The Bouba Effect: Sound-Shape Iconicity in Iterated and Implicit Learning

Matthew Jones[a] (john.jones.09@ucl.ac.uk), David Vinson[a] (d.vinson@ucl.ac.uk) , Nourane Clostre[a] (n.clostre@ucl.ac.uk), Alex Lau Zhu[a] (a.zhu@ucl.ac.uk), Julio Santiago[b] (santiago@ugr.es), Gabriella Vigliocco[a] (g.vigliocco@ucl.ac.uk)

[a]Division of Psychology and Language Sciences, University College London, London WC1E 6BT, UK
[b]Dept. de Psicología Experimental y Fisiología del Comportamiento, Facultad de Psicología, Universidad de Granada, Campus de Cartuja s/n 180179-Granada, SPAIN

Abstract

Although wordforms are often arbitrarily linked to their meaning, many exhibit iconicity (resemblance between form and meaning). This is especially visible in the lexica of non-Indo-European languages and signed languages. Iconicity has been argued to play a role in grounding linguistic form to real-world experience, rendering language more learnable (Perniss & Vigliocco, in press). Here we examine sound-shape iconicity, the 'kiki-bouba' effect, i.e. the tendency to associate bouba-type labels with round shapes, and kiki-type labels with spiky shapes. In a first experiment we show that this iconicity emerges in the course of iterated learning (presumably because it renders labels more learnable). However, it only emerges for the mapping between *round* shapes and bouba-type labels. In a second experiment (using cross-situational learning, see Monaghan et al., 2012) greater learnability is observed for mappings of the bouba-to-round type but not of the kiki-to-spiky type. We discuss possible mechanisms underlying this difference.

Keywords: cross-modal; cultural evolution; iconicity; iterated learning; language evolution; sound symbolism.

Introduction

Iconicity as a Widespread Feature of Language

The arbitrariness of the wordform has often had the status of a truism (de Saussure, 1983). However, it is also the case that wordforms are often *motivated* by iconic relationships with meaning. In English, iconicity can be found in onomatopoeia, (e.g. *bang, miaow*). However, in languages *outside* the Indo-European family, iconicity is more pervasive. Large iconic or sound-symbolic lexica have been reported for many languages of different families (including sub-Saharan African languages, Australian Aboriginal languages, and Japanese and Korean; see Perniss Thompson, & Vigliocco, 2010).

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

Given the hands' potential for mimesis, it is unsurprising that signed languages are extremely rich in iconicity (see Perniss, Thompson, & Vigliocco, 2010 for a review). E.g. in British Sign Language, the sign for hammer is a hammering gesture, and the sign for *lion* uses the hands as the pouncing cat's paws.

There is evidence, especially in sign languages, that iconic mappings facilitate learning and processing in adults learning novel labels (Perniss et al, 2010 for review). For spoken languages it has been shown that iconicity facilitates word processing in adults (Westbury, 2005); that 3- to 4-month-old infants are sensitive to iconic mappings (Walker et al., 2010); and that iconic words are easier for 3-year-olds to learn than non-iconic words (Imai et al., 2008). This is unsurprising given an embodied perspective: if the semantics of words depend on sensorimotor activation, then words that generate appropriate activation by virtue of their form will automatically be more learnable than arbitrary words, and will enjoy easier encoding, storage, and retrieval (Perniss and Vigliocco, in press).

Sound-Shape Iconicity

A seemingly universal form of iconicity is the association between certain sounds (e.g. back vowels and voiced consonants) with heavy, slow, rounded objects; and others (e.g. front vowels and voiceless consonants) with small, quick, jagged objects (Ramachandran & Hubbard, 2001). In standard demonstrations, participants (being adult speakers of English, 4 month old infants, or people in non-literate, non-industrial societies) are given images of two 2-dimensional shapes, one round, the other spiky. The majority prefers to associate 'kiki' with a spiky shape, and 'bouba' with a round shape. (Maurer, Pathman, & Mondloch, 2006; Ozturk, Krehm, & Vouloumanos, 2013; Bremner et al., 2012). Similar shape-sound associations obtain when the methodology used is implicit learning (Monaghan, Mattock, & Walker, 2012).

