can range from a single dot (i.e., the period symbol, '.') to complex shapes that could be difficult to draw by hand (e.g., the biohazard symbol, '☣'). Hence not only are symbols more distinct than words semantically but they also vary more perceptually, making them easier to identify and distinguish from each other.

It seems plausible that if physical distinctiveness was driving memory performance in Experiment 4, then our picture stimuli should have elicited the best performance. While the symbols and pictures used in that experiment were sized identically, and both were black and white, the symbols typically shared many simple geometric shapes and had solid fills whereas the images—with their multiple lines, shapes, shading, and sometimes distinct sub-parts—were arguably more physically distinct from one another. And yet, performance was practically identical between symbols and pictures in Experiment 4. This result is consistent with the idea from Ensor, Surprenant, and Neath (2019) that, because Experiment 4 used a free recall test, dual coding—rather than differences in physical distinctiveness—was likely driving the superior memory performance for symbols and pictures relative to words.

The results of Experiments 2 and 3, on the other hand, provide evidence more consistent with a distinctiveness explanation. In these two experiments, memory performance was measured using old/new recognition tests. Consequently, Ensor, Surprenant, and Neath (2019) would contend that the picture superiority effect should result from increased distinctiveness for images rather than from dual coding because, in these cases, access to the logogen pathway is not required to determine whether the item was studied. Therefore, the concreteness of the underlying concept should no longer affect memory in the case of recognition tests because there is no benefit of spontaneous imaging for concrete words. This is precisely what we found: In comparing symbol vs. word recognition performance in Experiments 2 and 3, the effect size in each case was identical (both $d = 0.44$), despite the word comparators being abstract and concrete, respectively. Thus, there was little evidence of dual coding at play in these particular experiments where recognition testing was used, suggesting that a different mechanism (physical distinctiveness, perhaps) was the main determinant of memory. Indeed, the intrinsically memorable visual properties of symbols highlighted in Experiment 5 further the idea that symbols may benefit memory as a result of physical distinctiveness.

Taken together, evidence from the present set of experiments is consistent with the idea that both dual coding and distinctiveness could be driving memory performance, depending upon the type of retrieval test used. This work was motivated by and is consistent with Paivio's dual coding theory, but other explanations are indeed viable as well. Although the current data do not allow us to disentangle theories of picture superiority, two novel theoretical contributions emerged: (1) The picture superiority effect in memory extends to symbols, even though most people would likely not think of symbols as images, and (2) that the concreteness of the underlying concept that symbols and pictures represent does not alter picture superiority in memory. This effect could be due to symbols offering unique visual referents for abstract words that are otherwise unlikely to be spontaneously imaged.

While our a priori reasoning was that symbols could be eliciting an image code for improved retention, for a verbal code to be provided one must also be familiar enough with the symbol to identify and label the imaged content. On top of that possibility, symbols also contain quite simple and recognizable visual features which could contribute to improved discriminability. As a result, both high-level familiarity and low-level visual attributes could play key roles in the memorability of symbols.

7.2. *Influence of familiarity and memorability*

Other studies certainly have investigated memory for abstract visuospatial stimuli (e.g., Fernandes & Guild, 2009) but, in those studies, the items used were novel shapes or patterns with no meanings. Consequently, familiarity could have played little role in those studies. Because previous studies of images have investigated only the extremes of familiar stimuli by using pictures of easily recognizable objects (e.g., Paivio & Csapo, 1973) or semantically void patterns (e.g., Smith et al., 1990), no prior investigations have explored the influence of familiarity with regard to memory for pictures.

This study provided the first investigation of familiarity's influence on memory for picture-like stimuli. Our prediction, based in dual coding theory, was that familiarity would correlate positively with memory for symbols. In the end, this prediction was not supported by the data: Familiarity was not consistently related to memory performance using within-subject comparisons in Experiments 1–3, and using aggregate data with familiarity ratings from naïve participants in Experiment 5 confirmed the same result. Nonetheless, the results reported here for familiarity are important to consider for theories of picture superiority because—insofar as symbols are pictures—it was shown that dual coding theory may be insufficient to explain the lack of a predicted relation between familiarity and memory for imaged content. The missing link between familiarity and memory for symbols suggests that knowing what a symbol *means* is not necessary to gain a memory benefit.

