Effects of Iconicity in Recognition Memory

David M. Sidhu (dmsidhu@gmail.com)
Carleton University, Ottawa, Canada
University College London, London, UK

Gabriella Vigliocco (g.vigliocco@ucl.ac.uk)
University College London, London, UK

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 iconicity would also affect 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 criterion C. We did not observe any interaction between iconicity and encoding condition. We also conducted an analysis of recognition memory megastudy data and found that iconicity was predictive of higher false alarms and a lower criterion C across 1,646 items. We interpret these results as suggesting that iconicity leads to a feeling of familiarity in recognition memory.

Keywords: iconicity; recognition memory; sound symbolism; levels of processing

Introduction

Iconicity

Iconicity refers to the presence of imagistic links between form and meaning in language (see Dingemanse et al., 2015; Murgiano et al., 2021; Perniss et al., 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*). However, iconicity can also manifest in analogical crossmodal links such as in *teeny* (whose high pitch is evocative of smallness) or *gooey* (which is evocative of stickiness). Iconicity stands in opposition to the *arbitrariness of the sign* (Saussure, 1916): the dictum that there is no special connection between the form of a word and its meaning.

Iconicity has now been shown to exist beyond onomatopoeia (which has long been accepted as an instance of iconicity; see Saussure, 1916). It has been observed in words for body parts (Johansson et al., 2020; Joo, 2020), objects of different shapes (Sidhu et al., 2021), and adjectives of size (Winter & Perlman, 2021) and texture (Winter et al., 2021). These demonstrations have included analyses within a single language, and across languages.

In addition to demonstrating that iconicity exists, the field has also begun to show that it has observable consequences for language users. Iconicity has been shown to affect language acquisition, with iconic words being acquired earlier (Perry et al., 2015; Sidhu et al., 2021), as well as used more often by adults when speaking to young infants (e.g., Vigliocco et al., 2020). It appears that the imitative link between form and meaning may help children acquire iconic words.

Neuroimaging work has shown that iconic words recruit different brain areas as compared to non-iconic words (Hashimoto et al., 2010; Kanero et al., 2014; Revill et al., 2014). In addition, EEG work has shown that processing iconic vs. non-iconic words leads to observable differences in brain signatures (e.g., Lockwood & Tuomainen, 2015; Peeters, 2016; Vigliocco et al., 2019). This has been interpreted by some as an integration of iconic words' phonologies with the sensory properties that they resemble (e.g., Lockwood & Tuomainen, 2015).

Recent work has also shown behavioural differences in responses to iconic vs. non-iconic words. Meteyard et al. (2015) found that iconic words were responded to faster on a naming as well as an auditory lexical decision task (i.e., is this a word or a nonword?). Sidhu et al. (2020) also found that participants were faster and more accurate when responding to iconic words on a visual lexical decision task as well as a phonological lexical decision task (i.e., does the *sound* of this letter string correspond to a real word?).

In the present study we follow up on this in the domain of 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 property that *depicts* sensory experience 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.

Memory

There has not been a great deal of work examining effects of iconicity in memory. Sonier et al. (2020) found that participants had better memory for nonword-shape pairs if the pairs were iconically related (e.g., a round-associated nonword like *bouba* with a round shape). Similarly, Lockwood et al. (2016) found that participants were better able to learn pairings between foreign words and meanings if they were iconic.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

In the present study we examined *recognition memory* for known words. This allowed us to examine the contribution of an individual word's iconicity to memory. We were also able to avoid complications arising from participants having to learn a new pairing.

There has been recent work showing that various lexical-semantic variables that affect lexical decision performance also play a role in recognition memory (e.g., Lau et al., 2018). Most relevant for the current study, Khanna and Cortese (2021) showed that imageability and concreteness both improved recognition memory performance. 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.

In addition, we made use of the levels of processing paradigm (see Craik, 2002), by manipulating the level at which items were encoded. In this paradigm, participants are either encouraged to encode items in a deep manner (i.e., focusing on word meaning) or a shallow manner (i.e., 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 manipulation may help pinpoint the mechanism of iconicity effects as 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 special in some way, and lead to a benefit in processing. However, experimental manipulations by Sidhu et al. (2020) aimed at directing participants towards phonology (vs. orthography) only marginally increased the iconicity effect.

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 vs. non-iconic 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 vs. 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.

The Present Study

In the present study we examined whether the special nature of iconic words would lead to better recognition memory. In addition, we explored whether this benefit would arise from words' phonologies. To that end we manipulated whether participants encoded words with a focus on their meanings or their forms.