The origin of this sound-shape iconicity is, however, unclear. Ramachandran and Hubbard (2001) suggest that the effect comes about as a reflection of cross-modal similarity between the articulatory gestures required to produce the labels and the visual properties of the shape, implying that non-visual representation of articulation mediates between sound and shape (p. 19). Alternatively, they also suggest that 'cross-wiring' (p. 21) of auditory and visual brain maps may create an unmediated link (with associations depending on features of brain architecture). Both of these accounts predict that the effect is present for both 'kiki'-spiky and 'bouba'-round.

There is however, yet another possible explanation: auditory-visual associations between speech sounds and lip shape. Words like 'bouba' involve literal rounding of the

lips - representations of such lip rounding would on this account mediate between 'round' sounds and round objects. Unlike the first two accounts, this latter predicts an asymmetry: round sound-shape associations should be stronger than spiky ones, because round sounds involve visible rounding of an articulator, whereas spiky sounds do not involve any comparable visible spikiness. Importantly, the very few studies in the literature *separately* assessing the strength of bouba-round and kiki-spiky associations suggest a stronger effect for the round association and less (or no) effect for the spiky association (Kovic, Plunkett, & Westermann, 2010). This is not something that the classic kiki-bouba experiment is able to test - because there are only two words and two shapes, the determination of one (hypothetically stronger) sound-shape pairing would automatically determine the other (weaker or absent) pairing. In this study we independently assess the two types of associations by asking whether both will emerge in the course of cultural evolution (iterated learning), suggesting that they render a 'language' more learnable, and - in a second experiment - whether they both provide a learning advantage, relative to neutral labels, in probabilistic learning across situations (Monaghan et al., 2012). The question of which (if either) of the labels is more important will bear crucially on our understanding of the mechanism of the effect.

These two studies do not make it impossible that an oppositional contrast between 'round' sounds and 'spiky' sounds will be set up (allowing spiky iconicity to piggyback on round or *vice versa*), but they do free the two types of iconicity from strict interdependence. Thus if one kind of iconicity emerges as stronger than another in spite of the possibility of an opposition, this will be particularly striking evidence for its primacy.

Experiment 1: Iterated Learning

The *iterated learning model* (Scott-Phillips & Kirby, 2010) approximates cultural evolution using *diffusion chains*, in which a succession of separate participants (or *generations*) each learn from the previous participant (rather like the children's game called 'broken telephone' or 'Chinese whispers'). An initial participant is taught a 'language' of mappings between novel words and visual stimuli, and then tested on names for stimuli they have seen and (unbeknownst to them) similar stimuli they haven't. Crucially, this compels the participant to innovate. Hence when the responses of the first generation are taught to the second generation (participants are unaware of being part of a chain), evolution of the set of names takes place. This process is repeated between the second and third generation, the third and the fourth, and so on.

When participants change the labels, these changes are passed on to the next generation. If these changes make the language more learnable they should be retained in the language (Christiansen & Chater, 2008). Therefore, if iconicity makes vocabulary more learnable, we would expect the process to favour iconic mappings. Crucially, by using an initial language that is neutral with respect to iconicity and seeing whether kiki-bouba like labels emerge for their respective appropriate shapes we can assess each mapping (spiky-to-kiki-like vs. round-to-bouba-like) separately. This is in the format of Experiment 1. Moreover, in order to test whether the paradigm is suitable to elicit iconicity effects in general, we introduced a further manipulation of our visual stimuli (length of motion) which could be realised in the labels in terms of length of word.

Methods

Participants were sixty native speakers of English (32 females, $M = 26.3 \pm 4.4$ years old).

