# The Interactive Origin of Iconicity

## Mónica Tamariz,[a] Seán G. Roberts,[b,c] J. Isidro Martínez,[d] Julio Santiago[d]

[a]*Department of Psychology, Heriot-Watt University*
[b]*Department of Archaeology and Anthropology, University of Bristol*
[c]*Language and Cognition Department, Max Planck Institute for Psycholinguistics*
[d]*Deparment of Psychology, University of Granada*

## Abstract

We investigate the emergence of iconicity, specifically a bouba-kiki effect in miniature artificial languages under different functional constraints: when the languages are reproduced and when they are used communicatively. We ran transmission chains of (a) participant dyads who played an interactive communicative game and (b) individual participants who played a matched learning game. An analysis of the languages over six generations in an iterated learning experiment revealed that in the Communication condition, but not in the Reproduction condition, words for spiky shapes tend to be rated by naive judges as more spiky than the words for round shapes. This suggests that iconicity may not only be the outcome of innovations introduced by individuals, but, crucially, the result of interlocutor negotiation of new communicative conventions. We interpret our results as an illustration of cultural evolution by random mutation and selection (as opposed to by guided variation).

*Keywords:* Iconicity; Iterated learning; Cultural evolution

## 1. Introduction

The associations between linguistic signals and their meanings are largely arbitrary (De Saussure, 1983), but many languages contain iconic elements, in which aspects of signals resemble in some respect the structure of meanings (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016; Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Perniss & Vigliocco, 2014). Iconic mappings in natural languages include cases of *relative iconicity* in which an analogical contrast between meanings is related to a contrast between forms (Dingemanse et al., 2015) by, for example, activating multimodal associations or perceptual analogies (Kanero, Imai, Okuda, Okada, & Matsuda, 2014). Studies have shown that

---

Correspondence should be sent to Seán G. Roberts, Department of Archaeology and Anthropology, University of Bristol, 43 Woodland Road, BS8 1UU Bristol, UK. E-mail: sean.roberts@bristol.ac.uk

iconicity serves acquisition: Naive learners find it easier to remember iconic novel words (Lockwood, Dingemanse, & Hagoort, 2016), Sign Language signs (Vinson, Thompson, Skinner, & Vigliocco, 2015), and ideophones (Dingemanse, Schuerman, Reinisch, & Mitterer, 2016; Kantartzis, Imai, &, 2011), and iconic words are acquired early in life (Perry, Perlman, & Lupyan, 2015). Learning syntactic categories is also facilitated by sound correspondences (Monaghan, Christiansen, & Fitneva, 2011).

However, iconicity also plays a role in communication, during which it can help establish new conventional meaning-signal mappings. In experiments that explore the origin of communication systems, participants need to communicate with each other, but they do not have (or are experimentally barred from using) a common language (Galantucci, 2005; Garrod, Fay, Lee, Oberlander, & Macleod, 2007). When a communication system is established *de novo*, the initial signals produced tend to be motivated (iconic or indexical) signals which help disambiguate mappings between signals and possible meanings during communication. Motivated signals, therefore, help comprehension by establishing new shared conventions and common ground between the interlocutors (Perlman, Dale, & Lupyan, 2015). Other studies have shown that improvised graphical communication systems begin as sets of detailed and iconic signals, but become simpler and more *arbitrary* over episodes of interaction and feedback between a pair of interlocutors (Fay, Garrod, Roberts, & Swoboda, 2010; Garrod et al., 2007). However, this appears to conflict with the prevalence of iconicity in modern languages and evidence that iconicity both increases and decreases over historical time (Blasi et al., 2016).

A number of recent studies have modeled experimentally the emergence of fundamental properties of language as an interaction between cognitive skills and biases from communication on the one hand and sociocultural processes in transmission on the other (see Kirby, Cornish, & Smith, 2008; Galantucci, Garrod, & Roberts, 2012; Tamariz & Kirby, 2016; Tamariz, 2017, for reviews). In these experiments, participants are trained on miniature artificial languages or sets of signals, which they then reproduce and/or use communicatively; crucially, their output (usually different from the input) is used to train new participants, who themselves produce output which is given as input to a third "generation," and so on. This *iterated learning* design amplifies the effects of systematic biases that change the input, but it may be too subtle to be revealed in a single episode. Analyses of the resulting languages reveal the impact of the specific social dynamics on the emergence of linguistic properties. Iterated learning and reproduction of unstructured, holistic miniature artificial languages leads to *simplicity* (Kirby et al., 2008). If communication is added to the design by having two participants per generation who play a communicative game, a kind of systematicity called *compositionality*, a key property of language, emerges (Kirby, Tamariz, Cornish, & Smith, 2015).

Jones et al. (2014) found that iconic signals emerge in an iterated learning experiment. Participants had to learn and reproduce signals associated with objects that varied in color, movement, and shape (round or spiky). Over generations of transmission, signals for round shapes became rated as sounding more round than spiky shapes, mirroring the so-called "bouba-kiki" effect (Köhler, 1929; Ramachandran & Hubbard, 2001). However, this experiment only tested the contribution of transmission, not its interaction with communication.

If iconic signals are beneficial to communication, they should also appear during the negotiation of conventions between interlocutors. Then, because they are easy to learn and reproduce, iconic signals should be retained in languages over cross-generational transmission, leading to an increase in iconicity. In this study we explore the emergence of iconicity under pressures from learning, communication, and transmission, focusing on communicative interaction versus individual reproduction. In our iterated learning experiment, initially random, noniconic miniature artificial languages are either (a) learned and reproduced or (b) learned and used communicatively. We measure the relative iconicity of the languages over the generations.