It may also be that symbols are simply efficient vehicles for conveying concepts, thanks to their visual properties. After all, symbols are often compact, use simple shapes and lines, and do not need color to be interpreted. To investigate this possibility, Experiment 5 made use of ResMem (Needell & Bainbridge, 2022), a newly built residual neural network that is capable of providing for any image a memorability score that corresponds to the likelihood of a person remembering it. The set of symbols used here received high memorability scores ($M = 0.91$ on a 0 to 1 scale). As expected, these memorability scores correlated significantly with memory performance for the current set of symbols. Hierarchical regression showed that, even when accounting for familiarity and frequency, memorability scores still significantly predicted memory performance. Therefore, it could be the case that symbols contain inherent visual properties that boost memory, apart from any influence of the observer's knowledge or personal experiences. Although research on memorability of symbols is still in its early days, there are many parallels to be drawn to existing concepts of visual imagery and distinctiveness accounts of memory more generally.

7.3. *Reconciling picture superiority with inherent memorability*

Dual coding and distinctiveness accounts of picture superiority share at least one critical parallel: the concept of an image trace from dual coding theory and the notion of physical distinctiveness for pictures. Both accounts posit that there is something inherently special about the visual nature of images. In dual coding theory, pictures are thought to elicit an 'image code.' In physical distinctiveness accounts, pictures are thought to be visually more distinct than words. But what about the low-level visual characteristics of images in relation to their memorability? Might image codes, physical distinctiveness, and intrinsic memorability represent different perspectives on the same mechanism?

Although the favorable visual aspects of symbols (and pictures) may be captured by the notion of an image trace, of physical distinctiveness, and of inherent memorability, each of these has yet another feature in common: meaning. In dual coding theory, this is the verbal trace, in distinctiveness accounts it is conceptual processing, and in memorability it is high-level semantic information such as category. Thus, in each case, encoding of low-level visual features is inherently tied to high-level semantic conceptualizations. One might assume that visual features drive higher-level abstractions of meaning—after all, one must perceive something before it can be identified. There has been work, however, showing that top-down meaning-based processes can affect bottom-up processing in a visual search task (Lupyan & Spivey, 2008).

Emerging conceptualizations of memorability have also suggested a critical role for semantic information in determining later memory (e.g., Koch, Akpan, & Coutanche, 2020). For instance, a recent study by Kramer, Hebart, Baker, and Bainbridge (2023) gathered over a million human ratings of memorability and found that whereas semantic and visual dimensions accounted for a combined 35% of the variance in the inherent memorability of images, the vast majority of that variance (31%) was due to semantic predictors alone. This finding also aligns well with the relative contributions of perceptual and semantic features thought to underlie memorability predictions from the ResMem neural network (Needell & Bainbridge, 2022). Thus, the memorability of a graphic symbol may depend little on its low-level visual features—such as whether it is comprised of curved or straight lines. It is likely, however, that there exists interplay between the distinct visual and semantic facets represented in each individual symbol.

8. Conclusion

This study sought to address how common graphic symbols are processed in human cognition. The major hypothesis was that graphic symbols serve to concretize abstract concepts. As a result, it was predicted that memory for symbols should be superior to memory for words. As well, if graphic symbols truly are akin to pictures, then memory for these two types of stimuli should be equivalent. Across four experiments, these predictions were confirmed: Symbols were indeed better remembered than words and did not differ from pictures. These findings remained stable in the face of changes to experiment setting, study design, retrieval test type, and even the nature of word-based comparators. A final experiment showed that the memorable properties that symbols possess predicted performance over and above ratings of familiarity and frequency. Thus, we conclude that graphic symbols likely are processed distinctly from words but that there seem to be moderating effects of the visual properties inherent in the designs of the symbols. Future work on symbols will elucidate factors underlying memory, cognitive faculties for abstract representation, and optimal techniques for efficient visual communication. While philosophers have debated the definition and use of symbols for centuries, modern psychology allows us to disentangle how symbols are represented in memory.

Declaration of Competing Interest

None.

Data availability

All data, analysis code, experiment programs, and other materials are listed on the Open Science Framework (OSF; https://osf.io/e53z4/).

Acknowledgements

This research was supported by a Natural Sciences and Engineering Research Council (NSERC) of Canada Postgraduate Scholarship to BRTR and by NSERC Discovery Grants A7459 to CMM and 2020-03917 to MAF. Early versions of this work were presented as talks at the University of Waterloo's annual Psychology Discovery Conference, the 6th annual meeting of the Toronto Area Memory Group (TAMeG), and the 32nd annual meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science (CSBBCS), in 2022. Portions of this work were also presented as conference posters during the 61st annual meeting of the Psychonomic Society in 2020 and the Images2Symbols pre-conference workshop during the 44th annual meeting of the Cognitive Science Society (CogSci) in 2022. We thank Saad Qadeer and Philip He for their help with scoring recall data in Experiment 4. Our pre-registrations, data, statistical code, and other materials are listed on the Open Science Framework (OSF): https://osf.io/e53z4/.