Materials

Visual Stimuli Eighteen video stimuli were used for the study, varying on the dimensions of shape (round vs. spiky), colour (red, green, and blue), and motion (still, upwards, bouncing up and down – see Figure 1 below). Colour was not expected to be of direct interest, but served to provide enough stimuli. Shapes were chosen to maximize the kind of contrast sound-shape iconicity is known to capture. In Figure 1 arrows represent motion (no arrow: still, duration = 0s; single arrow: single upwards motion, average duration = 0.5s; dual up-down arrows: repeated bouncing motion, average duration = 5s).

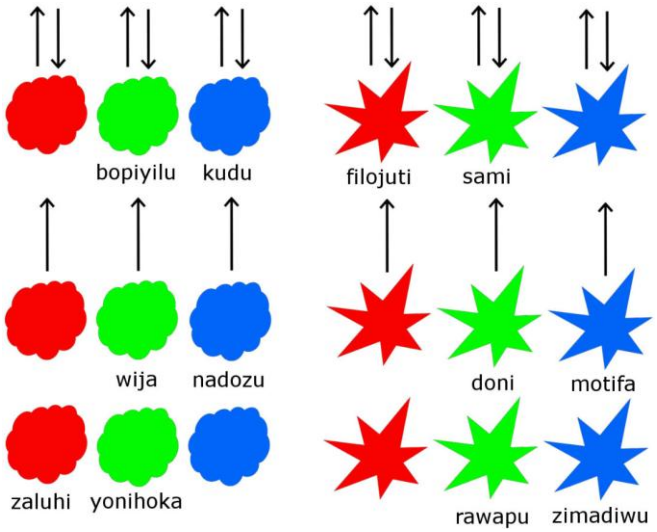

Figure 1: Stills from the stimuli for Experiment 1. Arrows denote motion. Stimuli in the original teaching set appear with their names below.

Labels For the initial language, iconically neutral names (letter strings) were constructed from normative data. All consonant-vowel pairings possible in English orthography were rated by monolingual English speakers who did not participate in the other studies ($N = 28$, 16 males, 28.46 ± 12.03 years old) on a ten-point scale anchored by a circle (1) and a star (10). A centered scale was created by redefining

the mean rating (5.04) as zero. Neutral names were generated by taking syllables whose score did not significantly differ from zero on the normalized scale and concatenating them. Figure 1 shows each of these names below the shape it was initially paired with. Stimuli without names were unseen by first generation participants in the training phase (see below).

Apparatus and Procedure The study was run on E-Prime 2.0 on an IBM compatible PC equipped with a 15" monitor (resolution: 1024×768). The procedure closely followed Kirby, Cornish, and Smith (2008). Participants were told that their task was to learn an 'alien language', which paired certain words with certain pictures. Each language consisted of 18 pairings, which were randomly divided into a SEEN set (12 items) and an UNSEEN set (six items). Participants were trained only on the SEEN set but (unbeknownst to them) tested on both sets to force innovation. A post-experiment questionnaire confirmed that participants typically did not notice the new items.

Participants learned the languages in three rounds of training, with breaks between rounds. Each round was followed by a testing phase. In each training round participants were exposed to the SEEN set in two randomized orders. The first frame of each video was displayed centrally for 1 second before the letter string was displayed below the video. The video + label were visible together for 5 seconds.

In each testing phase, participants were presented with videos and asked to produce the corresponding letter strings by using a standard keyboard. There was no time limit for answering. The first round's test phase contained only half the SEEN set and half the UNSEEN set, with the second round's testing phase containing the other half of each set. The final test phase contained all items. The responses in the final test phase were the only source for the next generation's language. Participants were assigned a position in one of 6 diffusion chains of 10 generations each. The first participant in each chain was trained on the named stimuli depicted in Figure 1. Subsequent generations were trained on a SEEN set randomly drawn from the output of the previous participant with the following constraint: where the previous participant assigned the same name to multiple stimuli, the SEEN set was chosen such as to minimize the number of uses of a name (see Kirby et al., 2008).

Results and Discussion

Analyses were conducted using repeated-measures ANOVAs with: 10/11 (generations) × 2 (shapes) × 3 (colours) × 3 (motion types). Generation was a within-subjects variable (1 to 10 for error, 0 to 10 for sound-shape iconicity and length), as were shape, colour, and motion type. As there is continuity of names across the generations of a particular chain, generations cannot be regarded as independent. Therefore, chain was treated as a random effect.