Methods

Participants

Based on an a priori power analysis, we aimed for a sample of 132 participants. This power analysis, along with the methods for this study, were preregistered and can be found at: https://osf.io/ujeyz. 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 we were left with a sample of 127 participants (85 male, M_{Age} = 25.79, SD_{Age} = 7.56).

Materials

Our stimuli consisted of 160 words: 80 iconic items and 80 non-iconic items. This was based on an unpublished set of iconicity ratings retrieved from here: (https://github.com/bodowinter/iconicity_ratings). Words were rated on a scale ranging from one (non-iconic; i.e., arbitrary) to seven (iconic). We selected our iconic items from words with a rating > 5.5, and our non-iconic items with from words with a rating < 2.5. Iconic and non-iconic items were also matched on log subtitle frequency (Brysbaert & New, 2009), length, orthographic Levenshtein distance, phonological Levenshtein distance (Yarkoni et al., 2008), concreteness (Brysbaert et al., 2014), valence (Warriner et al., 2013), mean bigram frequency, number of phonemes, number of syllables, number of phonological neighbours and ease of articulation (collected as pilot data). 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.

Table 1: Mean item properties (SD in parentheses)

Property	Non-Iconic	Iconic
Iconicity	2.15 (0.29)	5.98 (0.35)
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 Neighbours	5.43 (7.04)	5.48 (7.04)
Articulation Difficulty	2.00 (0.49)	1.91 (0.46)

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 one (very unpleasant) to five (very pleasant). In the shallow condition (henceforth *articulatory condition*) participants were asked to rate how difficult each word was to articulate from one (very easy) to five (very difficult). We chose this manipulation to direct participants to words' forms and expected this task to focus participants on words' articulation and phonology.

Following this, participants solved ten 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*).

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 et al., 2015), "lmerTest" (Kuznetsova et al., 2017), and "afex" (Singmann et al., 2021). 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, valence, bigram frequency, ease of articulation and age of acquisition (Kuperman et al., 2012) as control variables. 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.

Hits

The analysis of correct responses on old trials revealed a significant effect of encoding condition ($b = 0.26$, $p < .001$). Items encoded in the semantic condition were more likely to be correctly identified 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 Figure 1. Participants were also more accurate when responding to items learned at a younger age ($b = -0.18$, $p < .001$).

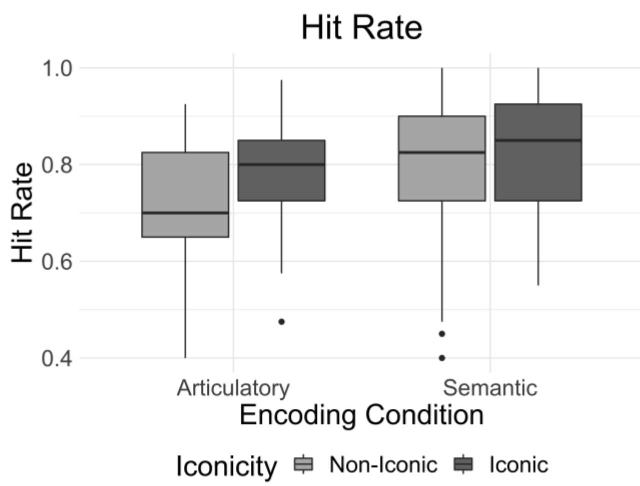

Figure 1: Mean participant hit rates by iconicity and encoding condition.

False Alarms

We next predicted the likelihood of an *incorrect* response on new trials. That is, the likelihood of incorrectly indicating that a new item had been previously seen. The random slope for encoding condition was removed to address a singular fit. This model revealed a significant effect of iconicity ($b = 0.46$, $p < .001$). Participants were more likely to false alarm to iconic items. There was not a significant effect of encoding condition ($b = -0.14$, $p = .14$), nor a significant interaction between iconicity and encoding condition ($b = -0.04$, $p = .23$).[1] See Figure 2. Participants were also more likely to false alarm to longer ($b = 0.49$, $p < .001$) and less phonologically distinct items ($b = -0.40$, $p = .002$).

[1] Note that we ran a version of these analyses that treated iconicity as a continuous predictor. This also did not change the pattern of results.

