COGNITIVE SCIENCE
A Multidisciplinary Journal

Check for updates

Cognitive Science 47 (2023) e13382

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13382

Effects of Iconicity in Recognition Memory

David M. Sidhu,[a,b] Nareg Khachatoorian,[b] Gabriella Vigliocco[b]

[a] *Department of Psychology, Carleton University*
[b] *Department of Experimental Psychology, University College London*

Received 19 May 2023; received in revised form 1 October 2023; accepted 7 November 2023

Abstract

Iconicity refers to a resemblance between word form and meaning. Previous work has shown that iconic words are learned earlier and processed faster. Here, we examined whether iconic words are recognized better on a recognition memory task. We also manipulated the level at which items were encoded—with a focus on either their meaning or their form—in order to gain insight into the mechanism by which iconicity would affect memory. In comparison with non-iconic words, iconic words were associated with a higher false alarm rate, a lower d' score, and a lower response criterion in Experiment 1. We did not observe any interaction between iconicity and encoding condition. To test the generalizability of these findings, we examined effects of iconicity in a recognition memory megastudy across 3880 items. After controlling for a variety of lexical and semantic variables, iconicity was predictive of more hits and false alarms, and a lower response criterion in this dataset. In Experiment 2, we examined whether these effects were due to increased feelings of familiarity for iconic items by including a familiar versus recollect decision. This experiment replicated the overall results of Experiment 1 and found that participants were more likely to categorize words that they had seen before as familiar (vs. recollected) if they were iconic. Together, these results demonstrate that iconicity has an effect on memory. We discuss implications for theories of iconicity.

Keywords: Iconicity; Memory; Sound symbolism; Levels of processing; Multimodality

Correspondence should be sent to David M. Sidhu, Department of Psychology, University of Carleton, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada. E-mail: david.sidhu@carleton.ca

1. Introduction

1.1. Iconicity

Iconicity refers to the presence of imitative links between form and meaning in language (see Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Murgiano, Motamedi, & Vigliocco, 2021; Perniss, Thompson, & Vigliocco, 2010). Our chief concern here is phonological iconicity—instances in which the forms of words (i.e., their sound and/or articulation) resemble their meaning in some way. An example of this is onomatopoeia, in which sound directly imitates meaning (e.g., *splash, bang, mumble*). Iconicity also exists for non-sound meanings, for example, *teeny* (whose high pitch is evocative of smallness) or *cactus* (whose voiceless stops are evocative of sharp edges). Iconicity stands in opposition to the arbitrariness of the sign (de Saussure, 1916): the dictum that there is no special connection between the form of a word and its meaning. In English, iconicity has been observed in words for body parts (Johansson, Anikin, Carling, & Holmer, 2020; Joo, 2020), objects of different shapes (Sidhu, Westbury, Hollis, & Pexman, 2021), adjectives of size (Winter & Perlman, 2021), and texture (Winter, Sóskuthy, Perlman, & Dingemanse, 2022). Importantly, recent work has demonstrated the behavioral consequences of iconicity. For example, iconic words are responded to faster on naming and lexical decision tasks (is this a word or a nonword?; Meteyard, Stoppard, Snudden, Cappa, & Vigliocco, 2015; Sidhu, Vigliocco, & Pexman, 2020)

In the present study, we examine effects of iconicity on recognition memory. Our motivation for doing so was twofold. For one, there is a tradition of sensory experience (as measured by concreteness and imageability ratings) affecting recognition memory (see Khanna & Cortese, 2021). This invites the question of how a lexical-semantic property that *depicts* sensory experience (i.e., iconicity) would affect memory. In addition, as we will discuss below, the encoding phase of a recognition study allows for manipulations of how stimuli are processed, creating the opportunity for theoretically incisive comparisons.

1.2. Learning and memory

There is a rich literature exploring the effects of iconicity on learning. Developmentally, the general pattern is that early acquired words (and signs) tend to be higher in iconicity (Perry, Perlman, & Lupyan, 2015; Sidhu, Williamson, Slavova, & Pexman, 2022; for a review, see Nielsen & Dingemanse, 2021, Thompson et al., 2012). Iconicity is also used more often by adults when speaking to young infants (Motamedi et al., 2021; Vigliocco et al., 2019). The benefit of iconicity to infant learning has also been shown experimentally. For example, Imai, Kita, Nagumo, and Okada (2008) demonstrated that 3-year-old infants were better able to generalize the meanings of novel verbs if they were iconic. Imai and Kita (2014) speculate that an iconic relationship between a word and its meaning could help infants establish a link between the two.

Iconicity has also been shown to affect learning in adults during a single experimental session. Sonier, Poirier, Guitard, and Saint-Aubin (2020) had participants learn nonword labels for simple shapes (round, spiky, or neutral). Participants then had their memory tested by being presented with a shape along with three nonword options. Responses were more accurate when shapes were paired with iconically related labels (e.g., a round-associated nonword like bouba with a round shape). Similarly, Lockwood, Dingemanse, and Hagoort (2016) had a group of Dutch speakers learn Japanese ideophones (iconic words which depict sensory events; e.g., *gorogoro* meaning “a large object rolling repeatedly”). They found participants were more accurate when taught words along with their correct definitions as compared with their antonyms. There has also been related work showing that iconic nonwords can aide category learning (Lupyan & Casasanto, 2015; Monaghan, Mattock, & Walker, 2012).

In contrast to research on learning, we are unaware of any work on the effects of iconicity for memory of encountering previously learned items (i.e., words that a person already knows as they enter the testing situation). Studies of learning have manipulated the iconicity of to be learned items, and then tested participants shortly after. Instead, here we examine whether iconic versus non-iconic words—existing representations learned years earlier—lead to different memory traces. In particular, we focus on *recognition memory*, the process by which an individual recognizes an item as being one that they previously encountered. This depends on both recollection (i.e., recalling the specific episode of encountering the item earlier) and familiarity (i.e., feeling that you have encountered the item previously, but not remembering specific details).

There are notable differences between the processes involved in a learning versus recognition task. In a learning task, iconicity could play a role in learning and/or memory for word-meaning pairs. Conversely, a recognition task using already known words would help pinpoint effects of iconicity to memory. In addition, previous iconicity learning tasks have involved a forced choice task, in which participants see words and multiple meaning options at test. This introduces another process at test. Namely, participants could be biased toward choosing the iconically congruent meaning, regardless of memory. A recognition task presents only a single stimulus at test, also helping to pinpoint memory as a process.

Recent work has shown that lexical-semantic variables affecting lexical decision performance also play a role in recognition memory (e.g., Lau, Goh, & Yap, 2018). Most relevant for the current study, Khanna and Cortese (2021) examined the effects of imageability, concreteness, perceptual strength, and action strength on recognition memory for over 4700 items in existing megastudy datasets (Cortese, Khanna, & Hacker, 2010, 2015; also analyzed in the present study). In these megastudies, participants were presented with a random 50 words at study, and then were tested on those 50 words along with a random 50 new words (i.e., “Did you see this word at study?”). Khanna and Cortese’s analysis found that memory was better for words that were more imageable and concrete. The authors interpreted these results as evidence that “recognition memory benefits to the extent to which a word can evoke a mental image” (p. 628). With this in mind, it is plausible to expect a benefit from iconicity as well. Murgiano et al. (2021) suggested that iconicity serves to bring referents “to the mind’s eye” (p. 2). For example, the word *splash* imitates the sound of water, bringing those properties into the linguistic context in the absence of any actual water. This could presumably lead to a richer encoding experience, and/or improved retrieval.

Using existing words also allows us to take advantage of the levels of processing paradigm (see Craik, 2002). In this paradigm, participants are either encouraged to encode items in a deep manner (e.g., focusing on word meaning) or a shallow manner (e.g., focusing on word form). Deeper encoding is expected to result in a higher-quality memory trace, that then has a higher chance of being retrieved at test. Indeed, this is what is typically observed.

This levels of processing manipulation may help pinpoint the mechanism of iconicity effects since this is still unknown. One proposal is that iconic words enjoy extra links between the semantic system and modality-specific features (Meteyard et al., 2015). That is, the word *splash* may be associated with the auditory sensations that it imitates. Another, non-mutually exclusive, possibility is that the links between phonology and semantics are more direct or robust in iconic words (Meteyard et al., 2015; Sidhu et al., 2020). That is, because an iconic word's phonology has an imitative link with its meaning, these links may be stronger in some way, and lead to a benefit in processing. Indeed there is evidence that iconic words may lead to greater activation in areas of the brain responsible for processing the features being imitated (see Aryani, Hsu, & Jacobs, 2019; Kanero, Imai, Okuda, Okada, & Matsuda, 2014).