First, following Kirby et al. (2008) we measured transmission error between generations, an index of learnability useful as a manipulation check. Transmission error was operationalized as Levenshtein edit distance. We found a significant effect of generation on transmission error ($F(9,45) = 5.410$, $p < .001$), i.e. error declined over the generations - the languages became more learnable.

Sound-Shape Iconicity was assessed using a metric called LetterScore, obtained through the norming studies that produced the syllables for the initial language. Each letter's LetterScore was the mean of the ratings of the syllables it appeared in. A letter that tended to appear in spiky-sounding syllables would receive a spiky score, a letter that appeared in round-sounding syllables would receive a round score. A word's LetterScore is the mean of its letters' LetterScores. Scores were centered on a neutral zero, with positive scores representing spikiness, negative scores roundness. As the initial generation-zero language was chosen so as to have scores of approximately zero, unnamed stimuli were deemed to have LetterScores of zero in the initial language for the purposes of this analysis.

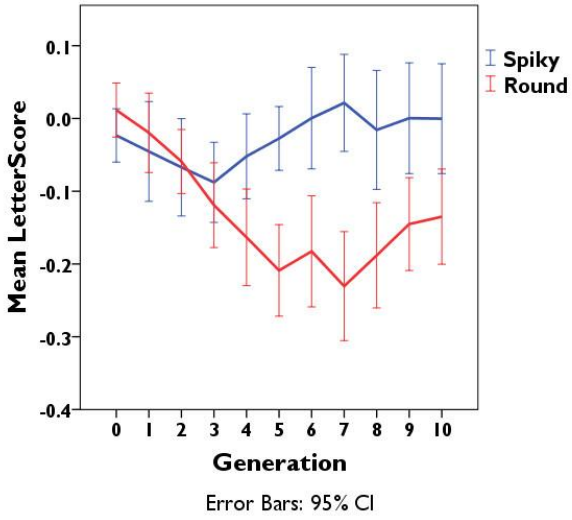

Figure 2 – LetterScore in Experiment 1, averaged across chains. A LetterScore of zero implies neutrality. Positive scores are spiky, negative scores round.

The only significant effect is an interaction between generation and shape ($F(10,50) = 3.486$, $p = .002$). The LetterScores for round and spiky shapes diverged in the expected direction over the course of the experiment (see Figure 2). Figure 2 suggests however that the emergence of iconicity was expressed among round but *not* spiky stimuli. This is confirmed when separate ANOVAs are run for round and spiky stimuli – for round stimuli there is a significant main effect of generation ($F(10,50) = 3.146$, $p = .003$), i.e. LetterScores changed systematically over time, but there is no significant effect for spiky stimuli ($F < 1$).

Length Iconicity The metric for length was number of letters per name. The videos that were not named in

generation 0 (i.e. in the original experimenter generated 'language') were deemed to have names six letters long (the mean for generation 0) for the purposes of this analysis.

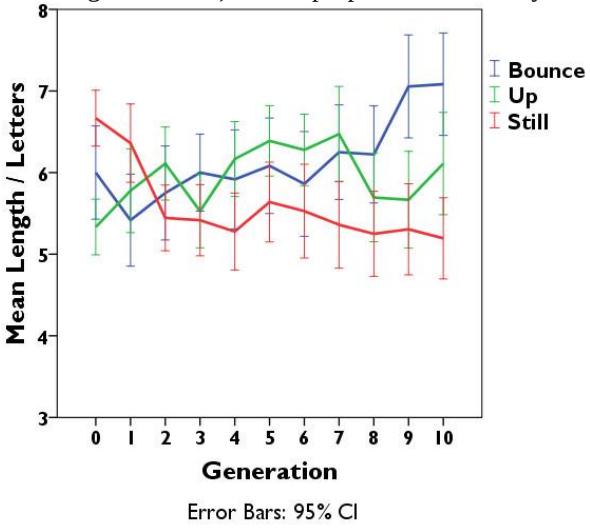

Figure 3 – Word Length in Experiment 1, averaged across chains.