## 2. Materials and methods

### 2.1. Participants

A total of 93 native speakers of Spanish were recruited from undergraduate and postgraduate psychology courses at the University of Granada in exchange for course credit. (The participant detail sheets containing the ages and genders were lost while moving buildings.) Data from 19 original participants were excluded and collected again using new participants, three due to software failure, and the rest because of noncompliance with the instructions which indicated that the use of words that sounded like Spanish were not permitted: In the Communication condition, one of the chains and the last two generations of another one contained mainly words that were readily identifiable as the name of colors in Spanish, for example, *azjulll*, *ajhul* for "azul" (blue); *bojo*, *roj* for "rojo" (red); or *veeeejheeee*, *vehe* for "verde" (green).

### 2.2. Initial languages

All the initial languages used as input to the first generation consisted of the same 12 typed signals (Fig. 1), each associated with a drawing of an object (the associations varied between chains). The set of 12 *signals* contained a diverse set of characters with approximately equal frequency for all vowels and all consonants, and were constructed to be as invariant as possible in terms of their perceived spikiness (see Appendix S1 for details). The 12 *objects* are all possible combinations of three features: two shapes (rounded, spiky), three colors (red, green, blue), and two borders (no border, border). The shapes and colors are identical to those used in Jones et al. (2014).

We generated 30 random mappings between the 12 signals and the 12 objects, from which we selected eight mappings to be the input for the first generation in each of our eight chains. These mappings had nonsignificant systematicity levels (Mantel test: $-0.70 <$ z-score $< 0.70$; 2-tailed $p > .24$) and iconicity values ($t$-tests comparing the spikiness values given in a norming study [see Appendix S1 and S2] to spiky vs. round objects returned $t < 1.62$, $p > .17$). (See Table 1 for details of metrics.)
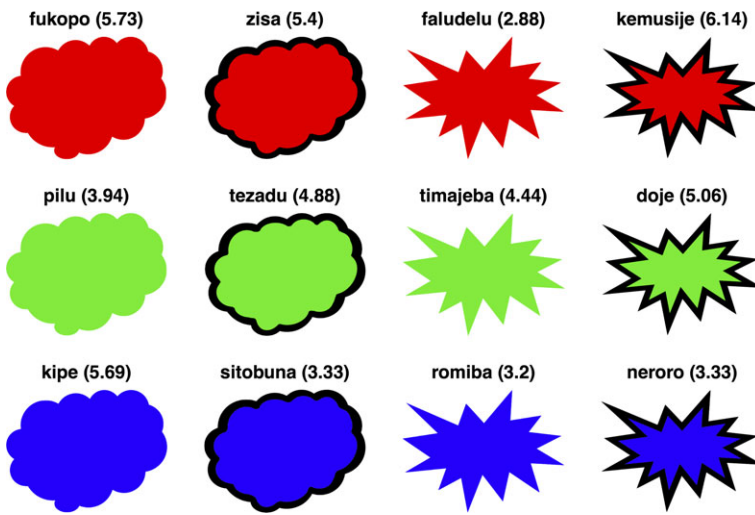
Figure 1. Example initial input for a chain, consisting of 12 objects each with an associated signal. Numbers in brackets show the mean spikiness ratings in a scale from 1 to 7.

Table 1
Definitions of concepts used in the study and how they were measured

| Concept | Definition | Measured by | Example |
|---|---|---|---|
| Absolute iconicity | A form directly resembles its meaning | Not measured in this study | Onomatopoeia (e.g. "bang") |
| Relative iconicity | An analogical contrast between meanings is related to a contrast between forms | Spikiness rating: Naive judges rated the spikiness/ roundness of each word on a Likert scale | "kiki" sounds more spiky than "bouba" |
| Systematicity | A systematic mapping between substrings of the signal and aspects of meaning | Mantel z-score of the correlation between differences in meaning and differences in form | The prefix pa- is used for all red objects |
| Innovation | A new or mutated signal | A signal that differs from its transmitted ancestor in form or mapping to its referent | Participant is taught "pilu" for a green object, but produces "pilo" |
| Increase in iconicity | The extent to which an innovation enhances the relative iconicity | Difference in the estimated spikiness ratings of the innovation and its ancestor | Participant is taught "bibi" for a spiky object, but produces "kiki" |
| Transmission error | Participants introduce innovations into the language that is transmitted | Average normalized Levenshtein distance between the forms in the input language and forms in the output language | |
| Task success | Interlocutors align on a referent | Whether the guesser chooses the correct target meaning | |