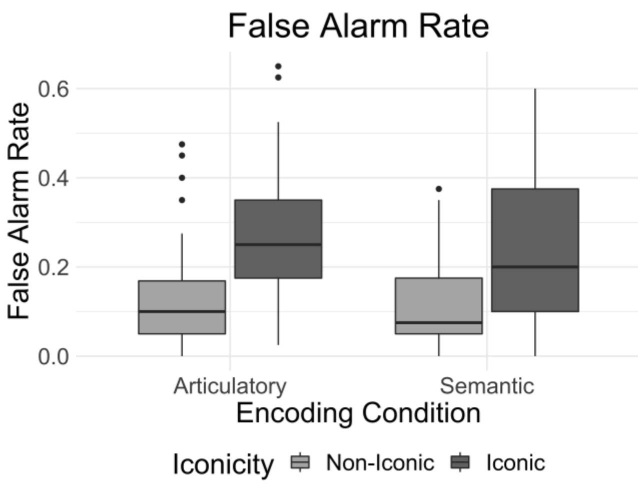

Figure 2: Mean participant false alarm rates by iconicity and encoding condition.

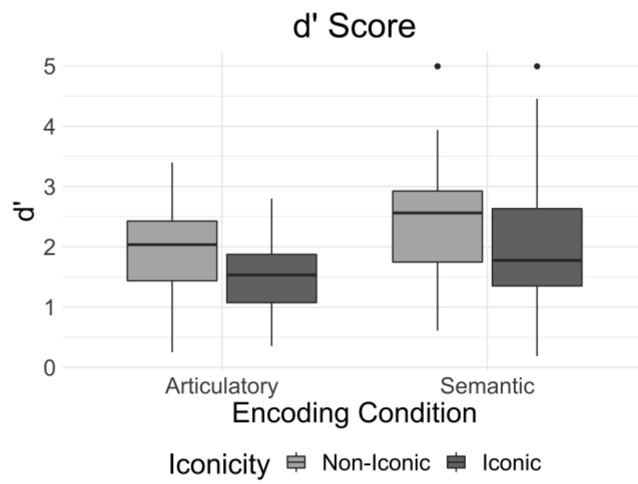

Figure 3: Mean participant d' Scores by iconicity and encoding condition.

d' Score

We next computed a d' score (1) for each participant, separately for iconic and non-iconic items. This serves as an overall measure of a participant's ability to distinguish old and new items.

d' = zHit Rate – zFalse Alarm Rate (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 ($F[1, 125] = 64.80$, $p < .001$) and encoding condition ($F[1, 125] = 10.39$, $p = .002$). 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 Figure 3.

Criterion C

We also calculated each participant's criterion C (2) for iconic and non-iconic items. 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 = - [z(Hit Rate) + z(False Alarm Rate)]/2 (2)

We then ran an ANOVA with criterion C as the dependent variable, with iconicity, encoding condition, and their interaction, as predictors. This revealed a significant effect of iconicity ($F[1, 125] = 190.14$, $p < .001$). 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 Figure 4.

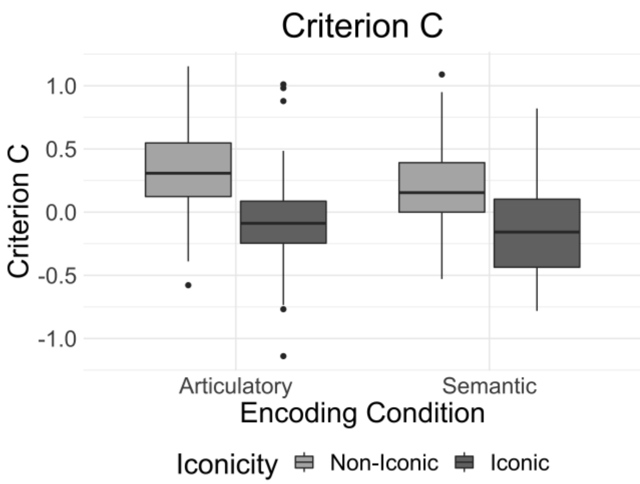

Figure 4: Mean participant Criterion C by iconicity and encoding condition.

Analysis of Memory Megastudies

In order to examine whether these patterns would emerge in a different list context, we made use of two existing recognition memory megastudies (Cortese et al., 2010; Cortese et al., 2015). In these studies, participants memorized one of several lists of 50 words, and then were tested on a list of 100 words. Depending on the study, words were either all monosyllabic or disyllabic. We examined whether published iconicity ratings (Perry et al., 2015; Winter et al., 2017), 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 1,646 words for which all values were available. All predictors were standardized.

We found that iconicity was predictive of a higher false alarm rate (b = 0.008, $p < .001$), and a lower criterion C (b = -0.02, $p < .001$). Iconicity was only a marginal predictor of hits (b = .004, p = .06), and not a significant predictor of d' scores (b = -0.02, p = .09).