References

Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *NeuroReport, 16*(16), 1769–1773. https://doi.org/10.1097/01.wnr.0000183905.23396.f1

Atkin, A. (2013). Peirce's theory of signs. In *The Stanford encyclopedia of philosophy (Summer 2013 edition)*. https://plato.stanford.edu/archives/sum2013/entries/peirce-semiotics/.

Bainbridge, W. A. (2019). Memorability: How what we see influences what we remember. In *, Vol. 70. Psychology of learning and motivation* (pp. 1–27). Elsevier. https://doi.org/10.1016/bs.plm.2019.02.001.

Bainbridge, W. A. (2020). The resiliency of image memorability: A predictor of memory separate from attention and priming. *Neuropsychologia, 141*, Article 107408. https://doi.org/10.1016/j.neuropsychologia.2020.107408

Bainbridge, W. A. (2021). Memorability: Reconceptualizing memory as a visual attribute. *PsyArXiv.*. https://doi.org/10.31234/osf.io/ckv92

Bainbridge, W. A. (2022). Shared memories driven by the intrinsic memorability of items. In B. Ionescu, W. A. Bainbridge, & N. Murray (Eds.), *Human perception of visual information* (pp. 183–206). Springer International Publishing. https://doi.org/10.1007/978-3-030-81465-6_7.

Bainbridge, W. A., Dilks, D. D., & Oliva, A. (2017). Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *NeuroImage, 149*, 141–152. https://doi.org/10.1016/j.neuroimage.2017.01.063

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General, 142*(4), 1323–1334. https://doi.org/10.1037/a0033872

Bainbridge, W. A., & Rissman, J. (2018). Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. *Scientific Reports, 8*(1), 8679. https://doi.org/10.1038/s41598-018-26467-5

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology. General, 133*(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29*(4), 639–647. https://doi.org/10.3758/BF03200465

Bellhouse-King, M. W., & Standing, L. G. (2007). Recognition memory for concrete, regular abstract, and diverse abstract pictures. *Perceptual and Motor Skills, 104*(3), 758–762. https://doi.org/10.2466/pms.104.3.758-762

Bergstrom, J. A. (1893). Experiments upon physiological memory by means of the interference of associations. *The American Journal of Psychology, 5*(3), 356–369. https://doi.org/10.2307/1410999

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123–144). Lawrence Erlbaum.

Brady, T. F., & Bainbridge, W. A. (2022). *Visual memory* (1st ed.). Routledge.

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences, 18*(2), 90–98. https://doi.org/10.1016/j.tics.2013.11.005

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term-memory recall of repeated and novel word lists. *Memory & Cognition, 21*(4), 539–545. https://doi.org/10.3758/BF03197185

Coppens, L., Verkoeijen, P., & Rikers, R. (2011). Learning Adinkra symbols: The effect of testing. *Journal of Cognitive Psychology, 23*(3), 351–357. https://doi.org/10.1080/20445911.2011.507188

da Costa e Silva, T. (2019). *The logic of design process: Invention and discovery in light of the semiotics of Charles S. Peirce.* Transcript-Verlag.

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology, 30*(1), 92–101. https://doi.org/10.5334/irsp.82

Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2020). Taking parametric assumptions seriously: Arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA. *International Review of Social Psychology, 32*(1), 1–12. https://doi.org/10.5334/IRSP.198

Deng, J., Li, L.-J., Li, K., Li, F. F., Dong, W., Socher, R., Fei-Fei, L., & Li, K. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). https://doi.org/10.1109/cvpr.2009.5206848

Ding, J., Liu, W., & Yang, Y. (2017). The influence of concreteness of concepts on the integration of novel words into the semantic network. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.02111

Ensor, T. M., Bancroft, T. D., & Hockley, W. E. (2019). Listening to the picture-superiority effect: Evidence for the conceptual-distinctiveness account of picture superiority in recognition. *Experimental Psychology, 66*(2), 134–153. https://doi.org/10.1027/1618-3169/a000437

Ensor, T. M., Surprenant, A. M., & Neath, I. (2019). Increasing word distinctiveness eliminates the picture superiority effect in recognition: Evidence for the physical-distinctiveness account. *Memory & Cognition, 47*(1), 182–193. https://doi.org/10.3758/s13421-018-0858-9

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.