There was a significant interaction between generation and motion ($F(20,100) = 4.647$, $p < .001$), indicating that word length did not change in the same way for all types of motion. By the end of the diffusion chains the length of videos' names positively correlated with the duration of their phases of motion (see Figure 3). An ANOVA conducted on the last five generations by way of confirmation revealed a significant main effect of motion ($F(2,10) = 4.310$, $p = .045$).

Thus in Experiment 1, iconicity emerged for shape and for duration of motion. These changes were accompanied by an increase in learnability. For shape, we found that iconicity emerged for round but, crucially, not spiky shapes. This is in spite of the fact that secondary spiky iconicity could have emerged through oppositional contrast to round iconicity (see introduction). In addition we found that iconicity emerged for motion duration: longer-moving videos had longer names. In Experiment 2 we test the role of shape iconicity in the learning of new words for shapes, again assessing the contribution of round and spiky sounds to sound-shape iconicity separately.

Experiment 2

Experiment 1 suggests that the mapping between bouba-type labels and round shapes may be more iconic than the mapping between kiki-type labels and spiky shapes. Experiment 2 aims to further test any difference between the two mappings, focusing on learning labels for visual stimuli in a cross-situational learning paradigm (Yu & Smith, 2007). Previous work by Monaghan et al. (2012) has used the methodology to demonstrate the existence of an advantage for learning names showing kiki-bouba iconicity (i.e. iconic *congruence*). However it did not assess the independent contributions of spiky sound-shape associations and round sound-shape associations. By including neutral labels, hence having either 'round' and 'neutral', or 'spiky' and 'neutral' names in each condition, this experiment assesses each association separately, without creating as strong an expectation of a round-sound vs. spiky-sound dichotomy (even implicitly). In addition, Experiment 2 complements Experiment 1 by assessing these effects using spoken, rather than written, labels.

Methods

Participants were thirty two adult native English speakers (17 females, $M = 23.3 \pm 4.4$), not including six participants who failed to learn and were excluded and replaced.

Materials

Visual Stimuli (Shapes) Sixteen rounded and sixteen spiky shapes were created and matched for size (see Figure 4 for examples).

Auditory stimuli (names) Thirty two names were generated using previously normed syllables (see Experiment 1) - eight composed of syllables normed as round, eight of syllables normed as spiky, and sixteen of syllables normed as neutral. Names were recorded by a female native speaker of North American English.

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). Each participant took part in two conditions, each being a separate task in the cross-situational learning paradigm (Yu & Smith, 2007).

Trials featured two shapes on screen (one to the left and one to the right – see Figure 4) and one name (played through headphones). The name belonged to one of the two shapes and the participant's task was simply to say which shape the name belonged to (by pressing the left or right arrow). Participants did not receive feedback and had to guess at first. However, over time it was possible to infer which shape the name referred to.

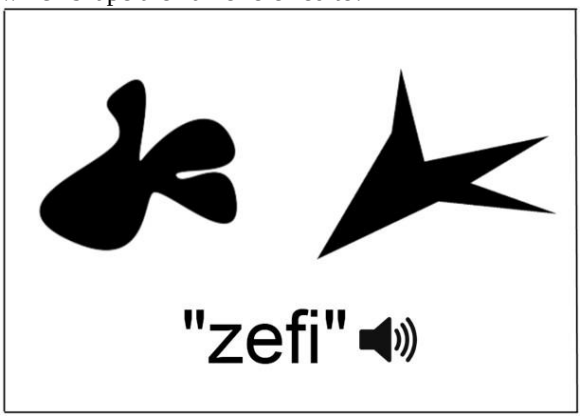

Figure 4 – A cross-situational learning trial (note that names are presented aurally).

