Iconicity in Word Learning: What Can We Learn from Cross-Situational Learning Experiments?

Matthew Jones (m.jones.12@ucl.ac.uk)
Gabriella Vigliocco (g.vigliocco@ucl.ac.uk)
Department of Experimental Psychology, University College London
26 Bedford Way, LONDON, WC1H 0AP UK

Abstract

Iconicity, i.e. resemblance between form and meaning, is a widespread feature of natural language vocabulary (Perniss, Thompson, & Vigliocco, 2010), and has been shown to facilitate vocabulary acquisition (Imai, Kita, Nagumo, & Okada). But what kind of advantage does iconicity actually give? Here we use cross-situational learning (Yu & Smith, 2007), to address the question for sound-shape iconicity (the so-called kiki-bouba effect, Ramachandran & Hubbard, 2001). In contrast to Monaghan, Mattock, and Walker (2012), Experiment 1 suggests that the iconicity advantage comes from referential disambiguation rather than more efficient memory encoding. Experiments 2 and 3 replicate this result, and moreover show that the kiki-bouba effect is roughly equally strong for sharp and rounded shapes, a property that classic experiments were unable to confirm, and which has implication for the effect's mechanism

Keywords: iconicity; cross-situational learning; kiki-bouba, vocabulary acquisition; artificial language learning; sound-symbolism

Introduction

Iconicity as Widespread in Natural Language

The meaning of a word does not determine its form, but wordforms are often *motivated* by iconic relationships with meaning. In English, iconicity can be found in onomatopoeia, (e.g. *bang, miaow*). *Outside* the Indo-European family, iconicity is more pervasive. Large iconic lexica are reported for many unrelated languages, signed and spoken (see Perniss et al., 2010).

Such iconicity is not limited to sounds. In Japanese reduplication of syllables indicates repetition of an event, and voicing of an initial consonant indicates object size (e.g. *gorogoro* – heavy object rolling repeatedly; *korokoro* – light object rolling repeatedly; Perniss et al., 2010).

Iconicity and Word Learning

Experimental work shows that Japanese sound-symbolic words are easier for 3-year-olds to learn than non-iconic words, whether the children are Japanese or English speakers (Imai et al., 2008; Kantartzis, Imai, & Kita, 2011; Yoshida, 2012).

Observational research suggests a role for iconicity in vocabulary acquisition outside the lab. Japanese children acquire iconic words early (Maeda and Maeda, 1983), and in keeping with this Saji and Imai (2013) find that Japanese caregivers use more sound-symbolic and onomatopoeic words speaking to their toddlers than to adults.

Perry, Perlman, and Lupyan (2015), analysing English and Spanish, found a negative correlation between iconicity and age of acquisition: even in Indo-European languages, it may play a role in acquisition.

However, substantial questions remain about what advantage iconicity confers on word learning. Does iconicity kick in after the problem of identifying a word's meaning has already been solved, with iconic words being encoded in memory more quickly or efficiently? Or does iconicity help by facilitating referential disambiguation? Experiment 1 will begin to address this question.

Sound-Shape Iconicity

A near-universal form of iconicity is the association between certain sounds (e.g. back vowels and high sonority consonants) with heavy, slow, rounded objects; and others (e.g. front vowels and low sonority consonants) with small, quick, jagged objects (Ramachandran & Hubbard, 2001). In standard demonstrations, participants are given images of two 2-dimensional shapes, one round, the other spiky. The majority pairs 'kiki' with a spiky shape, and 'bouba' with a rounded shape. (Dingemanse & Lockwood, 2015).

The mechanism of sound-shape iconicity is, however, uncertain. The effect could arise from correlated input from different sensory modalities. Alternatively, Ramachandran and Hubbard (2001) suggest it is a reflection of cross-modal analogy between the articulatory gestures required to produce the labels and the visual properties of the shapes (p. 19). They also suggest that 'cross-wiring' (p. 21) of auditory and visual brain maps may create an unmediated link.