## 2.3 Procedure

We used a naming game inserted in an iterated learning paradigm adapted from Kirby et al. (2015). In the following paragraphs we describe the game in the "Communication" and the individual "Reproduction" conditions (Fig. 2).
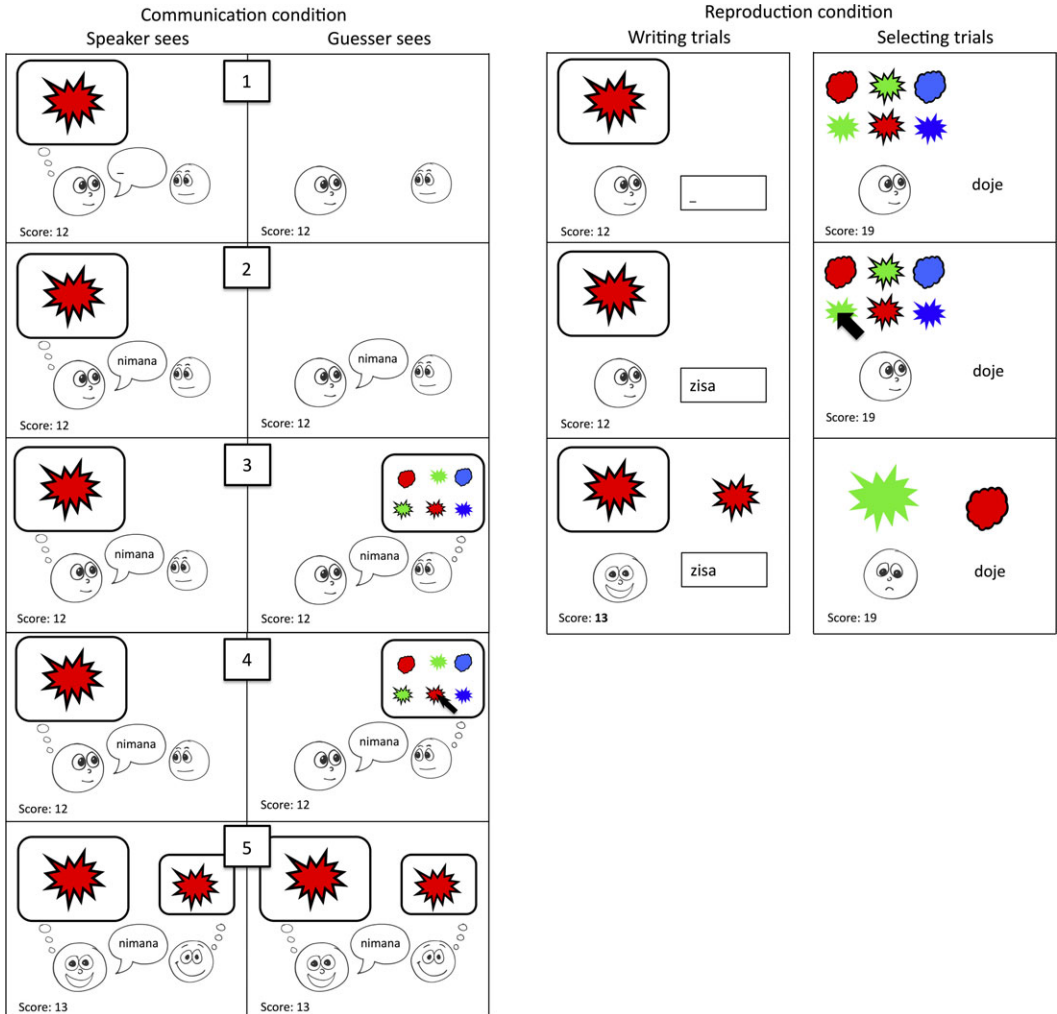


Figure 2. Illustrations of the game. On the left, five screen views from the Communication condition from the point of view of the speaker (left column) and the guesser (right column). The speaker is given a target object (1) and then he types a signal for it (2). The guesser sees the typed signal (2) and then is presented with an array of six objects (3), from which he selects the object he thinks corresponds to the signal (4). Finally, the target and guess are revealed to both players. In this case, the guess is correct, so the score is increased by 1 (5). On the right, six screen views from the Reproduction condition. In a Writing trial, the participant sees a target object, has to type the signal for it, and then gets feedback (see text). In a Selecting trial, she is shown an array of six objects, as well as one signal, and has to select the object associated with the signal. She then gets feedback.

In the Communication condition, two participants sit at separate networked computers and are told they will play with each other. Both are trained with the same initial language: They see each item in turn (the object for 1 s, then object plus its signal for 5 s, and followed by 3 s of blank screen), six times in different random orders. After training, they play a naming game (Fig. 2) using the language. In each trial of the game, a speaker names an object for his partner, the guesser, who then tries to pick the correct target object out of a set of six. Finally, feedback is given: Both players see the other's response. If the target and selected objects are the same, the two characters on the screen smile and the score is increased by 1. If they are different, they make a sad face and the score does not change. The experiment then moves on to the next trial, with the participants switching their roles as speaker and guesser. The playing stage includes two blocks. In each block all the items are presented twice in random order to each player, once as a guesser and once as a speaker.

In the Reproduction condition, each participant plays individually, not as part of a pair. The procedure is identical to the Communication condition, except for the following aspects: The on-screen presentation does not show a partner (Fig. 2). The participant is told she will play a memory game: First, she will be trained on a language, and then she will be tested on it. The participant alternates between writing and selecting trials where she (a) sees an object and writes the corresponding signal and (b) sees signal and selects the corresponding object. Feedback appears on the screen 1 s after she submits her answer. When the participant is given a signal and has to select its associated object from an array, the feedback is the correct word in the training language. When she is given an object and has to type its associated signal, then (a) if the participant types a word present in the training language, then the feedback is the corresponding object (or a randomly selected object if several objects had the same word in the training language) or (b) if she types a novel word, the feedback is the object whose word in the training language has the shortest edit distance (normalized Levenshtein distance) to the signal she typed. At every test trial, the score is increased by 1 point if the target and feedback are the same. The language used to train the next participant in the chain is formed by the 12 objects, each associated with the last word that the participant typed for it during testing. In the Communication condition, all signals come from one of the two participants, selected at random (as in Kirby et al., 2015). We ran four transmission chains of six generations in each condition.

Participants were asked to come in pairs for the Communication condition or independently for the Reproduction condition. They sat at a computer each, in separate laboratory cubicles in the case of pairs. They then read the instructions which explained the game (see Appendix S3), emphasized the joint goal of scoring as many points as possible, and asked them not to use recognizable words. The experiment lasted for about 45 min. After all chains were run, 16 naive judges rated each word in the initial languages and in the transmitted languages for spikiness/roundness on a 7-point Likert scale (see Appendix S1).