General Discussion

Our main goal was to examine whether iconicity would have an effect on recognition memory. We found that iconic words had a higher false alarm rate than non-iconic words. In addition, iconic words had a lower d' score and criterion C than non-iconic words. Notably, these results did not interact with encoding condition. In addition, a supplementary analysis of recognition memory megastudies found that across 1,646 items, iconicity was associated with a higher false alarm rate, and a lower criterion C, in line with the results of the experiment reported here.

Iconicity and Recognition Memory

We now consider these results in the context of existing theories of recognition memory. Such theories can be divided into dual- or single-process theories (Cortese et al., 2015). Dual-process theories propose that 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. This would explain why they were associated with increased false alarms. Such an interpretation would be consistent with the view that iconic words more readily activate sensory features. At first glance, this contrasts with previous reports that concreteness and imageability—other variables related to sensory features—improve memory by increasing hits and lowering false alarms (Khanna & Cortese, 2021). However, there are important differences between iconicity and these variables. Concreteness and imageability quantify the amount of sensory experience associated with a word's referent. Conversely, iconicity is a property that itself *depicts* sensory experience, serving to bring that experience "to the mind's eye" (Murgiano et al., 2021, p. 2) in the absence of a word's referent.

Another important point is that iconicity tends to evoke *auditory* sensory experience rather than visual experience (Winter et al., 2017).[2] This would certainly contrast with imageability. Note that Khanna and Cortese (2021) found perceptual strength (i.e., the amount of sensory experience associated with a concept in its dominant modality; thus including non-visual sensory experience) did not improve memory performance.

An ideal way for future work to test the proposal that iconicity results in a greater feeling of familiarity would be with a remember/know judgment task (see Yonelinas, 2002). In this task, after categorizing an item as old, participants make a further judgment as to whether they remember seeing the item (i.e., recollection) or simply know that they have seen the item, in the absence of a specific memory (i.e., familiarity). This would allow us to test for the separate effects of iconicity on recollection and familiarity.

In addition to increasing false alarms, iconicity also appeared to lower a participant's response criterion. This suggests that participants were more ready to indicate that they had previously seen an item when it was iconic. There is not a great deal of existing literature on item level properties affecting response bias. However, Adelman and Estes (2013) observed a lower criterion for negatively valanced items. This was interpreted as an over-confidence when responding to negative items. In addition, Cortese et al. (2010) observed a

[2] Indeed, our iconic items had significantly greater auditory perceptual strength (M = 1.96) compared to non-iconic items (M = 1.36; b = 0.59, p = .003; Lynott et al., 2020).

lower response criterion for words that were more imageable, more phonologically distinct, and had a greater number of phonological neighbours. These suggest that participants' response bias to individual items *can* be impacted by phonological and semantic variables.

It is also possible to view recognition memory results from a single-process account. In these accounts, items encountered at test produce some output value, and when this value is above a certain threshold the item is categorized as old. One example of such a theory is McClellend and Chappell's (1998) item noise account. According to this account, the encoding of a given item is based on a subset of its features being stored in memory. 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. When new items encountered at test have many features in common with those studied at test, this should lead to increased false alarms. This would be one way of explaining the increased false alarms for iconic items. Because half of the items studied were iconic, assuming there is a great deal of semantic overlap among these items[3], this should lead to more false alarms for iconic items at test. However, this would not explain why iconicity was also associated with increased false alarms in the megastudy datasets. In fact, previous work has shown that iconic items tend to be more semantically unique when compared to the lexicon in general (Sidhu & Pexman, 2018). This makes it unlikely that the higher false alarm rate in the megastudy datasets was due to semantic similarity among items.

The main conclusion from this experiment is that iconicity increases participants' willingness to respond that they have seen an item before, even if they have not. We interpret this as being due to a greater feeling of familiarity for iconic items, perhaps due to the activation of sensory experience via phonology.

Encoding Condition

We found an effect of encoding condition on both hits and d' scores, suggesting that this manipulation was effective. However, encoding condition did not interact with iconicity for any of our outcome measures. As a "sanity check", we tested whether valence interacted with encoding condition in the prediction of hits. Because the semantic encoding task explicitly asked about valence, this is an interaction that should be especially likely. 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 is theoretically possible for our encoding manipulation to interact with item-level semantic variables.