Another possible explanation is more literal correspondences between speech sounds and lip shape. 'Bouba' involves literal rounding of the lips – visual or motoric representations of lip rounding could mediate between 'round' sounds and objects. This account predicts an asymmetry: round sound-shape associations should be stronger than spiky ones, because round sounds involve literal rounding of an articulator, whereas spiky sounds do not involve any comparable spikiness. Some prior ERP evidence suggests that the round association may be stronger than the spiky one in processing (Kovic, Plunkett, & Westermann, 2010 – though their paradigm could not separate the associations behaviourally). This dissociation is not something that the classic kiki-bouba experiment is able to test: with two words and two shapes, one (hypothetically stronger) sound-shape pairing would automatically determine the other (weaker or absent) pairing. However our

Experiments 2 and 3 will represent some of the first work ever to address this question.

Cross-Situational Learning

Monaghan, Mattock, and Walker (2012) established that the classic kiki-bouba effect is found using the cross-situational learning (CSL) paradigm. CSL takes the form of a series of trials where a word is appears along with a number of possible referents (Yu & Smith, 2007). Any single trial is ambiguous, and initially participants must guess, but information can be integrated across trials to solve this trial-level referential ambiguity.

Monaghan et al.'s referents were round and spiky shapes, and their names were iconically round or iconically spiky nonwords. Half of shapes received iconically congruent names (e.g. rounded shape-round name), and the other half iconically *in*congruent (e.g. rounded shape-spiky name). In each trial the participant saw two shapes, heard one name, and indicated which shape the name belonged to. Would accuracy in choosing the correct referent would be higher for congruently named items?

Monaghan et al. found that congruence was no advantage in the first block, but became advantageous in later blocks. Moreover the advantage was only present in trials where the unnamed shape (the *foil*) was from the opposite category to the target. From the first result they concluded that iconicity indeed supports word learning (perhaps e.g. in the sense of facilitating more rapid or robust memory encoding of iconic names); from the second they that the advantage pertains to category level information, and not to information distinguishing individual words within categories.

These results are somewhat surprising. The classic kiki-bouba experiment involves guessing names. If iconicity is expressed there, then why wouldn't it be expressed in the first block, when participants are forced to guess name-referent pairings? If that bias were expressed from the start, then iconicity might support referential disambiguation. Experiment 1 takes up this question. Experiments 2 and 3 attempt to tease apart effects of round vs. spiky iconicity.

Experiment 1

Methods

Participants 24 adult native English monolinguals (13 women, $M = 29.7 \pm 10.0$).

Visual Stimuli (Shapes) Sixteen shapes were created using the GNU Image Manipulation Program. Eight 'spiky' shapes were created using randomised parameters. Eight 'rounded' shapes were created by taking each spiky shape and using its corners as fixed points for Bezier curves, then scaled by eye to match for perceived size (see Figure 1). Stimuli were 600*600 pixel images comprising the shape in black on a white background.

Auditory Stimuli (Names) Names were constructed on the basis of LetterScore, a text-based index of sound-shape iconicity: All consonant-vowel pairings in English orthography that feature consonants with only one canonical pronunciation ($N = 85$; c, g, q, and x were excluded) were rated by monolingual Anglophones who did not participate in other studies ($N = 28$, 12 women, 28.5 ± 12.0 years old) on a ten-point scale anchored by a circle (1) and a star (10).

Eight of the names were constructed using syllables that received the spikiest ratings (example: *tikiza*), eight using the syllables that received the roundest ratings (example: *mujo*). For each category of name, two were one syllable long, four were two, and two were three. Recordings were made by a female native speaker of North American English, pronouncing the words as she considered natural.

Subsequently, word recordings were normed as part of a wider norming study. 101 native English speakers ($M = 32.4 \pm 9.7$, 41 women) were each given 118 speech tokens to rate (largely from another study), meaning that each speech token was rated about ten times. The study was performed using Qualtrics (2015). In each trial, the participant saw a seven-point ratings scale. '1' represented the roundest rating, and '7' the spikiest (counterbalanced for half of participants). The mean of each token's ratings was then taken. This was its WordScore. Names for Experiments 2 and 3 were also rated for WordScore (see below). T-tests confirm that spiky names ($M = 4.71 \pm 0.53$) were rated as significantly spikier by WordScore than round names ($M = 2.90 \pm 0.43$) ($p < .001$, $t(13.4) = 7.54$, difference = 1.81, 95% CI [1.29 ,2.33]; Cohen's $d = 3.77$).