## 3. Results

We set out to examine the effect of communicative interaction on the emergence of iconicity in artificial miniature languages. This section reports the results relating to the emergence of iconicity, explores the possible functionality of iconicity for communication, and looks at possible cultural evolutionary mechanisms for its emergence.

### 3.1. The emergence of iconicity

Our *iconicity* metric for each word was based on naive judges' spikiness–roundness ratings (see Appendix S1). We performed a linear mixed-effect model analysis[1] predicting word spikiness ratings according to the following fixed effects: Condition (Communication, Reproduction), Shape of the object (Spiky, Round), and Generation (1–6). We included random intercepts for Chain (0–7) and Item (1–12); random effects for participant were negligible. The spikiness ratings were very bimodal, so were transformed into binary values (split halfway along the Likert scale), and a binomial model was used. Estimates of significance were obtained through model comparison (log likelihood ratio test).

There was a significant main effect of shape (*log likelihood difference* = 2.4, *df* = 1, $\chi^2$ = 4.85, *p* = .028) and a significant interaction between shape and condition (*log likelihood difference* = 6, *df* = 1, $\chi^2$ = 12.04, *p* = .00052). However, the effect sizes for these variables were small (shape: $\beta$ = 0.52, *SE* = 0.6, *t* = 0.87; shape × condition: $\beta$ = −0.061, *SE* = 0.83, *t* = −0.074). There was a marginal three-way interaction among shape, condition, and generation (*log likelihood difference* = 1.6, *df* = 1, $\chi^2$ = 3.2, *p* = .073), and the effect size for this was larger ($\beta$ = 0.39, *SE* = 0.22, *t* = 1.8). This suggests that, starting with similar spikiness ratings at generation 0,[2] in the Communication condition the spikiness ratings for spiky and round shapes diverge over generations (spiky shapes become more spiky, round shapes become less spiky), whereas in the Reproduction condition they do not diverge (Figs. 3 and 4, see Appendix S5).

To check the robustness of the result to the assumptions of the analysis, we ran a series of alternative analyses (see Appendix S4). Qualitatively similar results were obtained with a mixed effects model using the continuous spikiness ratings, an ANOVA, and a binary regression tree analysis. However, a Monte Carlo permutation test, which deals well with bimodal data but ignored aspects of meaning other than spikiness, did not find significant differences between the conditions.

We calculated the extent to which each output language encoded each aspect of meaning. For each dimension of shape, color, and border, we calculated the systematicity (see Appendix S4). Given that the question of interest in this study was iconicity with respect to shape, we measured the systematic structure of the languages with respect to each of our three meaning dimensions. In other words, we asked whether objects that are similar in shape (or color, or border) are associated with words that are similar. A linear mixed-effect model analysis predicting the systematicity of languages with Condition (Communication or Reproduction), Generation (0–6), and Meaning (Shape, Colour, Border) as
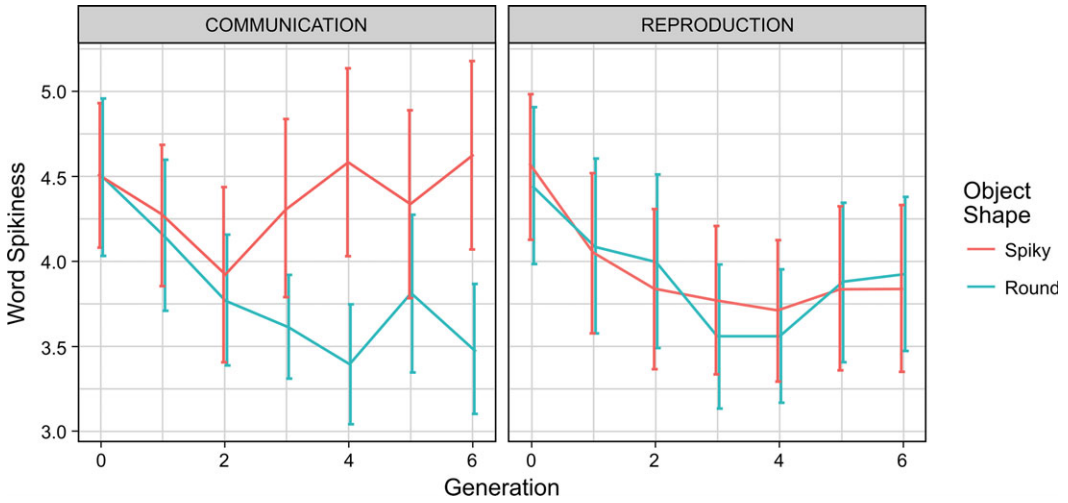
Figure 3. Spikiness values of the signals used to name the objects of the two different shapes, at the different generations, in the two conditions. Values are mean ratings on a 7-point Likert scale with 95% CI.
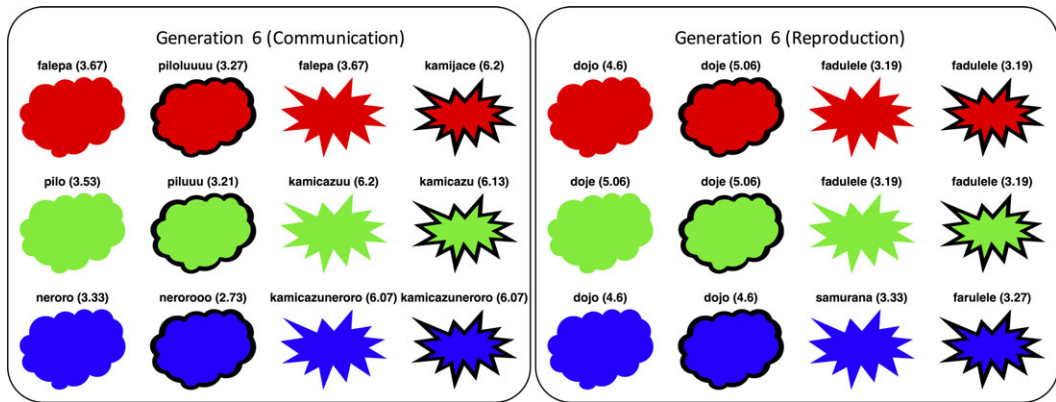