Trials were grouped into four blocks per condition, each of 64 trials. Within each block each name appeared four times, and concomitantly 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 to appear for two trials in a row. Within these constraints, trials and trial order were randomised.

One of the two conditions was the 'round' condition. In this condition half of the shapes were round and half spiky (eight of each), 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 (different to the ones used in the round condition), but by contrast had eight neutral names and eight *spiky* names (again, neutral names were new). Shapes and neutral names were counterbalanced across conditions between participants. Condition order was also counterbalanced between participants.

At the outset of each participant's experiment, each shape was assigned a name from its condition (specific assignment was counterbalanced between participants). Half of the shapes in each category were assigned iconically *congruent* names. The other half of each category were assigned iconically *incongruent* names.

There were equal numbers of congruent and incongruent pairings for each category of shape in each condition. Here congruence is defined *within* whichever half of the putative round-spiky spectrum of sounds the condition in question covers. 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 (as there are no spiky names in this condition, and also in contrast to the incongruent round-name-spiky shape pairings) and neutral name-round shape pairings were considered incongruent (as they are less congruent than round-round pairings). The converse applied for the spiky condition (see figure 5)

	Round Condition		**Spiky Condition**	
	Round Name	*Neutral Name*	*Neutral Name*	*Spiky Name*
Round Shape	Congruent	Incongruent	Congruent	Incongruent
Spiky Shape	Incongruent	Congruent	Incongruent	Congruent

Figure 5 – Relative Congruence

If only round sound-shape associations matter, then iconic congruence will only be an advantage in the round condition.

Results

Repeated-measures ANOVAs were conducted with: 2 (conditions) × 4 (blocks) × 2 (congruent/incongruent pairing) × 2 (foil from same or different category) × 2 (target rounded or spiky); and with a dependent variable of log-transform of accuracy across the eight-trial cells defined by the above structure. Within-subjects variables were condition, block, congruence, foil from same vs. different category as target, and category of target. Participant was treated as a random effect. There was an additional nuisance between-subjects variable of condition order (round first vs. spiky first).