Apparatus and Procedure The study was run using Matlab 7.4.0 on an IBM compatible PC equipped with a 15" monitor (resolution: 1024×768). For each participant, half the shapes in each category received congruent names (e.g. round names for round shapes). The other half received incongruent names (e.g. spiky names for round shapes). Assignment of names to shapes was counterbalanced between participants.

The experiment took the form of a series of 256 trials, each featuring two shapes on screen (one to the left and one to the right – see Figure 1) and one name (played through headphones). The name belonged to one of the two shapes (this shape was the target, the unnamed shape being the foil). The participants stated which shape the name belonged to (by pressing the left or right arrow). Participants received no feedback and had to guess at first, but in time could infer which name belonged to which shape by noting that each name only consistently appears with one shape.

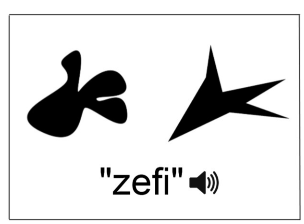

Figure 1: A cross-situational learning trial (note that names were presented aurally, not in text)

Trials were grouped into four blocks of 64 trials, as in Monaghan et al.. Within each block each name appeared four times, and each shape appeared four times as a target and four times as a foil. The number of times each shape appeared on each side of the screen in each role was counterbalanced, as was the number of appearances by each shape as a foil for a target from its own category vs. the opposite category. The same name was not permitted two trials in a row. Otherwise trials were randomised.

Results

Trials with reaction times of less than 0.5 seconds or more than 25 seconds were removed.

Statistical Methods Data was analysed using the LMEM package lme4, version 1.1-12 (Bates, Maechler, Bolker, & Walker, 2015) running in R version 3.2.1 (R Core Team, 2015). In addition to random intercepts for names and participants, we also included random slopes. We aimed for a design-driven maximal random effects structure (see Barr, Levy, Scheepers, & Tily, 2013), but were limited in the number of random effects we could fit. For participants we included random slopes for linear block, congruence, category of foil (coded as same or different to category of target), and the congruence-category of foil interaction. For names, we were limited to random effects slopes for congruence, category of foil, and their interaction. Block was coded linearly (1 = -1.5, 2 = -0.5, 3 = 0.5, 4 = 1.5), and both other variables were contrast coded (incongruent = -0.5, congruent = 0.5; same category foil = -0.5, different category foil = 0.5). Our predictor was accuracy: i.e. whether participants answered correctly on given trials.

Overall Analysis The omnibus model showed reliable effects only of linear block (β = 0.84, 95% CI [0.658, 1.022], z = 9.066): participants learned; and congruence (β = 0.417, 95% CI [0.128, 0.706], z = 2.826): participants performed better with congruent names (see Figure 2). The congruence-category of foil interaction was also significant (β = 0.702, 95% CI [0.203, 1.201], z = 2.759): congruence represents more of an advantage when the foil shape is from the opposite category to the target.

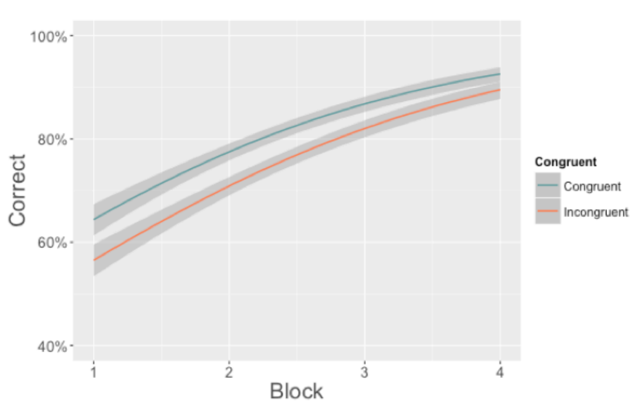

Figure 2: Graph of the predictions by block and congruence of the final omnibus model for Experiment 1. Error bars represent 95% CIs.

See https://github.com/JMJofficial/Jones_Vigliocco_2017 more graphs, and for graphs of Experiments 2 & 3.

Block 1 Monaghan et al. found that congruence interacted with block. Crucially, there was no congruence advantage in the first block, implying that the benefit of congruence was to memory encoding rather than kiki-bouba style response bias. By contrast, we found no interaction between congruence and block (z < 0.7), suggesting an advantage from the first block. To test this, we fitted a model for the first block only. There *were* reliable effects of congruence (β = 0.328, 95% CI [0.047, 0.609], z = 2.288): performance was better in congruent trials; and of the interaction between congruence and category of foil (β = 0.842, 95% CI [0.368, 1.316], z = 3.484): the benefit of congruence was stronger in different-category-foil trials. Note that this cannot be attributed to differences in design, as the number and structure of our trials was identical.