Figure 4. Examples of output languages from the sixth generation for a chain in the Communication condition (left) and the Reproduction condition (right). Numbers in brackets show the mean spikiness ratings.

fixed effects and Chain (0–7) as a random effect revealed a significant effect of Generation (*log likelihood difference* = 5.09, *df* = 1, $\chi^2$ = 10.2, *p* = .0014) and Meaning (*log likelihood difference* = 5.69, *df* = 2, $\chi^2$ = 11.4, *p* = .003), and significant interactions of Meaning × Condition (*log likelihood difference* = 7.75, *df* = 2, $\chi^2$ = 15.5, *p* = .0004) and Generation × Meaning x Condition (*log likelihood difference* = 13.75, *df* = 7, $\chi^2$ = 27.5, *p* = .0003) (Fig. 5, see also Appendix S5). The most interesting result here is the effect of Meaning: Systematic encoding of shape in the language had the highest average z-score (1.27), followed by color (0.91), whereas border had the lowest (0.18). This indicates that shape seems to be a salient object feature. In addition, the
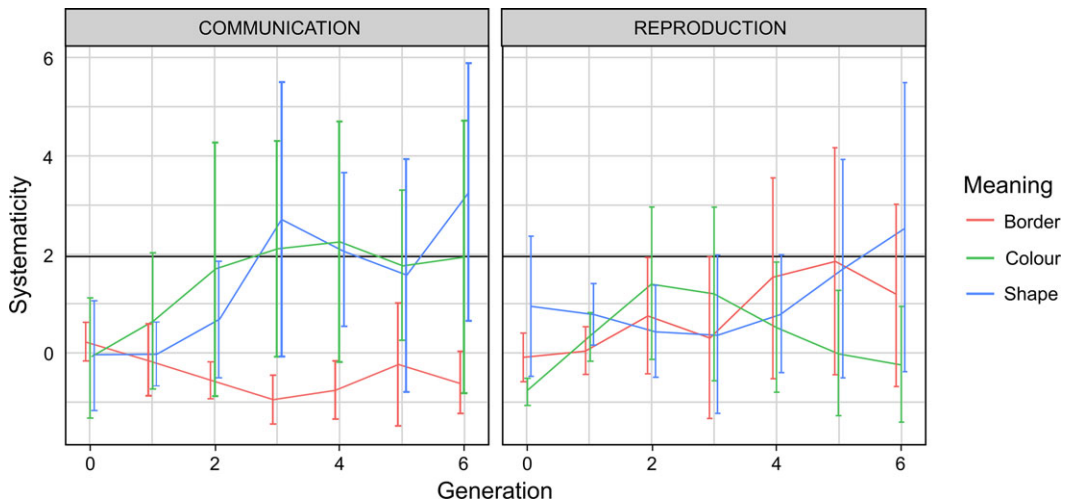
Figure 5. Systematicity with respect to shape, color, and border, over generations and in the Communication and Reproduction conditions. Z-score values larger than 1.96 are significant ($p < .05$).

systematicity of the language is driven by differences in word forms correlating with differences in shape, our feature of interest (and also with color); this is true to a significantly higher extent in the Communication than in the Reproduction condition, another indication that the iconicity effect is favored by communicative interaction.

Finally, as an additional test of iconicity, we obtained a spikiness coefficient for each letter based on its frequency in the words produced in the game, namely, the frequency of the letter in words for spiky objects over its frequency in words for round objects. We calculated its correlation with the spikiness values given by judges directly to the individual letters in the norming study. We found a marginally significant correlation when all words were taken into account ($r = .414$, $n = 22$; $p < .1$). When we considered the two conditions separately, however, we found a significant correlation in the Communication condition ($r = .525$, $n = 21$; $p = 0.017$) but not in the Reproduction condition ($r = .195$, $n = 20$; $p = .395$). This further indicates that the signals given to spiky objects tend to be more spiky, but only when there is communicative interaction.

The languages also increased in systematic structure over generations, the transmission error decreased, and the task success increased, consistent with previous results (Kirby et al., 2008, 2015), and there were no differences between conditions (see Appendix S2). However, we did find a difference in how expressivity changed: Participants in the Reproduction condition cumulatively introduced more homonyms than in the Communication condition (see Appendix S2).