If it is true that effects of iconicity derive from links between phonology and semantics, then directing participants to focus on words' phonologies may be a "less shallow" task for iconic versus noniconic words. That is, focusing on the sound of *splash* may also entail a greater amount of semantic processing, because of the special link between phonology and semantics. This would result in a better memory trace for iconic versus non-iconic items encoded while focusing on words' forms. Thus, if any memory benefits observed for iconic words derive from special phonology-semantic links, then the difference between deep and shallow encoding conditions should be attenuated for iconic items.

1.3. The present studies

In the present studies, we examined whether performance is better for iconic words in a recognition memory task. In addition, we explored whether this benefit would arise from words' phonologies. To that end, in Experiment 1, we manipulated whether participants encoded words with a focus on their meanings or their forms. We then explored effects of iconicity in the data from a recognition memory megastudy (Cortese et al., 2010, 2015). Finally, we attempt to pinpoint the process leading to memory differences for iconic words in Experiment 2, by having participants indicate *how* they recognize the words they classify as old (i.e., with a recollect/familiar judgment, introduced later). Our prediction is that the special links between phonology and semantics should result in better memory for iconic versus noniconic items. This difference should be amplified when participants are directed to attend to words' forms.

2. Data availability

All experiment materials, data, and analysis code can be found at: https://osf.io/ce6wb/.

Table 1
Mean item properties (SD in parentheses)

Property	Non-iconic	Iconic
Iconicity	2.47 (0.41)	5.77 (0.48)
Length	5.74 (1.25)	5.85 (1.25)
Frequency	2.13 (0.51)	2.04 (0.52)
OLD	2.10 (0.45)	1.99 (0.45)
PLD	1.94 (0.57)	1.80 (0.51)
Concreteness	3.73 (0.99)	3.67 (0.60)
Valence	5.21 (0.92)	5.05 (1.03)
Bigram frequency	1586 (616.07)	1451 (624.24)
Number of phonemes	4.76 (1.15)	4.66 (1.10)
Number of syllables	1.74 (1.74)	1.63 (1.62)
Phonological neighbors	5.43 (7.04)	5.48 (7.04)
Articulation difficulty	2.00 (0.49)	1.91 (0.46)

3. Experiment 1

3.1. Methods

3.1.1. Participants

Based on an a priori power analysis, we aimed for a sample of 132 participants. This power analysis, along with the methods and analyses for this study, were preregistered and the preregistration can be found at: https://osf.io/ce6wb/registrations. Participants were recruited through the online platform Prolific (https://www.prolific.co/). All participants reported being fluent in English, and normal or corrected to normal vision. After eliminating participants who failed our attention checks or said that their data should not be used (e.g., because they did not understand instructions), we were left with a sample of 127 participants (85 male, $M_{\text{Age}} = 25.79$, $SD_{\text{Age}} = 7.56$).

3.1.2. Materials

Our stimuli consisted of 160 words: 80 iconic items and 80 non-iconic items. This was based on iconicity ratings in Winter, Perlman, Perry, Lupyan, and Dingemanse (2022). Words were rated on a scale ranging from 1 (non-iconic; i.e., arbitrary) to 7 (iconic). Our iconic items had ratings ≥ 4.7, while our noniconic items had ratings < 3.5.[1] Iconic and non-iconic items were also matched on log subtitle frequency (Brysbaert & New, 2009), length, orthographic Levenshtein distance, phonological Levenshtein distance (Yarkoni, Balota, & Yap, 2008), concreteness (Brysbaert, Warriner, & Kuperman, 2014), valence (Warriner, Kuperman, & Brysbaert, 2013), mean bigram frequency, number of phonemes, number of syllables, number of phonological neighbors, and ease of articulation (collected as pilot data; $n_{\text{per item}} \geq 8$). See Table 1. Stimuli were further separated into Lists A and B—each of these lists contained 40 iconic and 40 non-iconic items, and were matched on all of the variables mentioned. A full list of our stimuli can be found at: https://osf.io/zrnv5.

3.1.3. Procedure

Participants took part online, through the platform Gorilla (https://gorilla.sc/). They were first presented with an encoding task, in which they saw items from either List A or List B (list assignment was random), one at a time. There were two encoding conditions: a deep and a shallow condition, assignment to which was random. In the deep condition (henceforth semantic condition), participants were asked to rate the valence of each word from 1 (very unpleasant) to 5 (very pleasant). In the shallow condition (henceforth articulatory condition), participants were asked to rate how difficult each word was to articulate from 1 (very easy) to 5 (very difficult). We chose this shallow condition to direct participants to words' forms and expected this task to focus participants on words' articulation and phonology.

Following this, participants solved 10 addition problems as a distractor task. They then took part in a recognition task in which they saw all 80 studied words, along with 80 unstudied words (e.g., List B if they had studied List A), in a random order. For each word, their task was to indicate if they had seen the word before (old) or not (new).

3.2. Results

The data were analyzed using R software (R Core Team, 2021). We first conducted logistic mixed effects regressions on recognition task performance, separately for old and new trials. Models were run using the packages "lme4" (Bates, Mächler, Bolker, & Walker, 2015), "lmerTest" (Kuznetsova, Brockhoff, & Christensen, 2017), and "afex" (Singmann et al., 2016). The data and code for all analyses can be found here: https://osf.io/ce6wb/. The dependent variable was whether a word was correctly classified as either old or new. Our predictors of interest were each word's iconicity (iconic vs. non-iconic), encoding condition (semantic vs. articulatory), as well as an interaction between these variables. Iconicity and encoding condition were effects coded to allow interpretation of main effects. We also included: length, frequency, phonological Levenshtein distance (Yarkoni et al., 2008), bigram frequency, valence, ease of articulation,[2] and age of acquisition (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) as control variables. Valence and ease of articulation were added because our manipulation specifically directed participants to these properties in our items. Age of acquisition was added because previous work has shown it to correlate with iconicity (e.g., Perry et al., 2015). Continuous predictors were scaled. Models also included a random subject slope for iconicity, a random item slope for encoding condition, as well as random subject and item intercepts. To deal with convergence issues and singular fits, we attempted these remedies, in the following order: increasing the number of iterations, switching to the "bobyqa" optimizer, removing the correlation between the random subject intercept and slope, removing the random subject slope, and removing the random intercept with the lowest associated variance (see Brauer & Curtin, 2018; Meteyard & Davies, 2020).

3.2.1. Hits

The analysis of correct responses on old trials revealed a significant effect of encoding condition ($b = 0.26$, $p < .001$). Participants in the semantic condition were more likely to

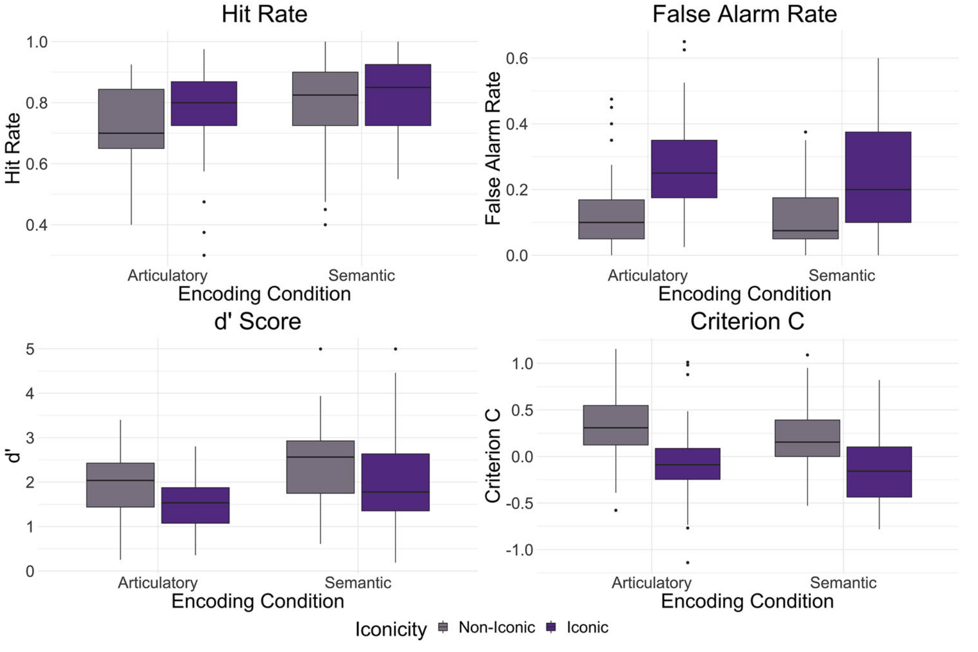

Fig. 1. Participant's average hit rate, false alarm rate, d' score, and criterion C, separated by iconicity and encoding condition, in Experiment 1.

correctly identify previously seen items as old. There was not a significant effect of iconicity ($b = 0.06$, $p = .25$), nor an interaction between iconicity and encoding condition ($b = -0.02$, $p = .50$). See Fig. 1. Participants were also more accurate when responding to items learned at a younger age ($b = -0.18$, $p < .001$). See Table 2.