To exclude the possibility that this is the result of learning within the first block, we took the 187 trials with a different category foil where a participant encountered a name for the first time, and fitted a LMEM featuring only a fixed intercept, and random intercepts by participant. On 56.1% of trials participants chose the iconically congruent referent for the name. The model's intercept was not reliably different from zero under two-tailed interpretation (β = 0.247, 95% CI [-0.048, 0.549], z = 1.678), but under a one-tailed interpretation, the intercept was significantly different from zero at p = .047. Thus though this analysis has low power, it suggests a sizable bias towards iconic matches before learning has taken place, which can only be explained by the bias/referential disambiguation account.

In conclusion, we largely replicated Monaghan, Mattock, and Walker's (2012) findings, but found an advantage of iconicity from the first block. This difference with Monaghan et al. – and the fact that iconic congruence is only an advantage when the foil is from the opposite category – is consistent with the possibility that iconicity biased participants towards the right answer in trials where they were forced to guess, effectively assisting with referential disambiguation. The discrepancy with Monaghan et al. may be due to name stimuli: while we tailored ours to maximize iconicity, they created theirs on the basis of phonetic features, which do not correlate perfectly with iconicity (e.g. Monaghan et al. used plosives as spiky sounds, but [b] – a plosive – is widely deemed to sound round, cf. bouba). Next we move on to two further experiments aimed at testing the relative contribution of roundness and spikiness to sound-shape iconicity.

Experiments 2 and 3

Experiment 2 and 3 aim to clarify the mechanism of sound-shape iconicity by modifying Experiment 1 in order to test the effect of round-to-round and spiky-to-spiky iconicity separately, something previous experiments have been unable to do. This is achieved by using iconically neutral names as well as round and spiky names. Experiment 2 yielded marginally significant results, so Experiment 3 was a replication to attempt to verify whether the effect was real. Both were then submitted to omnibus Bayesian statistics.

We opted for a two-condition design. Each condition is of the same format as Experiment 1, and each features both round and spiky shapes, but one condition features round and neutral names only, the other features spiky and neutral names only, thus avoiding problems related to tasks that involve discrimination between a round and spiky alternative. If one class of name is less iconic then we would expect minimal benefit of one class of shape being paired with that class of name vs. a neutral name.

Experiment 2: Methods

Participants were 32 adult native English monolinguals (17 women, $M = 23.3 \pm 4.4$).

Visual Stimuli (Shapes) The eight round and eight spiky shapes used in Experiment 1 were combined with an additional eight of each, created in the same manner.

Auditory Stimuli (Names) 32 names were generated using previously normed syllables (see Experiment 1) - eight from round syllables, eight from spiky, and 16 from neutral; and recorded as in Experiment 1. T-tests confirm that the spiky names ($M = 4.71 \pm 0.53$) were rated as spikier than the round names ($M = 2.90 \pm 0.43$) ($p < .001$, $t(13.4) = 7.54$, difference = 1.81, 95% CI [1.29, 2.33]; Cohen's $d = 3.77$). Moreover, neutral names ($M = 3.77 \pm 0.80$) were rated as less spiky than spiky names ($p = .002$, $t(20.0) = 3.46$, difference = 0.94, 95% CI [0.37, 1.51]; Cohen's $d = 1.31$), and less round than round names ($p = .002$, $t(21.8) = 3.46$, difference = 1.04, 95% CI [0.35, 1.39]; Cohen's $d = 1.24$).

Apparatus and Procedure Every participant took part in a round and a spiky condition. Each condition was of identical form to Experiment 1. One of the two conditions was the 'round' condition. In this condition half of the shapes were round and half spiky (eight of each), and, crucially, half of the names were round and half neutral. The other condition was the 'spiky' condition – which again had eight round and eight spiky shapes, but by contrast had eight neutral names and eight spiky names (fresh shapes and neutral names were used in the second condition). Shapes, neutral names, and condition order were counterbalanced across participants.