In the introduction we outlined two general ways in which iconicity could affect processing: through special links between the semantic system and modality-specific features, or through special links between phonology and semantics. The fact that iconicity did not interact with encoding condition goes against the idea that effects of iconicity emerge via phonological-semantic links. However, it is important to note that the effects of iconicity that we did observe (i.e., in false alarms and criterion C) are localized to the *recognition* phase (as opposed to the encoding phase). False alarms occur to items only seen in the recognition phase, and criterion C is set during the recognition phase. Thus, we would not in fact expect iconicity to interact with a manipulation of encoding for either of these measures. Indeed, there was not a significant effect of encoding condition on either false alarms or criterion C. An informative follow up would be to manipulate the conditions under which participants *retrieve* information. Perhaps having participants focus on words' forms at retrieval would amplify the effect of iconicity on false alarms.

Conclusions

There is clearly much work to be done understanding the effects of iconicity on recognition memory. The present work was a first step in this regard, by showing that participants are more likely to false alarm to, and set a lower response criterion for, iconic items. This may be indicative of iconicity's ability to bring sensory features to the mind's eye. These results add to the growing body of work showing that iconicity has a real and measurable impact on language processing.

References

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, *129*, 530-535.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1-48.

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*, 977-990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904-911.

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, *18*, 595-609.

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology*, *68*, 1489-1501.

Craik, F. I. (2002). Levels of processing: Past, present... and future?. *Memory*, *10*, 305-318.

[3] Indeed, there was a smaller average cosine distance (a proxy for distance in meaning; Mandera et al., 2017) among iconic items ($M = 0.86$) as compared to non-iconic items ($M = 0.94$; $b = 0.07$, $p < .001$).

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*, 603-615. doi:10.1016/j.tics.2015.07.013

Hashimoto, T., Usui, N., Taira, M., Nose, I., Haji, T., & Kojima, S. (2006). The neural mechanism associated with the processing of onomatopoeic sounds. *Neuroimage*, *31*, 1762-1770.

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*, 253-310.

Joo, I. (2020). Phonosemantic biases found in Leipzig-Jakarta lists of 66 languages. *Linguistic Typology, 24*, 1–12.

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*, 1-8.

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*, 1-15.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978-990.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*, 1-26.

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*, 2207-2222.

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

Lockwood, G., & Tuomainen, J. (2015). Ideophones in Japanese modulate the P2 and late positive complex responses. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00933

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*, 1271-1291.

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.

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.

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.

Murgiano, M., Motamedi, Y., & Vigliocco, G. (2021). Situating language in the real-world: the role of multimodal iconicity and indexicality. *Journal of Cognition, 4*.

Peeters, D. (2016). Processing consequences of onomatopoeic iconicity in spoken language comprehension. In *38th Annual Meeting of the Cognitive Science Society*. Cognitive Science Society, Philadelphia, PA.

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*. doi: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*, e0137147.

R Core Team (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Revill, K. P., Namy, L. L., DeFife, L. C., & Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain and Language*, *128*, 18-24.

Saussure, F. (1916). The nature of the linguistic sign. In P. Meisel, & H. Saussy (Eds.), *Course in General Linguistics* (pp. 65–70). New York, NY: Coumbia University Press.

Sidhu, D. M., & Pexman, P. M. (2018). Lonely sensational icons: semantic neighbourhood density, sensory experience and iconicity. *Language, Cognition and Neuroscience*, *33*, 25-31.

Sidhu, D. M., Vigliocco, G., & Pexman, P. M. (2020). Effects of iconicity in lexical decision. *Language and Cognition*, *12*, 164-181.

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, 1390–1398.

Sidhu, D., Williamson, J., Slavova, V., & Pexman, P. M. (2021). An Investigation of Iconic Language Development in Four Datasets. *Journal of Child Language.*

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M.S. (2021). afex: Analysis of Factorial Experiments. R package version 0.28-1.

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,* 776-782.

Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C. R., Milan Maillo, I., & Perniss, P. (2020). Onomatopoeias, gestures, actions and words in the input to children: How do caregivers use multimodal cues in their communication to children? *Proceedings of the 41st Annual Conference of the Cognitive Science Society.* Cognitive Science Society, Austin, TX

Vigliocco, G., Zhang, Y., Del Maschio, N., Todd, R., & Tuomainen, J. (2020). Electrophysiological signatures of

English onomatopoeia. *Language and Cognition, 12,* 15-35.
Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*, 1191-1207.
Winter, B., & Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa: a journal of general linguistics*, *6*.
Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic?: Iconicity in English sensory words. *Interaction Studies*, *18*, 443-464.
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*, 1-11.
Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971-979.
Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441-517.