3.2.2. False alarms

We next modeled the likelihood of incorrectly recognizing a new item as old. This model revealed a significant effect of iconicity ($b = 0.46$, $p < .001$). Participants were more likely to falsely recognize a new item as old if it was iconic. There was not a significant effect of encoding condition ($b = -0.14$, $p = .14$), nor a significant interaction between iconicity and encoding condition ($b = -0.04$, $p = .23$).[3] See Fig. 1. Participants were also more likely to make false alarms to longer ($b = 0.49$, $p < .001$) and less phonologically distinct items ($b = -0.40$, $p = .002$). See Table 3.

3.2.3. d' score

We next computed a d' score (1) for each participant, separately for iconic and noniconic items (see Macmillan & Creelman, 1991). Here and in the analysis of criterion C below, zHit Rate (zFalse Alarm Rate) represents participants' hit rate (false alarm rate) after being transformed into a z-score. This serves as an overall measure of a participant's ability to distinguish old and new items. We corrected hit rates and false alarm rates of 0 or 1 with the

Table 2
Results of mixed effects logistic regression predicting hits in Experiment 1

Fixed effect	*b*	*SE*	Wald's *Z*	*p*
Intercept	1.52	0.08	18.17	<.001***
Encoding condition	0.26	0.08	3.41	<.001***
Iconicity	0.06	0.06	1.14	.25
Length	0.04	0.07	0.61	.54
Log subtitle frequency	0.01	0.05	0.25	.80
PLD	0.02	0.07	0.31	.76
Emotional valence	0.04	0.05	0.91	.36
Mean bigram frequency	−0.08	0.05	−1.57	.12
Articulation difficulty	−0.01	0.05	−0.13	.90
Age of acquisition	−0.18	0.06	−3.34	<.001***
Encoding condition × Iconicity	−0.02	0.03	−0.67	.50
Random effect		s^2		*r*
Item intercept		0.20		
Item encoding slope		0.01		.01
Participant intercept		0.62		
Participant iconicity slope		0.02		.19

Note. PLD is phonological Levenshtein distance (Yarkoni et al., 2008).
***$p < .001$.

Table 3
Results of mixed effects logistic regression predicting false alarms in Experiment 1

Fixed effect	*b*	*SE*	Wald's *Z*	*p*
Intercept	−2.05	0.12	−16.72	<.001***
Encoding condition	−0.14	0.09	−1.47	.14
Iconicity	0.46	0.11	4.36	<.001***
Length	0.49	0.14	3.62	<.001***
Log subtitle frequency	0.03	0.10	0.26	.80
PLD	−0.40	0.13	−3.03	.002**
Emotional valence	0.08	0.09	0.90	.37
Mean bigram frequency	−0.06	0.09	−0.69	.49
Articulation difficulty	−0.10	0.10	−1.00	.32
Age of acquisition	0.00	0.10	−0.03	.98
Encoding condition × Iconicity	−0.04	0.03	−1.19	.23
Random effect		s^2		*r*
Item intercept		0.95		
Participant intercept		0.95		
Participant iconicity slope		0.03		.09

Note. PLD is phonological Levenshtein distance (Yarkoni et al., 2008).
***$p < .001$;
**$p < .01$.

correction suggested by Macmillan and Creelman (1991; i.e., adding or subtracting one-half trial, respectively).

$$d' = \text{zHit Rate} - \text{zFalse Alarm Rate} \tag{1}$$

We then ran an ANOVA with d' as the dependent variable, with iconicity, encoding condition, and their interaction, as predictors. Note that because ANOVAs must be done on subject means, we were not able to control for item-level properties. Nevertheless, this analysis revealed a significant effect of iconicity ($F[1, 125] = 64.80$, $p < .001$, $\eta^2{}_G = 0.06$) and encoding condition ($F[1, 125] = 10.39$, $p = .002$, $\eta^2{}_G = 0.07$). The interaction was not significant ($F[1, 125] = 1.28$, $p = .26$). The nature of these main effects was that participants had a higher d' score for non-iconic items ($M = 2.15$, $SD = 0.83$) than iconic items ($M = 1.76$, $SD = 0.85$), and that those in the semantic encoding condition had a higher d' score ($M = 2.17$, $SD = 0.92$) than those in the articulatory encoding condition ($M = 1.73$, $SD = 0.56$). See Fig. 1.

3.2.4. *Criterion C*

We calculated each participant's criterion C (2) for iconic and noniconic items (see Macmillan & Creelman, 1991). This is a measure of response bias, capturing an individual's willingness to say that they have previously seen an item. Lower values indicate a more liberal response threshold.

$$C = -\left[z\,(\text{Hit Rate}) + z\,(\text{False Alarm Rate})\right]/2 \tag{2}$$

We then ran an ANOVA with criterion C as the dependent variable, with the same predictors as above. This revealed a significant effect of iconicity ($F[1, 125] = 190.14$, $p < .001$, $\eta^2{}_G = 0.21$). The effect of encoding condition ($F[1, 125] = 2.63$, $p = .11$) and the interaction were not significant ($F[1, 125] = 2.91$, $p = .09$). The nature of the main effect of iconicity was that participants set a significantly lower response criterion for iconic items ($M = -0.10$, $SD = 0.37$) than for non-iconic items ($M = 0.27$, $SD = 0.35$). See Fig. 1.

3.3. *Discussion*

In contrast to our predictions, we observed a higher false alarm rate, and lower d' score and criterion C for iconic as opposed to non-iconic items. This seems to suggest that iconicity can actually make recognition memory decisions more difficult, because of a tendency to incorrectly identify iconic items as previously seen. While we did observe effects of encoding condition in hits and d' score, with better memory for items learned in the deep encoding condition, this did not interact with iconicity. This result goes against the hypothesis that iconic words have stronger connections between their forms and semantic representations.

An alternative interpretation of the results we observed (increased false alarms to iconic items), which does not include a different kind of processing for iconic items per se, is that it is an effect of list context. That is, half of the words that participants learned were iconic. This contrasts with their frequency in everyday language—iconic words tend to be less frequent (Sidhu & Pexman, 2018; Winter et al., 2023). Thus, we speculate that participants may have picked up on this, and been more likely to think that they had previously seen an iconic word when presented at test. We investigated this possibility in a recognition memory megastudy, which would not have provided the same list context to participants.

4. Analysis of memory megastudies

4.1. Method

We made use of two existing recognition memory megastudies (Cortese et al., 2010, 2015; including all experiments from each norms set). In these studies, participants memorized one of several lists of 50 words, and then were tested on a list of 100 words (half old and half new). Depending on the study, words were either all monosyllabic or disyllabic. We examined whether continuous iconicity ratings (Winter et al., 2023) would be predictive of memory performance, after controlling for: length, log subtitle frequency (Brysbaert & New, 2009), orthographic Levenshtein distance (Yarkoni et al., 2008), valence (Warriner et al., 2013), concreteness (Brysbaert et al., 2014), and age of acquisition (Kuperman et al., 2012). Analyses were performed on a list of 3880 words for which all values were available. All predictors were standardized.

4.2. Results

We found that iconicity was predictive of a higher hit rate ($b = 0.008$, $p < .001$), false alarm rate ($b = 0.008$, $p < .001$), and a lower criterion C ($b = -0.02$, $p < .001$). Iconicity was not a significant predictor of d’ scores ($b = -0.02$, $p = .70$). See Fig. 2.