## 3.2. The functionality of iconicity for communication

In an effort to understand why iconicity emerges in the Communication condition, we examined how the iconicity of innovations affected learning the mappings between

signals and meanings. An innovation is introduced in the language when the signal produced contains differences with respect to the corresponding signal in the input language, or when it is associated with a different object. Innovations can be either more or less iconic than the signal or mapping they replace. A total of 689 innovations were produced by pairs in the Communication condition and 282 by single participants in the Reproduction condition. About three quarters of innovations in both conditions were changes in mapping rather than previously unattested words (see Appendix S5). Human ratings for all innovations were not available, so we estimated the spikiness by extrapolating from the human ratings based on unigrams and bigrams using a random forests model (see Appendix S1 and S5). The difference in systematicity was calculated as the systematicity of the language before the innovation and the systematicity after the innovation. Note that, in general, systematicity and iconicity are independent. If words are holistically iconic, conveying multiple aspects of meaning at once, then iconicity does not serve systematicity. However, if an iconic contrast picks out just one aspect of meaning and is expressed in a substring of the meaning, then this will also increase systematicity. That is, in contrast to an arbitrary systematic language, the forms in an iconic systematic language are motivated.

We found that the iconicity of an innovation did not help guessers guess the exact object correctly during the communicative naming games (Fig. 6, left; log likelihood difference = 0.00018, $df = 1$, $\chi^2 = 0$, $p = .98$, see Appendix S5), but it did help guess correctly the shape of the object (Fig. 6, right, *log likelihood difference* = 5, *df* = 1, $\chi^2 = 10.0$, $p = .0016$, see Appendix S5). In other words, guessers are more likely to select a spiky object when presented with a signal that sounds more spiky, and a round object when presented with a signal that sounds round.
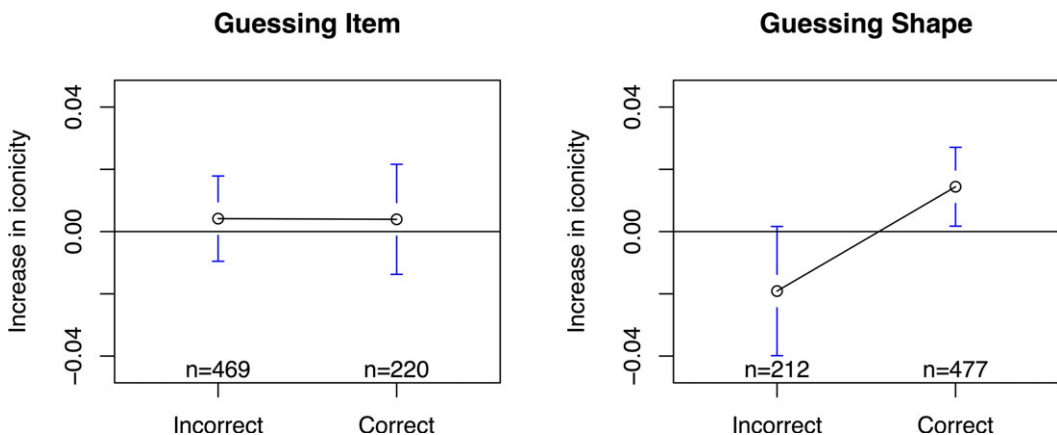


Figure 6. The increase in iconicity of innovations that led the guesser to guess correctly or incorrectly in trials of the guessing game the exact object (left) or the shape of the object (right). These data are only for the communication condition where both speaker and listener produce and receive innovations. Points are means with 95% confidence intervals.

We also looked at the relationship between whether an innovation increases systematicity and whether it increases iconicity. When predicting the increase in iconicity, there is a significant interaction between the increase in systematicity and condition (*log likelihood difference* = 2.4, *df* = 1, $\chi^2$ = 4.77, *p* = .029, see Appendix S5). In the Communication condition, innovations which increase iconicity also tend to increase systematicity, and decreases in iconicity tend to decrease systematicity, whereas in the Reproduction condition most innovations increase systematicity regardless of the change in iconicity. Although the effect size is small (correlation between systematicity and iconicity changes in Communication condition: *r* = .096), it provides another indication that both iconicity and systematicity are selected in the Communication condition.

## 3.3. Cultural-evolutionary mechanisms

In the Communication condition there are two possible (not mutually exclusive) processes that may increase iconicity: First, when speakers *introduce an innovation*, they may be biased toward increasing the iconicity of the mappings, either to facilitate processing of the language or to facilitate comprehension by the hearer. Second, when the interlocutor decides whether to *adopt an innovation*, there may be a bias to adopt and reproduce preferentially the innovations that increase iconicity. In contrast, in our Reproduction task, participants may introduce and subsequently reproduce their own innovations, but adoption of an interlocutor's innovation is not a possible mechanism.

We estimated whether innovations increased or decreased the iconicity and systematicity compared with the words they replaced. Fig. 7 shows the distribution of changes in
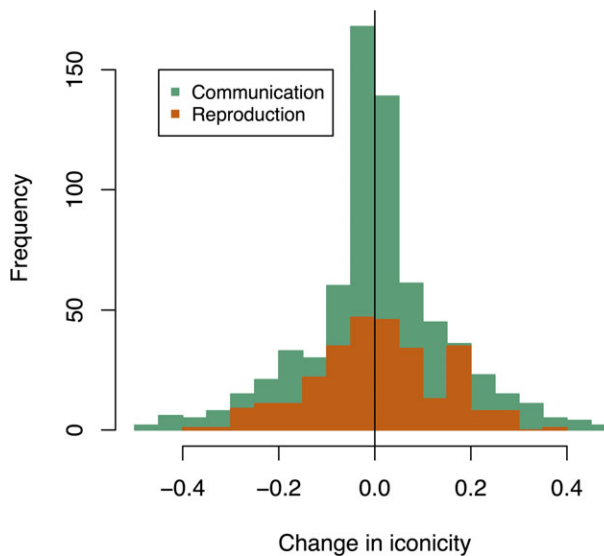