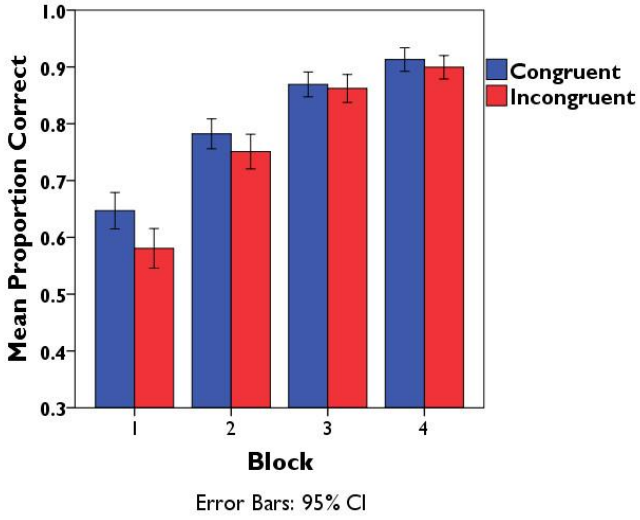

*Figure 6 – Results for the **round** condition in Experiment 2.*

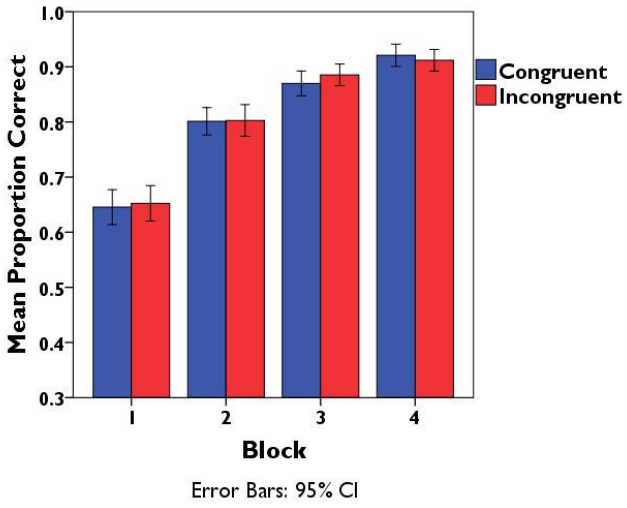

*Figure 7 – Results for the **spiky** condition in Experiment 2.*

As expected there was a main effect of block ($F(1.921,57.643) = 232.449$, $p < .001$; Greenhouse-Geisser corrected) indicating that participants learned over the course of the experiment. Furthermore, although there was no significant main effect of congruence, there was an interaction between condition and congruence ($F(1,30) = 4.893$, $p = .035$), indicating that congruence was an

advantage for the round but not the spiky condition (see Figures 6 and 7).

Discussion

These experiments explored the role lexical iconicity might play in diachronic language change, and the origins of a particular kind of iconicity, namely the so called 'kiki-bouba' effect.

In Experiment 1, an initially arbitrary 'language' developed iconicity for shape and duration of motion over the course of multiple generations. This is in line with the theory that because iconicity makes words easier to learn, remember, and process, iconicity is one pressure shaping long-term language change (among a number of others which include arbitrariness, see Perniss et al., 2010).

Intriguingly, both in Experiment 1 and 2 we found that sound-shape iconicity emerged for round but not spiky objects. This fact has not been widely reported before, possibly because the previous literature on the kiki-bouba effect has always relied on mutually-determining binary comparisons between pairs of words and shapes. Our experiments allowed us to independently assess iconic mappings to both round and spiky shapes. We found that iconic effects emerged in Experiment 1 only for bouba-round type mappings and that in Experiment 2 bouba-round associations show an iconic learning advantages in the absence of kiki-spiky pairings, but not vice versa.

This asymmetry is important with regards to the potential mechanisms underscoring these iconic effects. Ramachandran and Hubbard (2001) suggest that the kiki-bouba effect comes about either as cross-modal mappings that respect (non-visible) similarity between articulatory gestures and visual shapes, or that the effects may come about as cross-wiring between neural maps involved in audition and vision. Either way, however, an asymmetry is not predicted. Our speculative explanation for the effect is that it is driven by lip shape, perhaps implying a mechanism that maps similarity in visual shape between the lips of a speaker producing a bouba label and the corresponding shape. If this is the case, then this phenomenon does not require cross-modal analogies but comes about via similarities within the visual modality only, in line with substantial iconicity in sign languages. Future research will address the extent of the uni- vs. multimodality of the effect.

Acknowledgments

Supported by UK ESRC grants RES-062-23-2012 to Vigliocco & RES-620-28-6002 to the Deafness, Cognition and Language Research Centre (DCAL); UCL Impact; and the Spanish Ministry of Economy and Competitivity grant PSI2012-32464, to Santiago and Vigliocco. Paper written while Santiago was Leverhulme Visiting Professor (VP-1-2012-032) at UCL, hosted by Vigliocco.

References

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linell, K. J., & Spence, C. (2012). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition, 126*(2), 165-172.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31*(5), 489–558.

de Saussure, F. (1983). *Course in general linguistics*. La Salle, IL: Open Court.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*(1), 54-65.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America, 104*(12), 5241-5245.

Köhler, W. (1929). *Gestalt Psychology*. New York: Liveright.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition, 114*(1), 19-28.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science, 9*(3), 316-322.

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.

Ozturk, O., Krehm, Madelaine, & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology, 114*(2), 173-186.

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*(227). doi:10.3389/fpsyg.2010.00227

Perniss, P., & Vigliocco, G. (in press). The bridge of iconicity: From a world of experience to the experience of language. *Proceedings of the Royal Society B.*

Ramachandran, V. S. & Hubbard E. M. (2001). Synesthesia: a window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3-34.

Scott-Phillips, T., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Science, 14*(9), 411-417.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21*(1), 21-25.

Westbury, C. (2005). Implicit sound symbolism in lexical access: evidence from an interference task. *Brain and Language, 93*(1), 10-19.

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