Here congruence is defined within whichever half of the putative round-spiky spectrum of sounds the condition in question covers. E.g. in the round condition, round name-round shape pairings were considered congruent and round name-spiky shape pairings were considered incongruent. However, neutral name-spiky shape pairings were considered congruent for the purposes of the following analysis. The reverse was done for the opposite condition. This format was so that we could apply the same kinds of analysis as for Experiment 1 to keep results comparable.

Experiment 2: Results

Data were analysed as in Experiment 1. The additional variable of condition was coded Round = -0.5, Spiky = +0.5.

In the omnibus model, both linear ($\beta = 0.722$, 95% CI [0.619, 0.824], $z = 13.835$) and quadratic block ($\beta = -0.11$, 95% CI [-0.179, -0.041], $z = -3.116$) were reliable predictors: performance improved over the blocks, with improvement being faster between early than late blocks. Category of foil was also a reliable predictor ($\beta = 0.162$, 95% CI [0.015, 0.309], $z = 2.163$): performance was better on trials with foils from the opposite category to the target. Finally, the interaction between congruence and category of foil was reliable ($\beta = 0.286$, 95% CI [0.013, 0.56], $z = 2.05$): performance was better on congruent trials as long as the target and foil were from different categories (the main effect of congruence was not reliable: $\beta = 0.027$, 95% CI [-0.14, 0.194], $z = 0.317$; perhaps because iconic contrast was less pronounced than in Experiment 1).

The crucial interaction between condition and congruence was marginally reliable in the expected direction, implying that congruence was more of an advantage in the round condition ($\beta = -0.208$, 95% CI [-0.482, 0.066], $z = -1.486$): an inconclusive result (though the same was not true for the three way interaction adding category of foil: $\beta = -0.047$, 95% CI [-0.377, 0.283], $z = -0.277$).

As with Experiment 1, we analysed **Block 1** in isolation. There was a reliable effect of condition ($\beta = 0.177$, 95% CI [0.001, 0.354], $z = 1.967$): performance was better in the spiky condition. Though there was no overall effect of congruence ($z < 1.3$), there was a reliable effect of category of foil ($\beta = 0.173$, 95% CI [0.022, 0.325], $z = 2.246$): performance was better when the foil and target were from different categories, and a reliable interaction between congruence and category of foil ($\beta = 0.39$, 95% CI [0.105, 0.675], $z = 2.679$): that congruence was advantageous when foil and target were from different categories. Given that previous results suggest that the congruence advantage is in different-category-foil trials, this means that as in Experiment 1, and in contrast to the results of Monaghan et al., iconic congruence was an advantage from the outset. Finally, there was a reliable interaction between congruence and condition: for Block 1 (as was marginally the case for the omnibus model), the effect of congruence was stronger in the round condition ($\beta = -0.315$, 95% CI [-0.628, -0.002], $z = -1.975$).

To summarise, Experiment 2 largely replicated the results of Experiment 1. Additionally, we did not find an unambiguously reliable difference between round and spiky conditions in terms of iconicity advantage. However, the marginally reliable interaction is possible evidence for round iconicity being stronger. Experiment 3 is a near-replication aimed at clarifying this.

Experiment 3: Methods

Participants were 32 adult native English monolinguals (21 women, $M = 21.8 \pm 3.2$).

Visual Stimuli (Shapes) Were as Experiment 2.

Auditory Stimuli (Names) A fresh set of 32 names (eight round, eight spiky, and 16 neutral) were generated as in Experiment 2. An additional factor was controlled: number and distribution of phonemes in each category of name. Names were recorded as in Experiments 1 and 2.

T-tests confirm that the spiky names ($M = 4.80 \pm 0.39$) were rated for WordScore as spikier than the round names ($M = 2.90 \pm 0.63$) ($p < .001$, $t(11.7) = 7.19$, difference = 1.90, 95% CI [1.32, 2.47]; Cohen's $d = 3.60$). Moreover, neutral names ($M = 3.80 \pm 0.86$) were rated as less spiky than spiky names ($p < .001$, $t(22.0) = 3.91$, difference = 1.00, 95% CI [0.47, 1.53]; Cohen's $d = 1.35$), and less round than round names ($p = .01$, $t(18.5) = 2.89$, difference = 0.90, 95% CI [0.25, 1.55]; Cohen's $d = 1.24$).