Fernandes, M., & Guild, E. (2009). Process-specific interference effects during recognition of spatial patterns and words. *Canadian Journal of Experimental Psychology, 63*(1), 24–32. https://doi.org/10.1037/a0012870

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

Goetschalckx, L., Moors, P., & Wagemans, J. (2018). Image memorability across longer time intervals. *Memory, 26*(5), 581–588. https://doi.org/10.1080/09658211.2017.1383435

Goodman, N. (1976). *Languages of art: An approach to a theory of symbols*. Hackett Publishing.

Hamilton, M., & Geraci, L. (2006). The picture superiority effect in conceptual implicit memory: A conceptual distinctiveness hypothesis. *The American Journal of Psychology, 119*(1), 1–20. https://doi.org/10.2307/20445315

Hamilton, M., & Rajaram, S. (2001). The concreteness effect in implicit and explicit memory tests. *Journal of Memory and Language, 44*(1), 96–117. https://doi.org/10.1006/jmla.2000.2749

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1217–1232. https://doi.org/10.1037/0278-7393.23.5.1217

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*(4), 421–445. https://doi.org/10.1006/jmla.1993.1023

Huttenlocher, J., & Higgins, E. T. (1978). Issues in the study of symbolic development. In W. A. Collins (Ed.), *Minnesota symposia on child psychology: Vol. 11* (pp. 98–138). Erlbaum. https://doi.org/10.4324/9780203780961-8/issues-study-symbolic-development-collins.

Isola, P., Jianxiong, X., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(7), 1469–1482. https://doi.org/10.1109/TPAMI.2013.200

Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 145–152.

Jessen, F., Heun, R., Erb, M., Granath, D. O., Klose, U., Papassotiropoulos, A., & Grodd, W. (2000). The concreteness effect: Evidence for dual coding and context availability. *Brain and Language, 74*(1), 103–112. https://doi.org/10.1006/brln.2000.2340

Juhasz, B. J., Lai, Y.-H., & Woodcock, M. L. (2015). A database of 629 English compound words: ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods, 47*(4), 1004–1019. https://doi.org/10.3758/s13428-014-0523-6

Kassambara, A. (2021). *rstatix: Pipe-friendly framework for basic statistical tests*.

Khanna, M. M., & Cortese, M. J. (2021). How well imageability, concreteness, perceptual strength, and action strength predict recognition memory, lexical decision, and reading aloud performance. *Memory, 29*(5), 622–636. https://doi.org/10.1080/09658211.2021.1924789

Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and predicting image memorability at a large scale. In *2015 IEEE international conference on computer vision (ICCV)* (pp. 2390–2398). https://doi.org/10.1109/ICCV.2015.275

Koch, G. E., Akpan, E., & Coutanche, M. N. (2020). Image memorability is predicted by discriminability and similarity in different stages of a convolutional neural network. *Learning & Memory, 27*(12), 503–509. https://doi.org/10.1101/lm.051649.120

Kramer, M. A., Hebart, M. N., Baker, C. I., & Bainbridge, W. A. (2023). The features underlying the memorability of objects. *Science advances, 9*(17), eadd2981. https://doi.org/10.1126/sciadv.add2981.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*. https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W., & Ladurner, G. (2004). The visual word form area and the frequency with which words are encountered: Evidence from a parametric fMRI study. *NeuroImage, 21*(3), 946–953. https://doi.org/10.1016/j.neuroimage.2003.10.021

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Lloyd, E. C., Shehzad, Z., Schebendach, J., Bakkour, A., Xue, A. M., Assaf, N. F., Jilani, R., Walsh, B. T., Steinglass, J., & Foerde, K. (2020). Food folio by Columbia center for eating disorders: A freely available food image database. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.585044

Lupyan, G., & Spivey, M. J. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology, 18*(10), R410–R412. https://doi.org/10.1016/j.cub.2008.02.073

Mintzer, M. Z., & Snodgrass, J. G. (1999). The picture superiority effect: Support for the distinctiveness model. *American Journal of Psychology, 112*(1), 113–146. https://doi.org/10.2307/1423627

Modrak, D. K. W. (2001). *Aristotle's theory of language and meaning*. Cambridge University Press.