4.3. Discussion

We replicated the effect of iconicity on false alarm rate and criterion C using data from a megastudy in which list context could not have caused the effect. We also observed an effect of iconicity on hit rate, which we did not find in Experiment 1. We now turn to an investigation of how this might be taking place. According to dual-process theories of memory, responses on a recognition task are determined by how familiar an item feels, and whether participants can recollect having seen the item earlier (see Yonelinas, 2002). Familiarity is understood to be driven by the ease with which an item’s features can be activated at test, while recollection requires an individual to identify the episodic memory of encountering the item. Viewing our results from this perspective, we might propose that iconic words resulted in a greater feeling of familiarity. Such an interpretation would be consistent with the view that iconic words more readily activate sensory features (Meteyard et al., 2015). In Experiment 2, we give participants the opportunity to respond whether the items they have seen before are familiar or are being recollected.

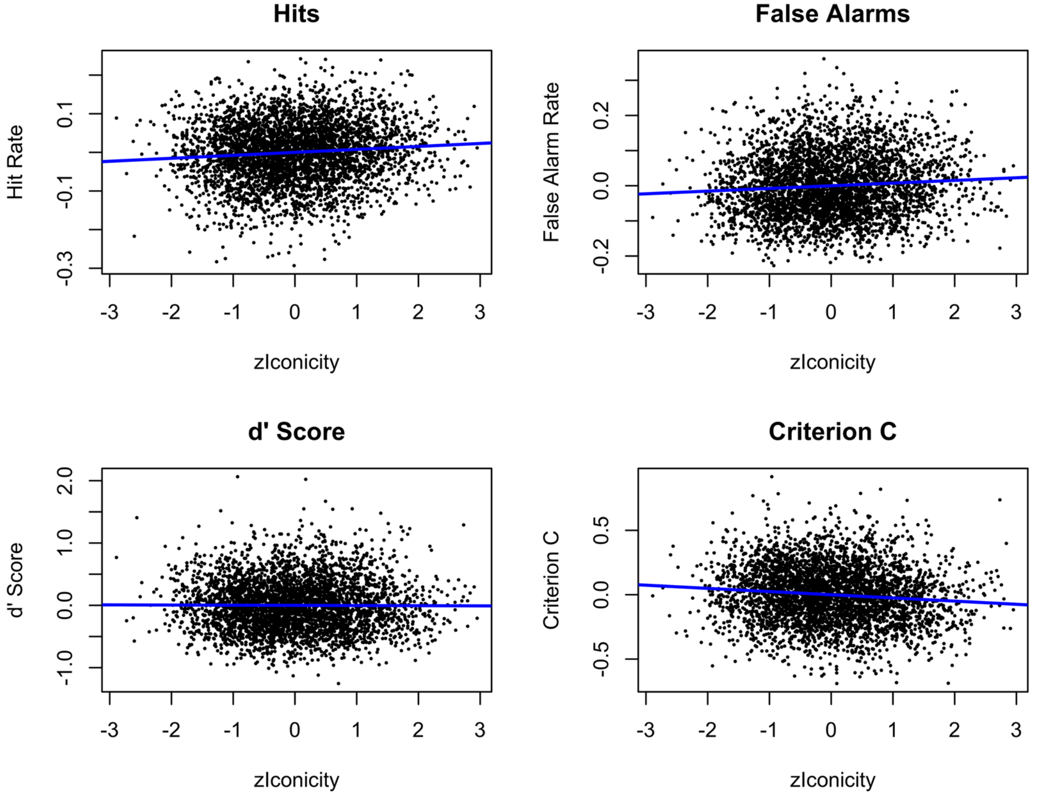

Fig. 2. Partial regression plots showing the relationship between standardized iconicity and hit rate, false alarm rate, d’ score, and criterion C for items in the recognition memory megastudy datasets (Cortese et al., 2010, 2015) after accounting for length, log subtitle frequency (Brysbaert & New, 2009), orthographic Levenshtein distance (Yarkoni et al., 2008), valence (Warriner et al., 2013), concreteness (Brysbaert et al., 2014), and age of acquisition (Kuperman et al., 2012). These plots show a given predictor on the *x*-axis (after partialing out variance associated with control variables), and the dependent variable on the *y*-axis (after again partialing out variance associated with control variables). Thus, they specifically show the effect of iconicity on each of the outcome measures. $N = 3880$.

5. Experiment 2

5.1. Methods

5.1.1. Participants

We aimed for a sample of 110 participants. This was based on the fact that the main analysis in this study would be a 2 (iconic vs. noniconic) × 2 (old vs. new item) ANOVA, with the percentage of old responses driven by familiarity as the dependent variable. In particular, we wished to be able to detect an effect of iconicity in new items (i.e., false alarms). Brysbaert (2019) suggests a sample size of 110 for this particular analysis. In addition, using the data from Experiment 1, we performed a power analysis to ensure that we would be able to detect an effect of iconicity on false alarms using a mixed effects model with this sample size. This suggested we would have 100% power to detect such an effect. A preregistration of these considerations, as well as the method and analyses for this experiment, can be found at: https://osf.io/ce6wb/. Participants were recruited through the online platform Prolific (https://www.prolific.co/). All participants reported being fluent in English, and normal or corrected to normal vision. We collected data until reaching a sample of 110 who passed our attention checks (72 male, $M_{\text{Age}} = 26.39$, $SD_{\text{Age}} = 6.89$; demographic info not collected for two participants).

5.1.2. Materials

The stimuli were identical to Experiment 1.

5.1.3. Procedure

The procedure was identical to Experiment 1, except that participants now had three response options on the recognition task: "recollect," "familiar," and "new." Participants were shown the following instructions, adapted from Brown and Bodner (2011):

> Your task will be to report whether you recognize each word from the rating task as outlined below. Please read the instructions carefully.
>
> Press "R" if you RECOLLECT the word.
> Recollection occurs when you are sure that you recognize the word and ONLY when you remember some specific detail about your experience with it on the rating task. Reading the word triggers a specific thought, image, or feeling that you experienced at the time you rated it. In other words, you remember why you recognize that word (e.g., ''I remember thinking what a funny-looking word it is'' or "I remember stretching my neck as I saw the word").
>
> Press "F" if the word is FAMILIAR.
> Familiarity occurs when you are sure you recognize the word but you can't remember any specific details about your experience with it on the rating task. In other words, you're sure you recognize it, but you're not sure why (e.g., ''I just know I saw that one'').
>
> Press "N" if the word is NEW.
> If you did not see the word in the rating task you should press "N."

5.2. Results

5.2.1. Hits

Here and in the subsequent analyses of false alarms, d' score, and criterion C, we treated "recollect" and "familiar" responses as "old" responses. These are both responses that indicate a participant thinks they have seen a word before. This allowed us to perform the same

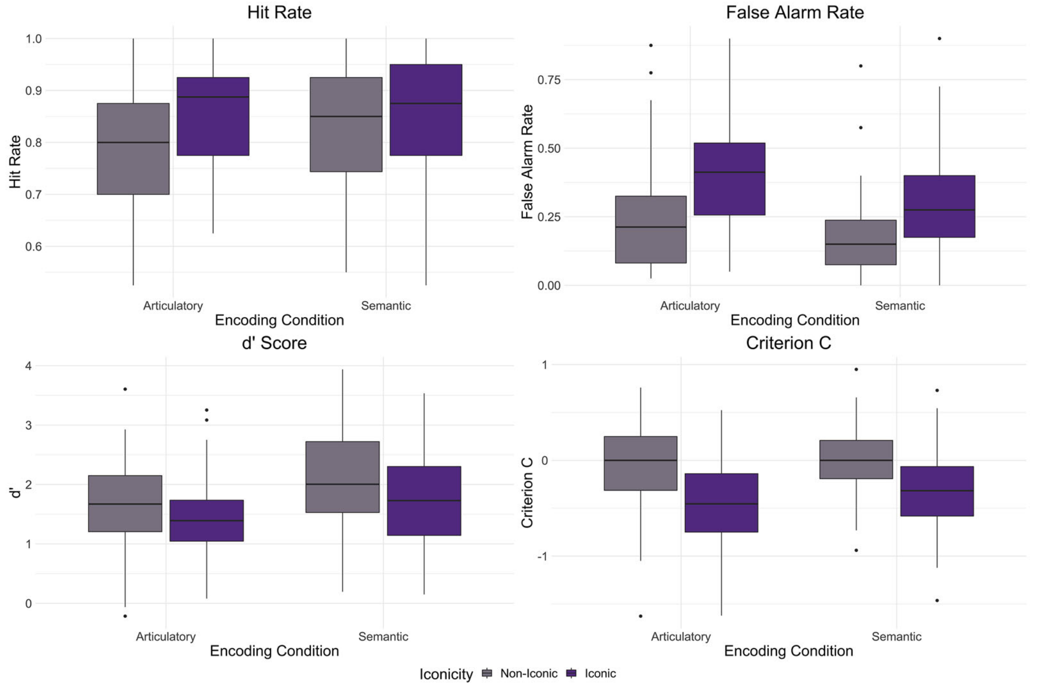

Fig. 3. Participant's average hit rate, false alarm rate, d' score, and criterion C, separated by iconicity and encoding condition, in Experiment 2.

analyses as in Experiment 1. The analysis of correct responses on old trials revealed a marginal effect of iconicity ($b = 0.22$, $p = .053$). The pattern was such that participants were more likely to correctly identify previously seen items as old if they were iconic. There was not a significant effect of encoding condition ($b = 0.13$, $p = .43$), nor a significant interaction between iconicity and encoding condition ($b = -0.25$, $p = .07$). See Fig. 3. Participants were also more accurate when responding to items learned at a younger age ($b = -0.24$, $p < .001$). See Table 4.