Apparatus and Procedure Were as in Experiment 2.

Experiment 3: Results

Data were analysed as in Experiment 2. The omnibus model featured reliable effect of linear block ($\beta = 0.739$, 95% CI [0.589, 0.889], $z = 9.641$): participants learned. However, it featured no other significant predictors ($|z| < 2.0$). In this respect, it was different from Experiments 1 and 2, both of which showed some advantage of congruence. However, the coefficients for both congruence ($\beta = 0.177$, 95% CI [-0.003, 0.357], $z = 1.929$) and the congruence-category of foil interaction ($\beta = 0.178$, 95% CI [-0.081, 0.437], $z = 1.346$) were in the expected direction, with congruence qualifying as marginally reliable. Crucially, the congruence-condition interaction did not approach reliability ($\beta = -0.047$, 95% CI [-0.326, 0.233], $z = -0.327$), and neither did the interaction between congruence, category of foil, and condition ($\beta = -0.147$, 95% CI [-0.464, 0.17], $z = -0.908$) suggesting that if there was an effect of congruence, it was no stronger in the round condition.

Again we analysed **Block 1** in isolation. This time there were no reliable predictors ($z < 1.3$ in each case). However note that this parallels the omnibus model, which featured no reliable predictors except block. Thus these results are silent on the question of the nature of the iconic advantage as the advantage failed to show up overall (probably due to a smaller iconic differences between words than in Experiment 1 leading to a weaker effect and Type II error).

Thus Experiments 2 and 3 gave somewhat contradictory results, with Experiment 2 showing a marginally reliable interaction between condition and congruence in the expected direction, and Experiment 3 showing no such thing. To attempt to resolve this, we submitted both sets of results to Bayesian statistics, which have the capability to confirm the null, and make it unproblematic to add more data to an analysis as one goes along (Kruschke, 2011).

Experiments 2 and 3: Bayesian Analysis

Models Not having a clear prior for the alternative hypothesis, we opted for Bayesian parameter analysis (Kruschke, 2011). We use the R package rstanarm (Gabry & Goodrich, 2016). We examined 95% Highest Density Intervals for parameter estimates (HDIs): the highest average density continuous interval containing 95% of posterior probability distribution. If this region excludes zero we can treat a predictor as reliable.

We based our priors on Gelman, Jakulin, Pittau, and Su's (2008) recommendations. All variables were centred at zero and scaled so as to have a standard deviation of 0.5. Priors (which were defined for the log odds ratios used as the models' parameters rather than for raw probabilities) took the form of Cauchy distributions.

Models were similar to the models used for Experiments 2 and 3, but a predictor and a by-subjects random slope were added for condition order. All two- and three-way interactions were included.

Results There were credible effects of linear ($\beta = 1.638$, 95% HDI [1.45, 1.835]) and quadratic ($\beta = -0.155$, 95% HDI [-0.255, -0.06]) block, condition order ($\beta = 0.472$, 95% HDI [0.322, 0.620]), and category of foil ($\beta = 0.091$, 95% HDI [0.019, 0.164]): participants performed better on trials where the target and foil shapes came from different categories. The HDIs for the main effect of congruence encompass zero ($\beta = 0.087$, 95% HDI [-0.022, 0.198]), albeit narrowly. However, there is a credible interaction between congruence and category of foil ($\beta = 0.242$, 95% HDI [0.114, 0.369]), indicating that an advantage for congruence is present when the target and foil are from different categories. There were interactions between condition order and both linear ($\beta = 0.329$, 95% HDI [0.196, 0.462]) and quadratic ($\beta = -0.213$, 95% HDI [-0.333, -0.094]) block, indicating that initial learning was faster in the second condition. There was also a difficult-to-interpret interaction between quadratic block, condition order, and congruence ($\beta = -0.253$, 95% HDI [-0.498, -0.009]). However, overall, the Bayesian analyses confirm the earlier inferential statistics.

Turning to the crucial interaction of congruence with condition type: the posterior distribution for the interaction between congruence and condition type is narrow compared to the prior, and centred close to zero ($\beta = -0.022$, 95% HDI [-0.165, 0.125]). If we assume the largest absolute value in the HDI, and take the intercept as our baseline, the difference between the levels of the interaction is 84.2% versus 85.3%. This is the same as the difference when the main effect of congruence is examined in the same way, assuming the mean of the posterior.