Figure 7. The distributions of change in iconicity for innovations in the two conditions. The data for the reproduction condition include only the innovations produced by the human participant.
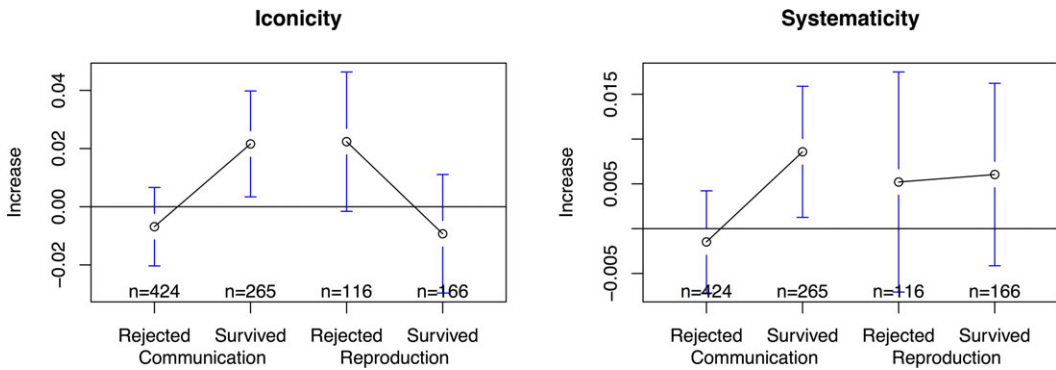
Figure 8. Increases in iconicity and systematicity for innovations in each condition, split by whether the innovation survived to be transmitted to the next generation. Points are means, with 95% confidence intervals.

iconicity. For both conditions, the means are not significantly different from zero (Wilcoxon signed rank test, Communication condition: $V = 117,360$, $p = .36$; Reproduction condition: $V = 20,314$, $p = .64$) and the distributions are not significantly asymmetric (MGG test of symmetry, see Miao, Gel, & Gastwirth, 2006, Communication condition: $p = .35$, Reproduction condition: $p = 0.67$, see Appendix S5). That is, speakers are equally likely to produce innovations which increase or decrease iconicity.

Fig. 8 shows the increase in iconicity and systematicity for innovations which survived to be transmitted to the next generation, and innovations which did not survive. Survival relies on adoption and repetition. For the increase in iconicity, there is a significant interaction between survival and condition ($\beta = -0.06$, $SE = 0.02$, *Wald t* $= -3.0$; *log likelihood difference* $= 4.4$, $df = 1$, $\chi^2 = 8.72$, $p = .003$). In the Communication condition, innovations that survive tend to increase both iconicity and systematicity. However, in the Reproduction condition, the innovations only contribute to systematicity but not to iconicity.

## 4. Discussion

Our results show that relative iconicity (similar to the bouba-kiki effect) emerges when initially noniconic languages are used communicatively and then transmitted to a new generation, but not when they are reproduced and then transmitted (Fig. 3). This suggests that something in the communicative interaction drives iconicity. The main difference between conditions lies in the opportunity for accommodation between interlocutors. Whereas in the Reproduction condition a single participant could innovate but received exclusively corrective feedback, in the Communication condition two interacting participants could both innovate and adopt each other's innovations. Evolution in the latter condition is therefore driven by innovations in the signals (which can be thought of as *mutations*), produced by participants which may then be transmitted to the following

generation, and by the adoption of innovations by the interlocutor. Depending on the point at which the mutation is produced, there are up to three chances to reproduce that mutation. This reproduction can be neutral, if all mutations are equally likely to be reproduced, or biased, if some kind of mutations are more likely to be reproduced than others. In the latter case we can talk of *selection*, which results, over generations, in a higher prevalence of the kind of mutations that are selected for. Our results show that the mutations produced are equally likely to increase or decrease the iconicity of the languages (Fig. 7). The results also show that similar numbers of innovations are produced in the Communication and the Reproduction conditions, but the transmission of these innovations varies across conditions. In Communication there is a preference for mutations that increase both iconicity and systematicity (Fig. 8). In Reproduction, in contrast, mutations that increase systematicity are selected for, but mutations that increase iconicity are not. Taken together, these results can be interpreted as evidence in favor of cultural evolutionary dynamics being driven by random mutation and selection, rather than by guided variation (Richerson & Boyd, 2005). The latter would involve an increased chance of innovations that increased iconicity, which is not attested (Fig. 7); the former requires unbiased production, shown in Fig. 7, and biased adoption, shown in Fig. 8.

The emergence of iconicity through random mutation and selection is in line with some other findings. For example, Verhoef, Roberts, and Dingemanse (2015) ran a communication game with only four meanings, two of which had obvious iconic mappings with the signaling medium. Iconic mappings emerged, but not always in the first generation. Blasi et al. (2016) find consistent sound symbolic patterns in the world's languages, but they do not align with language family structures, suggesting that the patterns emerge, are lost, and reemerge many times throughout history, rather than being conserved through time. This could happen if random mutations perturb the existing patterns, before selection brings them back again.

The cause of the preferential adoption of innovations that increase iconicity by interlocutors during communicative interaction could be (a) iconic mappings favor processing and learning, or (b) interlocutors "think" iconic mappings will be better for aligning/negotiating a new convention. Our result from the Reproduction condition seems to go against (a) so it will be interesting to look for the cause of our increased effect in the communicative condition in the benefits of iconicity for alignment. This connection between iconicity and the initial negotiation of new conventions is already apparent in studies such as Galantucci (2005), Scott-Phillips, Kirby, and Ritchie (2009), Garrod et al. (2007), and Fay et al. (2010). It is interesting to note that these four studies actually find a decrease in iconicity over rounds of communicative use, which contrasts with the increase in iconicity observed in our study. The nature of the tasks may be behind this discrepancy: the former starts with signals improvised by participants, whereas the latter starts by training participants on a language designed by the experimenters.