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2011). *BayesFactor: Computation of Bayes factors for common design*.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Needell, C. D., & Bainbridge, W. A. (2022). Embracing new techniques in deep learning for estimating image memorability. *Computational Brain & Behavior, 5*(2), 168–184. https://doi.org/10.1007/s42113-022-00126-5

Nelson, D. L., Bajo, M. T., & Casanueva, D. (1985). Prior knowledge and memory: The influence of natural category size as a function of intention and distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(1), 94–105. https://doi.org/10.1037/0278-7393.11.1.94

Nelson, D. L., Cañas, J., Casanueva, D. M., & Castaño, D. (1985). Prior knowledge and recognition. *The American Journal of Psychology, 98*(3), 379–397. https://doi.org/10.2307/1422624

Nelson, D. L., & McEvoy, C. L. (1979). Encoding context and set size. *Journal of Experimental Psychology: Human Learning and Memory, 5*(3), 292–314. https://doi.org/10.1037/0278-7393.5.3.292

Nelson, D. L., Reed, V. S., & McEvoy, C. L. (1977). Learning to order pictures and words: A model of sensory and semantic encoding. *Journal of Experimental Psychology: Human Learning and Memory, 3*(5), 485–497. https://doi.org/10.1037/0278-7393.3.5.485

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76*(3), 241–263. https://doi.org/10.1037/h0027272

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

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255–287. https://doi.org/10.1037/h0084295

Paivio, A., & Csapo, K. (1969). Concrete image and verbal memory codes. *Journal of Experimental Psychology, 80*, 279–285. https://doi.org/10.1037/h0027273.

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., 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

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology Section A, 48*(2), 384–404. https://doi.org/10.1080/14640749508401396

Prada, M., Rodrigues, D., Silva, R. R., & Garrido, M. V. (2016). Lisbon symbol database (LSD): Subjective norms for 600 symbols. *Behavior Research Methods, 48*(4), 1370–1382. https://doi.org/10.3758/S13428-015-0643-7

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.r-project.org/.

Raven, J. C. (1958). *Guide to using the Mill Hill vocabulary scale with the progressive matrices scales* (p. 64). H. K. Lewis & Co.

Reinke, K., Fernandes, M. A., Schwindt, G., O'Craven, K., & Grady, C. L. (2008). Functional specificity of the visual word form area: General activation for words and symbols but specific network activation for words. *Brain and Language, 104*(2), 180–189. https://doi.org/10.1016/j.bandl.2007.04.006

Roberts, B. R. T., & Wammes, J. D. (2021). Drawing and memory: Using visual production to alleviate concreteness effects. *Psychonomic Bulletin & Review, 28*(1), 259–267. https://doi.org/10.3758/s13423-020-01804-w

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210.

Roodenrys, S., & Quinlan, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory, 8*(2), 71–78. https://doi.org/10.1080/096582100387623

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioral Ecology, 17*(4), 688–690. https://doi.org/10.1093/beheco/ark016

Sakurai, Y., Mimura, I., & Mannen, T. (2008). Agraphia for kanji resulting from a left posterior middle temporal gyrus lesion. *Behavioural Neurology, 19*(3), 93–106.

de Saussure, F., & Harris, R. (1998). *Course in general linguistics (reprint edition)*. Open Court (Original work published 1916).

Shakespeare, W. (2011). In B. A. Mowat, & P. Werstine (Eds.), *Romeo and Juliet*. Simon & Schuster. https://www.simonandschuster.com/books/Romeo-and-Juliet/William-Shakespeare/Folger-Shakespeare-Library/9781451621709 (Original work published 1597).

Smith, A. D., Park, D. C., Cherry, K., & Berkovsky, K. (1990). Age differences in memory for concrete and abstract pictures. *Journal of Gerontology, 45*(5), P205–P210. https://doi.org/10.1093/geronj/45.5.P205

Stenberg, G., Radeborg, K., & Hedman, L. R. (1995). The picture superiority effect in a cross-modality recognition task. *Memory & Cognition, 23*(4), 425–441. https://doi.org/10.3758/BF03197244

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., …

Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language, 51*(2), 247–250.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373. https://doi.org/10.1037/h0020071

Vinson, D., Jones, M., Sidhu, D. M., Lau-Zhu, A., Santiago, J., & Vigliocco, G. (2021). Iconicity emerges and is maintained in spoken language. *Journal of Experimental Psychology: General, 150*(11), 2293–2308. https://doi.org/10.1037/xge0001024

Vogt, S., & Magnussen, S. (2005). Hemispheric specialization and recognition memory for abstract and realistic pictures: A comparison of painters and laymen. *Brain and Cognition, 58*(3), 324–333. https://doi.org/10.1016/j.bandc.2005.03.003

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

Xie, W., Bainbridge, W. A., Inati, S. K., Baker, C. I., & Zaghloul, K. A. (2020). Memorability of words in arbitrary verbal associations modulates memory retrieval in the anterior temporal lobe. *Nature Human Behaviour, 4*(9), 937–948. https://doi.org/10.1038/s41562-020-0901-2