5.2.2. *False alarms*

The analysis of incorrect responses on new trials revealed a significant effect of encoding condition ($b = -0.55$, $p = .01$). Participants in the semantic condition were more likely to falsely recognize a new item as old. There was also a significant effect of iconicity ($b = 0.75$, $p < .001$). Participants were more likely to make false alarms to iconic items. The interaction between iconicity and encoding condition was not significant ($b = -0.05$, $p = .72$).[4] See Fig. 3. Participants were also more likely to make false alarms to longer ($b = 0.57$, $p < .001$) and less phonologically distinct items ($b = -0.43$, $p < .001$). See Table 5.

Table 4
Results of mixed effects logistic regression predicting hits in Experiment 2

Fixed effect	*b*	*SE*	Wald's *Z*	*p*
Intercept	1.83	0.09	20.94	< .001***
Encoding condition	0.13	0.16	0.79	.43
Iconicity	0.22	0.11	1.93	.054
Length	0.01	0.07	0.10	.92
Log subtitle frequency	0.01	0.05	0.24	.81
PLD	0.04	0.07	0.68	.50
Emotional valence	0.07	0.05	1.55	.12
Mean bigram frequency	−0.06	0.05	−1.26	.21
Articulation difficulty	0.06	0.05	1.12	.26
Age of acquisition	−0.24	0.05	−4.59	<.001***
Encoding condition × Iconicity	−0.25	0.14	−1.83	.07
Random effect		s^2		*r*
Item intercept		0.15		
Item encoding slope		0.06		−.70
Participant intercept		0.59		
Participant iconicity slope		0.08		.37

Note. PLD is phonological Levenshtein distance (Yarkoni et al., 2008).
***$p < .001$.

Table 5
Results of mixed effects logistic regression predicting false alarms in Experiment 2

Fixed effect	*b*	*SE*	Wald's *Z*	*p*
Intercept	−1.29	0.13	−10.07	<.001***
Encoding condition	−0.55	0.21	−2.56	.01*
Iconicity	0.74	0.19	4.00	<.001***
Length	0.57	0.12	4.73	<.001***
Log subtitle frequency	0.05	0.09	0.53	.60
PLD	−0.43	0.12	−3.64	<.001***
Emotional valence	−0.05	0.08	−0.64	.52
Mean bigram frequency	−0.08	0.08	−1.00	.32
Articulation difficulty	−0.06	0.09	−0.72	.47
Age of acquisition	−0.10	0.09	−1.09	.28
Encoding condition × Iconicity	−0.05	0.13	−0.36	.72
Random effect		s^2		*r*
Item intercept		0.75		
Participant intercept		1.16		
Participant iconicity slope		0.13		−.36

Note. PLD is phonological Levenshtein distance (Yarkoni et al., 2008).
***$p < .001$
*$p < .05$

5.2.3. d' score

This analysis revealed a significant effect of iconicity ($F[1, 108] = 33.90$, $p < .001$, $\eta^2_G = 0.04$) and encoding condition ($F[1, 108] = 7.43$, $p = .007$, $\eta^2_G = 0.06$). The interaction was not significant ($F[1, 108] = 1.23$, $p = .27$). The nature of these main effects was that participants had a higher d' score for non-iconic items ($M = 1.91$, $SD = 0.82$) than iconic items ($M = 1.61$, $SD = 0.74$), and that those in the semantic encoding condition had a higher d' score ($M = 1.94$, $SD = 0.84$) than those in the articulatory encoding condition ($M = 1.57$, $SD = 0.70$). See Fig. 3.

5.2.4. Criterion C

This revealed a significant effect of iconicity ($F[1, 108] = 132.08$, $p < .001$, $\eta^2_G = 0.15$). The effect of encoding condition ($F[1, 108] = 1.37$, $p = .25$) and the interaction were not significant ($F[1, 108] = 2.50$, $p = .12$). The nature of the main effect of iconicity was that participants set a significantly lower response criterion for iconic items ($M = -0.38$, $SD = 0.46$) than for non-iconic items ($M = -0.01$, $SD = 0.42$). See Fig. 3.

5.2.5. Recollect/know judgments

We conducted a logistic mixed effects regression on trials on which the participants indicated they had seen the item before. The dependent variable was whether participants indicated the item was familiar or recollected. We included all of the control variables present in previous mixed effects models. Our predictors of interest were a word's iconicity (iconic vs. non-iconic), word status (new vs. old), and their interaction.[5] These predictors were effects coded. There was a significant interaction between iconicity and word status ($b = 0.49$, $p = .002$). Follow-up analyses with estimated marginal means suggested that for old items, participants were significantly more likely to respond that an item was familiar if it was iconic ($EMM = 0.30$) than if it was non-iconic ($EMM = 0.24$; $p = .02$). This difference was not significant for new items ($p = .13$). See Fig. 4. In addition, words were more likely to be responded to as familiar if they: were new ($b = -2.12$, $p < .001$), less orthographically distinct ($b = -0.15$, $p = .01$), more negatively valenced ($b = -0.11$, $p = .009$), had more frequent bigrams ($b = 0.16$, $p < .001$), and were acquired at a later age ($b = 0.17$, $p < .001$). See Table 6.

6. Combined analysis of Experiments 1 and 2

In order to conduct a high-powered test of our hypotheses, we conducted an exploratory analysis that combined the data from Experiments 1 and 2. The analyses were the same as for Experiments 1 and 2 except that we added an effects-coded predictor for experiment.

6.1. Hits

The analysis of correct responses on old trials revealed a significant effect of encoding condition ($b = 0.36$, $p = .001$) and iconicity ($b = 0.32$, $p < .001$). Participants in the semantic

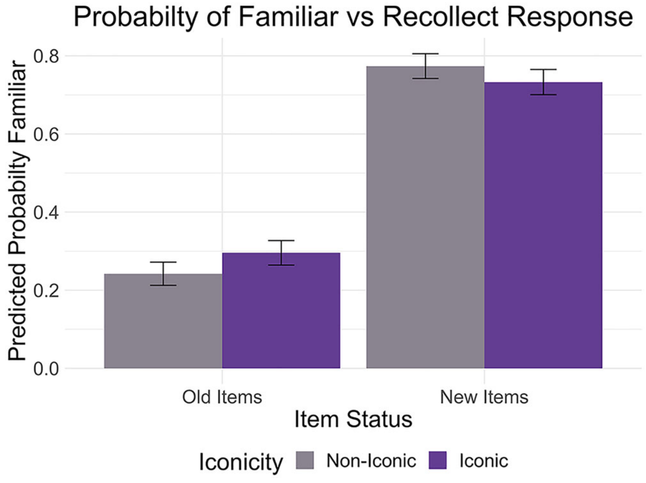

Fig. 4. Predicted probability of an item being categorized as "familiar" (vs. "recollected) at test in Experiment 2, for items that are not categorized as "new." Results are separated by iconicity and item status.