The absence of iconicity in the Reproduction condition conflicts with Jones et al.'s (2014) results. Our study is very similar, but there are differences in the proportion of the language exposed in training (50% vs. 100% in our study); the number of generations (10 vs. 6); the meaning space (motion instead of border); iconicity metric (estimates from

individual letters vs. direct ratings of whole words); and feedback (absent vs. present). The absence of feedback and the 50% bottleneck both disrupt transmission; this should increase the pressure for compressibility (Kirby et al., 2015). If our experiment had run for longer with a stronger pressure for compressibility, then perhaps iconicity would also have emerged in the Reproduction condition, implying that communication simply speeds up the process.

In experimental graphical communication systems, as well as in writing systems, initially complex, iconic signals tend to become reduced and more efficient to produce. In experiments, this happens both over episodes of communication (Garrod et al., 2007) and transmission (Caldwell & Smith, 2012). The signals become simpler and less transparent, and therefore more arbitrary for new learners. In our Communication condition, however, iconicity emerges and persists over communication and transmission. Why do we see a divergence from previous studies? Graphical iconicity is costly in terms of time and effort —the details need to be drawn accurately when the conventions are being established, but once the conventions are entrenched, simplified forms work just as well. By contrast, the relative iconicity that emerges in our experiment is not costly. The iconic signals that reflect the shape of the objects are not longer or more complex than the initial random signals. This kind of iconicity therefore persists over generations because it favors learning and it is not threatened by a reduction process in response to efficiency pressures.

In conclusion, in our iterated learning chains of orthographic languages referring to objects with a spikiness/roundness contrast, relative iconicity emerges when the languages are used communicatively, which suggests communication may be one of the mechanisms that explain the presence of iconicity across the world's languages.

## Acknowledgments

## Notes

1. We used R (R Core Team, 2016) and lme4 (Bates, Maechler, Bolker, & Walker, 2015) to perform linear mixed effects analyses. Unless otherwise stated, estimates of significance were obtained through model comparison (log likelihood ratio test).
2. The average spikiness value of the words at generation 0 was 4.5, above the middle of the 7-point Likert scale used to rate them. The same words had averaged a spikiness value of 5.1, only slightly below the middle of the 10-point

Likert scale during the norming study used to select the initial languages (see Section 2.2 Initial languages, and Appendix S1). Differences between the two ratings include the different Likert scales (10 vs. 7 points) as well as the fact that the same 12 words were accompanied by different words in each case: In the first case, by other five sets of 12 words with homogeneous character frequencies, and in the second case, the words transmitted to the other six generations in the experimental chains.

# References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, *113*(39), 10818–10823. doi:10.1073/pnas.1605782113

Caldwell, C. A., & Smith, K. (2012). Cultural evolution and the perpetuation of arbitrary communicative conventions in experimental microsocieties. *PLoS ONE*, *7*, e43807.

De Saussure, F. (1983). *Course in general linguistics*. LaSalle: Open Court.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615.

Dingemanse, M., Schuerman, W., Reinisch, E., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, *92*(2), e117–e133.

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, *34*, 351–386.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, *29*, 737–767.

Galantucci, B., Garrod, S., & Roberts, G. (2012). Experimental semiotics. Language and linguistics. *Compass*, *6*(8), 477–493.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & Macleod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, *31*, 961–987.

Jones, M., Vinson, D., Clostre, N., Zhu, A. L., Santiago, J., & Vigliocco, G. (2014). The bouba effect: Sound-shape iconicity in iterated and implicit learning. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 1114–1119). Austin, TX: Cognitive Science Society.

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*, e97905.

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

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the lab: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication drive the evolution of language. *Cognition*, *141*, 87–102.

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

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

Miao, W., Gel, Y. R., & Gastwirth, J. L. (2006). A new test of symmetry about an unknown median. In A. Hsiung, C.-H. Zhang, & Z. Ying (Eds.), *Random walk, sequential analysis and related topics — A festschrift in honor of yuan-shih chow*. Singapore: World Scientific Publisher.

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology. General*, *140*, 325–347.

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, *2*(8), 150152. doi:10.1098/rsos.150152

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Proceedings of the Royal Society B*, *369*, 20130300. doi:10.1098/rstb.2013.0300

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE*, *10*(9), e0137147.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. (2009). Signalling signalhood and the emergence of communication. *Cognition*, *113*(2), 226–233.

Tamariz, M. (2017). Experimental studies on the cultural evolution of language. *Annual Review of Linguistics*, *3*, 389–407.

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, *8*, 37–43.

Verhoef, T., Roberts, S. G., & Dingemanse, M. (2015). Emergence of systematic iconicity: Transmission, interaction and analogy. In D. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 2481–2486). Austin, Tx: Cognitive Science Society.

Vinson, D., Thompson, R. L., Skinner, R., & Vigliocco, G. (2015). A faster path between meaning and form? Iconicity facilitates sign recognition and production in British Sign Language. *Journal of Memory and Language*, *82*, 56–85.

---

**Supporting Information**

Additional Supporting Information may be found online in the supporting information tab for this article:

**Appendix S1.** Measuring iconicity.

**Appendix S2.** Systematicity, error, expressivity, and task success.

**Appendix S3.** Task instructions given to participants.

**Appendix S4.** Extra tests of iconicity: ANOVA, Binary Regression Tree analysis, and Monte Carlo analysis.

**Appendix S5.** R code for statistical tests.