Even if we assume that the extreme values of the HDI are correct, the effect of the congruence-condition interaction is no bigger than that of congruence *tout court* (which applies to both conditions). Thus there is clearly less support for the interaction between congruence and condition type than for the effect of congruence across conditions, and given the

HDIs encompass zero, our results are consistent with their being no congruence-condition interaction.

Discussion

We presented evidence that iconicity enhances performance in a statistical learning paradigm. Experiment 1 was a replication of Monaghan et al. (2012), and thus in a sense not novel, but (as small but theoretically interesting differences in our respective results underscore) the value of replication is increasingly recognised in cognitive science. Close analysis of the beginning of Experiment 1 (supported by the results of Experiment 2), and the consistent tendency for iconicity to be a greater advantage when the foil presented during the trial does not also match the name, suggest that the benefit of iconicity in these experiments is to do with picking out the right referent during a particular trial rather than in to do with learning in some other sense (*contra* Monaghan et al.). Thus one role of iconicity in vocabulary learning may be in referential disambiguation, in line with evidence that people guess iconic word meanings in unfamiliar languages above chance (Imai et al., 2008).

The second set of findings relate to the relative importance of rounded-rounded and spiky-spiky mappings in sound-shape iconicity. One hypothesis is that rounded sounds are associated with rounded lip shape, and that the iconicity arises from sound-shape correspondences during speech production and comprehension (Ramachandran and Hubbard, 2001). If this is indeed the mechanism for sound-shape iconicity, we would expect rounded associations to be primary, and spiky associations to arise later through something like a principle of contrast. If this were the case then rounded associations should be stronger than spiky associations (as suggested in Kovic. et al., 2010).

Experiment 2 and 3 tested this possibility by separating round and spiky iconicity into two separate conditions, and seeing whether iconic congruence exerted a stronger effect in one or the other. Experiment 2 appeared to suggest (with marginal reliability) that iconicity improved performance in the round condition, but not the spiky. However, this asymmetry failed to replicate in Experiment 3. We therefore submitted data from both experiments to a Bayesian analysis. Though the results were somewhat inconclusive, they suggest that any asymmetry between the conditions is a smaller effect than the overall influence of iconic congruence, and indeed they are consistent with there being no asymmetry at all. However, this may be different in the case of production, which would force motoric and perceptual engagement with lip shape (Jones et al., in prep.).

Our results advance our understanding of iconicity's role, suggesting it supports referential disambiguation.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

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

Gabry, J. and Goodrich, B. (2016). rstanarm: Bayesian Applied Regression Modeling via Stan. R package version 2.9.0-3

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics, 2*(4), 1360-83.

Imai, M. Kita, S., Nagumo, M., and Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*, 54–65.

Jones, M., Vinson, D., Clostre, N., Lau Zhu, A., Santiago, J., and Vigliocco, G. (in prep.). Iconicity emerges in a model of language change.

Kantartzis, Imai, & Kita (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science, 35*, 575-586.

Kovic, V., Plunkett K., and Westermann, G. (2010). The shape of words in the brain. *Cognition, 114*, 19–28.

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science 6*(3), 299-312.

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: a review of behavioural, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology, 6*(1246).

Maeda, T., and Maeda, K. (1983). *Yoji no goihattatsu no kenkyu [Investigation of a Child's Lexical Development].* Tokyo: Musashino Shoin.

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in word learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 38*(5), 1152-1164.

Perniss, P., Thompson, R. L., & G. Vigliocco (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Language Science, 1*(227).

Perry, L. K., Perlman, M., Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS 1, 10*(9): e0137147. DOI:10.1371/journal.pone.0137147

Ramachandran, V., & Hubbard, E. (2001). Synaesthesia: A window into perception, thought, and language. *Journal of Consciousness Studies, 8*(1), 3-34.

Saji, N. & Imai, M. (2013). Onomatope kenkyu no shatei – chikadzuku oto to imi [Sound Symbolism and Mimetics]. In K. Shinohara & R. Uno (Eds.), *Goishutoku ni okeru ruizousei no kouka no kentou* (pp. 151-166). Tokyo: Hituji Syobo.

Yu, C., & Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*(15), 414–420.