Table 6
Results of mixed effects logistic regression predicting "familiar" versus "recollect" responses, excluding "new" responses, in Experiment 2

Fixed effect	*b*	*SE*	Wald's *Z*	*p*
Intercept	0.06	0.14	0.41	.68
Item status	−2.12	0.12	−18.22	<.001***
Iconicity	0.03	0.11	0.25	.80
Length	0.07	0.06	1.07	.29
Log subtitle frequency	−0.04	0.05	−0.82	.41
PLD	−0.15	0.06	−2.47	.01*
Emotional valence	−0.11	0.04	−2.61	.009**
Mean bigram frequency	0.16	0.04	3.68	<.001***
Articulation difficulty	−0.07	0.05	−1.63	.10
Age of acquisition	0.17	0.05	3.57	<.001***
Item status × Iconicity	0.49	0.16	3.17	.002**

Random effect	s^2	*r*	
Item intercept	0.12		
Item status slope	0.30	.15	
Participant intercept	1.86		
Participant iconicity slope	0.17	−.18	
Participant status slope	0.56	.02	−.32

Note. PLD is phonological Levenshtein distance (Yarkoni et al., 2008).
***$p < .001$;
**$p < .01$.
*$p < .05$

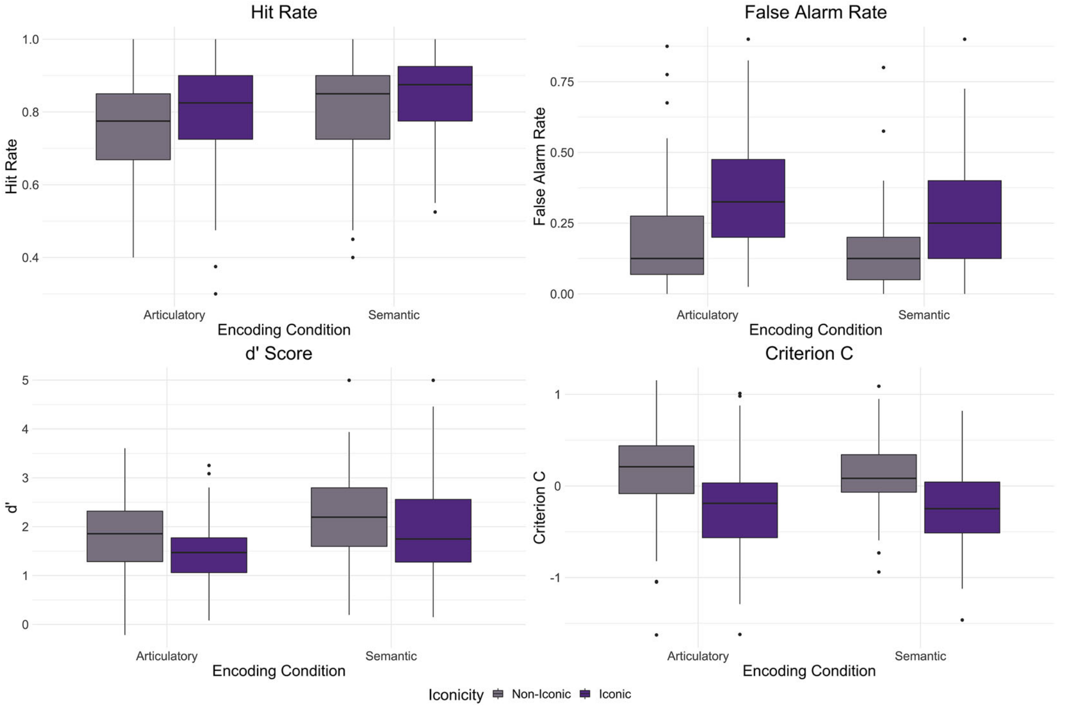

Fig. 5. Participant's average hit rate, false alarm rate, d' score, and criterion C, separated by iconicity and encoding condition, in Experiments 1 and 2 combined.

condition were more likely to correctly identify items as old. In addition, iconic items were more likely to be correctly identified as old. There was not a significant interaction between iconicity and encoding condition ($b = -0.14$, $p = .10$). See Fig. 5. See Table 7.

6.2. *False alarms*

The analysis of incorrect responses on new trials revealed a significant effect of encoding condition ($b = -0.42$, $p = .003$) and iconicity ($b = 1.01$, $p < .001$). Participants in the semantic condition were less likely to incorrectly identify an item as old (i.e., false alarm). In addition, iconic items were more likely to be incorrectly identified as old. There was not a significant interaction between iconicity and encoding condition ($b = -0.10$, $p = .32$). There were also more false alarms to longer words ($b = 0.45$, $p = .04$) and for participants in Experiment 2 ($b = 1.57$, $p = .01$). See Fig. 5. See Table 8.

6.3. *d' score*

This analysis revealed a significant effect of iconicity ($F[1, 235] = 96.58$, $p < .001$, $\eta^2_G = 0.05$) and encoding condition ($F[1, 235] = 17.69$, $p = .007$, $\eta^2_G = 0.06$). The interaction was not significant ($F[1, 235] = 0.01$, $p = .92$). The nature of these main effects was that

Table 7
Results of mixed effects logistic regression predicting hits in Experiments 1 and 2 combined

Fixed effect	b	SE	Wald's Z	p
Intercept	1.69	0.07	24.78	<.001***
Encoding condition	0.36	0.11	3.25	.001**
Iconicity	0.32	0.10	3.33	<.001***
Length	0.00	0.13	−0.02	.99
Log subtitle frequency	0.11	0.07	1.55	.12
PLD	0.07	0.07	0.94	.35
Emotional valence	0.22	0.12	1.85	.06
Mean bigram frequency	−0.04	0.05	−0.69	.49
Articulation difficulty	0.02	0.06	0.30	.76
Age of acquisition	−0.16	0.09	−1.77	.08
Experiment	0.64	0.39	1.63	.10
Encoding condition × Iconicity	−0.14	0.09	−1.66	.10

Random effect	s^2	r
Item intercept	0.21	
Participant intercept	0.61	
Participant iconicity slope	0.08	.27

Note. PLD is phonological Levenshtein distance (Yarkoni et al., 2008).
**$p < .01$;
***$p < .001$.

participants had a higher d' score for noniconic items (M = 2.04, SD = 0.83) than iconic items (M = 1.67, SD = 0.80), and that those in the semantic encoding condition had a higher d' score (M = 2.06, SD = 0.92) than those in the articulatory encoding condition (M = 1.65, SD = 0.69). See Fig. 5.

6.4. *Criterion C*

This analysis revealed a significant interaction between iconicity and encoding condition ($F[1, 235] = 5.44$, $p = .02$, $\eta^2_G = 0.003$). This interaction was such that iconicity had a numerically larger effect in the articulatory encoding condition (Estimated Marginal Mean Difference = 0.42, $p = < .001$) than the semantic encoding condition (Estimated Marginal Mean Difference = 0.32; $p = < .001$). There was also an overall main effect of iconicity ($F[1, 235] = 321.43$, $p < .001$, $\eta^2_G = 0.16$). The nature of this main effect was that participants had a higher criterion for noniconic items (M = 0.14, SD = 0.41) than iconic items (M = −0.23, SD = 0.44). See Fig. 5.

7. General discussion

In two experiments and an analysis of 3880 items from a recognition memory megastudy (Cortese et al., 2010, 2015), we found effects of iconicity on four measures of recognition

Table 8
Results of mixed effects logistic regression predicting false alarms in Experiments 1 and 2 combined

Fixed effect	*b*	*SE*	Wald's *Z*	*p*
Intercept	−1.62	0.11	−15.38	<.001***
Encoding condition	−0.42	0.14	−2.93	.003**
Iconicity	1.01	0.17	6.12	<.001***
Length	0.46	0.22	2.10	.04*
Log subtitle frequency	0.11	0.10	1.13	.26
PLD	−0.13	0.10	−1.20	.23
Emotional valence	0.00	0.22	−0.02	.99
Mean bigram frequency	−0.04	0.06	−0.77	.44
Articulation difficulty	0.07	0.09	0.84	.40
Age of acquisition	−0.01	0.13	−0.08	.94
Experiment	1.57	0.62	2.54	.01*
Encoding condition × Iconicity	−0.10	0.10	−1.00	.32
Random effect		s^2		*r*
Item intercept		0.88		
Item encoding slope		0.01		.31
Participant intercept		1.06		
Participant iconicity slope		0.13		−.12

Note. PLD is phonological Levenshtein distance (Yarkoni et al., 2008).
$^*p < .05$; $^{**}p < .01$; $^{***}p < .001$.

memory. Iconic words (vs. non-iconic words) were associated with a higher hit rate and false alarm rate. That is, participants were both more accurate in identifying previously seen iconic items, as well as more likely to incorrectly respond that they had seen a previously unseen item (i.e., a false alarm) if it was iconic. Further, we observed an overall lower d' Score for iconic items, indicating an overall worse ability to distinguish old and new items when they were iconic. Lastly, participants set a lower response criterion for iconic words: they required less evidence before indicating that they had seen a word before if it was iconic. Notably, apart from one exception discussed below, these effects did not interact with a manipulation of encoding condition (i.e., whether participants focused on the articulation or semantics of words).

Our observation of behavioral differences associated with iconic and non-iconic words adds to a growing appreciation that iconicity is an important property of language (see Murgiano et al., 2021), whose presence can affect a wide variety of cognitive processes. Previous work has shown that individuals have an easier time connecting a word to a meaning if the relationship between the two is iconic. This is true both for infants learning language (e.g., Imai et al., 2008) and adults learning words in an experimental setting (e.g., Lockwood et al., 2016). The present work adds to this by exploring iconicity in *memory*, for items that have been learned previously. That is, unlike studies that have taught participants novel word meanings and then tested them on those pairings, here we have tested memory for existing representations. Apart from any effects iconicity has on learning word meanings, we have shown that iconicity has an effect on recognition memory.

In the introduction, we outlined two general ways in which iconicity could affect processing: through special links between phonology and semantics, and through special links between semantics and modality-specific features. We included a manipulation of encoding task (i.e., a focus on articulation vs. semantics) as a way of directing some participants toward phonology-semantic links. Our manipulation of encoding condition was effective: we found main effects of encoding condition such that a focus on semantics improved memory. However, encoding condition did not interact with iconicity for any of our outcome measures, except for criterion C in the combined analysis (discussed below). This suggests that effects of iconicity, at least in this task, do not arise from links between phonology and semantics. Though, as a caveat, it is worth noting that our most robust finding (i.e., in false alarms) is an effect that occurs at test (as opposed to at encoding, when the manipulation took place). Nevertheless, this is consistent with the effect of iconicity in lexical decision not being enhanced by a greater emphasis on phonology (through degraded orthography and instructions to respond based on the *phonology* of a word; Sidhu et al., 2020).

As a sanity check that our manipulation of encoding condition *could* interact with item properties, we tested whether encoding condition interacted with item valence in the prediction of hits in Experiment 1. Because the semantic encoding task explicitly asked about valence, this is an interaction that we would expect to have an especially large effect size. Indeed, there was a significant interaction between the two ($b = 0.07$, $p = .01$), with valence having a larger effect in the semantic encoding condition. Thus, it was possible for our encoding manipulation to affect items differently based on their semantic properties, but this did not arise for iconicity.

If the effects of iconicity on memory do not depend on phonology-semantic links, where might the locus of such effects be? One possibility is that they reside in semantics alone. That is, the *meanings* of iconic words may be special in some way. Meteyard et al. (2015) proposed that iconic items could "enjoy additional connections from the semantic system to modality-specific features" (p. 273), a possibility that could have implications for embodied theories of language (see Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012). Dual-process theories propose that responses on a recognition memory task are determined by how familiar an item feels, and whether participants can recollect having seen the item earlier (see Yonelinas, 2002). Importantly, familiarity is understood to be driven by the ease with which an item's features can be activated at test.[6] Thus, we might have expected that the additional connections of iconic words from semantics to modality-specific features would allow these features to be more easily activated, and lead to more "familiar" responses. This is indeed what we observed with regard to previously seen items in Experiment 2. However, this does not explain the greater number of false alarms for iconic words, because new iconic words did not lead to more "familiar" responses.

Future research may explore this question using a task beyond a familiar/recollect judgment. Indeed, this judgment relies on participants' introspection and forces them to make a binary decision. Yonelinas (2002) suggests that this may "underestimate the probability that an item is familiar" (p. 453). Instead, an avenue for future research would be to have participants make a confidence judgment of their memory along a continuous scale. This could then be used in a receiver operator characteristic (ROC; Yonelinas & Parks, 2007) analysis to better understand the effect of response bias on memory accuracy for iconic versus noniconic items.

We did observe one instance of iconicity interacting with encoding condition, namely, in the prediction of criterion C (i.e., an individual's willingness to say that they have previously seen an item) in the combined analysis. While participants in general set a lower response criterion for iconic items, this tendency was increased for those in the articulation condition. This is a somewhat puzzling effect, as criterion C is a measure of responses during the test phase. It would suggest that a focus on articulation during the study somehow led to participants setting a lower criterion during recognition for iconic items. One might posit that focusing on articulation during encoding made participants more attentive to iconicity in general, and that this carried forward to the recognition task. However, without a straightforward interpretation, we should be wary of overinterpreting this result until it is replicated.

We also speculated that the effects we observed could be due to list context. That is, the high number of iconic items study list could have led to a general tendency to identify iconic items as previously seen. A related possibility is that the iconic items inadvertently created a Deese−Roediger−McDermott effect (Deese, 1959; Roediger & McDermott, 1995): a list-based false memory effect in which a word list related to a target (e.g., *glass*, *pane*, *curtain*; for the target *window*) creates a false memory of having seen the target. Perhaps iconic words created a false memory of having seen other iconic words. However, such effects could not have been the case in the megastudy data, suggesting that iconicity can affect memory even beyond list context effects. Though, note that there was not an effect of iconicity on d' score in the megastudy data, perhaps suggesting that this effect is especially dependent on list context.

A related possibility is that the effects arise from a greater semantic overlap among the iconic items as compared to the noniconic items. Indeed, a supplementary analysis found that there was a smaller average cosine distance (a proxy for distance in meaning; Mandera, Keuleers, & Brysbaert, 2017) among iconic items ($M = 0.86$) as compared to non-iconic items ($M = 0.94$; $b = 0.07$, $p < .001$). One account of recognition memory is that the encoding of a given item is based on a subset of its features being stored in memory (McClelland & Chappell, 1998). Then, at test, an item is responded to as previously seen if a certain number of its features match those previously stored. Notably, this account predicts a large effect of item similarity on recognition memory, and when new items at test have many features in common with those previously studied, this should lead to increased false alarms. This would be one way of explaining the increased false alarms for iconic items.

It is possible that a version of this explanation could also apply to the greater number of false alarms in the megastudy. We found that there was also a smaller cosine similarity among the most iconic items in the megastudy dataset ($M = 0.87$) as compared to the most arbitrary items ($M = 0.91$; defined as items with a standardized iconicity value above 1 or below -1, respectively; $b = -0.05$, $p < .001$). Thus, perhaps even when presented with a random 100 items at test (half old and half new; i.e., in Cortese et al., 2010, 2015), the new iconic items that appear will be confused with the old iconic items. There is some evidence that iconicity improves category learning but not the learning of individual labels (e.g., Monaghan et al., 2012). From this, we might conclude that iconic items are processed in a less individuated manner, without arriving at a specific representation of their unique meanings. This could also lead to greater confusion. Of course, this is admittedly speculative at this point. An avenue for future work on iconicity and memory could be to manipulate the number of semantic categories to which iconic and non-iconic items belong.

In addition, we intended ratings of ease of articulation to focus participants on words' phonology. However, it is possible that this was not successful, and that participants instead focused on words' articulation. Because phonology and articulation are so closely linked, it is often unclear which aspect of phonemes are more responsible for a word's iconicity. For example, the vowel in *teeny* has a high pitch as well as a small space in the mouth as it is articulated. If it were the case that the items used here relied more on phonological rather than articulatory iconicity, our manipulation of encoding condition may not have been ideal. This could be one explanation for the lack of an interaction between encoding condition and iconicity.

The study of iconicity in memory represents a fertile opportunity for future work. It allows researchers to distinguish between effects of iconicity on learnability and retrieval. In addition, previous work has suggested that iconic vocabularies may be problematic due to greater confusability among items (e.g., Gasser, 2004). A memory paradigm could provide insight into whether this difficulty arises in the storage and organization of the lexicon, or the retrieval of information from the lexicon. Taking a broader perspective, this work demonstrates another item-level variable that can impact recognition memory (see also Khanna & Cortese, 2021).

The main contribution of the present work was to show that iconicity has an effect on recognition memory, both in studies with carefully manipulated items sets, as well as megastudy data involving nearly 4000 items. There is clearly much work yet to be done understanding the mechanisms by which iconicity could have an effect on recognition memory. Nevertheless, this work adds to our growing appreciation for the fact that iconicity is far from a linguistic curiosity, and rather a factor that has a wide-ranging psychological impact.

Acknowledgments

David Sidhu was supported by a Social Sciences and Humanities Research Council of Canada (SSHRC) postdoctoral fellowship (756-2019-0557). Gabriella Vigliocco was supported by a European Research Council Advanced Grant (ECOLANG, 743035), and a Royal Society Wolfson Research Merit Award (WRM/R3/170016). The authors would like to thank Drs. Glen Bodner and Alex Taikh for their advice regarding memory task instructions.

Open Research Badges

This article has earned Open Data Research Design badge. Data is available at https://osf.io/ce6wb/.

Notes

1 Item selection was performed using an in-progress version of these ratings. However, in the paper, we report final values. This explains why cutoffs were not equally spaced—at the time of item selection, using in progress ratings, we used the cutoffs of > 5.5 and < 2.5. Also note that 21 of our 160 items did not appear in the final set of items. However, the raw data were available for these items, which we used to compute their iconicity values.

2 As a sanity check, we correlated each item's ease of articulation from pilot ratings with the average rating it received in this experiment. There was a significant positive correlation between the two ($r = .55$, $p < .001$).

3 Note that we ran a version of this analysis, and the above analysis of hits, treating iconicity as a continuous predictor. This did not change the pattern of results. We report categorical analyses of hits and false alarms in text to parallel the analyses of d' and criterion C.

4 Note that we ran a version of this analysis, and the above analysis of hits, treating iconicity as a continuous predictor. This did not change the pattern of results except that the effect of iconicity was a significant predictor of hits ($p = .04$).

5 We ran a version of this analysis that included encoding condition, along with all interactions. The threeway interaction was not significant ($p = .62$), nor was the interaction between encoding condition and iconicity ($p = .54$). There was a significant interaction between encoding condition and item status ($b = -0.29$, $p = .049$). Follow-up analyses with estimated marginal means suggested that familiar responses ($EMM = 1.48$) were more likely than recollect responses ($EMM = 0.80$) for new items (Estimate $= -0.69$, $p = .02$), while there was no effect for old items ($p = .28$).

6 In a Supplementary Analysis, we examined whether we would observe more "familiar" responses for words with more associated perceptual experience. This was operationalized using the overall perceptual strength of words in the Lancaster sensorimotor norms (Lynott, Connell, Brysbaert, Brand, & Carney, 2020), in which participants were asked to rate the extent to which words were experienced in different modalities. In particular, we used the Minkowski 3 measure that combined ratings in different senses. We found that words with more perceptual experience actually lead to more "recollect" responses ($p < .001$).

References

Aryani, A., Hsu, C.-T., & Jacobs, A. M. (2019). Affective iconic words benefit from additional sound–meaning integration in the left amygdala. *Human Brain Mapping*, *40*(18), 5289–5300. https://doi.org/10.1002/hbm.24772

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. https://doi.org/10.1037/met0000159

Brown, A. A., & Bodner, G. E. (2011). Re-examining dissociations between remembering and knowing: Binary judgments vs. independent ratings. *Journal of Memory and Language*, *65*(2), 98–108. https://doi.org/10.1016/j.jml.2011.04.003

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1), 16. https://doi.org/10.5334/joc.72

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

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

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, *18*(6), 595–609. https://doi.org/10.1080/09658211.2010.493892

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology (2006)*, *68*(8), 1489–1501. https://doi.org/10.1080/17470218.2014.945096

Craik, F. I. M. (2002). Levels of processing: Past, present… and future? *Memory*, *10*(5–6), 305–318. https://doi.org/10.1080/09658210244000135

de Saussure, F. (1916). *Course in general linguistics*. Columbia University Press.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615. https://doi.org/10.1016/j.tics.2015.07.013

Gasser, M. (2004). The origins of arbitrariness in language. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130298. https://doi.org/10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Johansson, N. E., Anikin, A., Carling, G., & Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology*, *24*(2), 253–310. https://doi.org/10.1515/lingty-2020-2034

Joo, I. (2020). Phonosemantic biases found in Leipzig-Jakarta lists of 66 languages. *Linguistic Typology*, *24*(1), 1–12. https://doi.org/10.1515/lingty-2019-0030

Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How sound symbolism is processed in the brain: A study on Japanese mimetic words. *PLoS ONE*, *9*(5), e97905. https://doi.org/10.1371/journal.pone.0097905

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

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lau, M. C., Goh, W. D., & Yap, M. J. (2018). An item-level analysis of lexical-semantic effects in free recall and recognition memory using the megastudy approach. *Quarterly Journal of Experimental Psychology*, *71*(10), 2207–2222. https://doi.org/10.1177/1747021817739834

Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1274–1281. https://doi.org/10.1037/xlm0000235

Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition*, *7*(2), 167–193. https://doi.org/10.1017/langcog.2014.21

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*, 724–760. https://doi.org/10.1037/0033-295X.105.4.734-760

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

Meteyard, L., Stoppard, E., Snudden, D., Cappa, S. F., & Vigliocco, G. (2015). When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia*, *76*, 264–275. https://doi.org/10.1016/j.neuropsychologia.2015.01.042

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1152–1164. https://doi.org/10.1037/a0027747

Motamedi, Y., Murgiano, M., Perniss, P., Wonnacott, E., Marshall, C., Goldin-Meadow, S., & Vigliocco, G. (2021). Linking language to sensory experience: Onomatopoeia in early language development. *Developmental Science*, *24*, e13066.

Murgiano, M., Motamedi, Y., & Vigliocco, G. (2021). Situating language in the real-world: The role of multimodal iconicity and indexicality. *Journal of Cognition*, *4*(1), 38. https://doi.org/10.5334/joc.113

Nielsen, A. K., & Dingemanse, M. (2021). Iconicity in word learning and beyond: A critical review. *Language and Speech*, *64*(1), 52–72. https://doi.org/10.1177/0023830920914339

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 1–15. https://doi.org/10.3389/fpsyg.2010.00227

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE*, *10*(9), e0137147. https://doi.org/10.1371/journal.pone.0137147

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Sidhu, D. M., & Pexman, P. M. (2018). Lonely sensational icons: Semantic neighbourhood density, sensory experience and iconicity. *Language, Cognition and Neuroscience*, *33*(1), 25–31. https://doi.org/10.1080/23273798.2017.1358379

Sidhu, D. M., Vigliocco, G., & Pexman, P. M. (2020). Effects of iconicity in lexical decision. *Language and Cognition*, *12*(1), 164–181. https://doi.org/10.1017/langcog.2019.36

Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review*, *28*(4), 1390–1398. https://doi.org/10.3758/s13423-021-01883-3

Sidhu, D. M., Williamson, J., Slavova, V., & Pexman, P. M. (2022). An investigation of iconic language development in four datasets. *Journal of Child Language*, *49*(2), 382–396. https://doi.org/10.1017/S0305000921000040

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *afex: Analysis of Factorial Experiments* (R package version 0.16-1) [Computer software].

Sonier, R.-P., Poirier, M., Guitard, D., & Saint-Aubin, J. (2020). A round Bouba is easier to remember than a curved Kiki: Sound-symbolism can support associative memory. *Psychonomic Bulletin & Review*, *27*(4), 776–782. https://doi.org/10.3758/s13423-020-01733-8

Thompson, R. L., Vinson, D. P., Woll, B., & Vigliocco, G. (2012). The road to language learning is iconic: Evidence from British Sign Language. *Psychological Science*, *23*, 1443–1448.

Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C., Milán-Maillo, I., & Perniss, P. (2019). *Onomatopoeia, gestures, actions and words: How do caregivers use multimodal cues in their communication to children?* PsyArXiv. https://doi.org/10.31234/osf.io/v263k

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Winter, B., & Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa: A Journal of General Linguistics*, *6*(1), 1–13. https://doi.org/10.5334/gjgl.1646

Winter, B., Lupyan, G., Perry, L. K., Dingemanse, M., & Perlman, M. (2023). Iconicity ratings for 14,000+ English words. *Behavior Research Methods*, 1–16.

Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled /r/is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-021-04311-7

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979. https://doi.org/10.3758/PBR.15.5.971

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, *133*(5), 800–832. https://doi.org/10.1037/0033-2909.133.5.800

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information
Supporting Information
Supporting Information
Supporting Information
Supporting Information