

Repeated imitation makes human vocalizations more word-like

Edmiston, Pierce; Perlman, Marcus; Lupyan, Gary

DOI:

[10.1098/rspb.2017.2709](https://doi.org/10.1098/rspb.2017.2709)

License:

None: All rights reserved

Document Version

Peer reviewed version

Citation for published version (Harvard):

Edmiston, P, Perlman, M & Lupyan, G 2018, 'Repeated imitation makes human vocalizations more word-like', *Royal Society of London. Proceedings B. Biological Sciences*, vol. 285, no. 1874.
<https://doi.org/10.1098/rspb.2017.2709>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

Final Version of Record available at: <http://dx.doi.org/10.1098/rspb.2017.2709>

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

1 Repeated imitation makes human vocalizations more word-like

2 Pierce Edmiston¹, Marcus Perlman², & Gary Lupyan¹

3 ¹ University of Wisconsin-Madison

4 ² University of Birmingham

5 Author Note

6 Pierce Edmiston and Gary Lupyan, Department of Psychology, University of
7 Wisconsin-Madison, Madison, Wisconsin. Marcus Perlman, Department of English Language
8 and Applied Linguistics, University of Birmingham, United Kingdom.

9 Correspondence concerning this article should be addressed to Pierce Edmiston, 1202
10 W. Johnson St., Madison, WI, 53703. E-mail: pedmiston@wisc.edu

11

Abstract

12 People have long pondered the evolution of language and the origin of words. Here, we
13 investigate how conventional spoken words might emerge from imitations of environmental
14 sounds. Does the repeated imitation of an environmental sound gradually give rise to more
15 word-like forms? In what ways do these forms resemble the original sounds that motivated
16 them (i.e., exhibit iconicity)? Participants played a version of the children’s game
17 “Telephone”. The first generation of participants imitated recognizable environmental sounds
18 (e.g., glass breaking, water splashing). Subsequent generations imitated the previous
19 generation of imitations for a maximum of 8 generations. The results showed that the
20 imitations became more stable and word-like, and later imitations were easier to learn as
21 category labels. At the same time, even after 8 generations, both spoken imitations and their
22 written transcriptions could be matched above chance to the category of environmental
23 sound that motivated them. These results show how repeated imitation can create
24 progressively more word-like forms while continuing to retain a resemblance to the original
25 sound that motivated them, and speak to the possible role of human vocal imitation in
26 explaining the origins of at least some spoken words.

27

Keywords: language evolution, iconicity, vocal imitation, transmission chain

28

Word count: 6964

29 Repeated imitation makes human vocalizations more word-like

30 Most vocal communication of non-human primates is based on species-typical calls that
31 are highly similar across generations and between populations [1]. In contrast, human
32 languages comprise a vast repertoire of learned meaningful elements (words and other
33 morphemes) which can number in the tens of thousands or more [2]. Aside from their
34 number, the words of different natural languages are characterized by their extreme diversity
35 [3,4]. The words used within a speech community change relatively quickly over generations
36 compared to the evolution of vocal signals [5]. At least in part as a consequence of this rapid
37 change, most words appear to bear a largely arbitrary relationship between their form and
38 their meaning — seemingly, a product of their idiosyncratic etymological histories [6,7]. The
39 apparently arbitrary nature of spoken vocabularies presents a quandary for the study of
40 language origins. If words of spoken languages are truly arbitrary, by what process were the
41 first words ever coined?

42 While the origin of most spoken words remains opaque, the situation is somewhat
43 different for signed languages for which much is known regarding the origins of many signs.
44 Although signed languages rely on the same type of referential symbolism as spoken
45 languages, many individual signs have clear iconic roots, formed from gestures that resemble
46 their meaning [8–10]. For instance, [11] noted the iconic origins of the American Sign
47 Language (ASL) sign for “bird”, which is formed with a beak-like handshape articulated in
48 front of the nose. Another example is “steal”, derived from a grabbing motion to represent
49 the act of stealing something. [12] identified about 25% of ASL signs to be iconic, and
50 reviewing the remaining 75% of ASL signs, [13] determined that about two-thirds of these
51 seemed plausibly derived from iconic origins. Further support for iconic origins of signed
52 languages comes from observations of deaf children raised without exposure to a signed
53 language, who develop homesign systems to use with their family. In these communication
54 systems, children frequently use pantomimes and various iconic and indexical gestures some
55 of which may become conventionalized [14]. Participants in laboratory experiments utilize a

56 similar strategy when they cannot rely on existing words [15].

57 In contrast to the visual gestures of signed languages, many have argued that iconic
58 vocalizations could not have played a significant role in the origin of spoken words because
59 the vocal modality simply does not afford much form-meaning iconicity [16–21]. It has also
60 been argued that the human capacity for vocal imitation is a domain-specific skill, geared
61 towards learning to speak, rather than the representation of environmental sounds. For
62 example, [22] suggested that, “most humans lack the ability... to convincingly reproduce
63 environmental sounds... Thus ‘capacity for vocal imitation’ in humans might be better
64 described as a capacity to learn to produce speech” (p. 209). Consequently, it is still widely
65 assumed that vocal imitation — or more broadly, the use of any sort of resemblance between
66 form and meaning — cannot be important to understanding the origin of spoken words.

67 Although most words of contemporary spoken languages are not clearly imitative in
68 origin, there has been a growing recognition of the importance of iconicity in spoken
69 languages [23,24] and the common use of vocal imitation and depiction in spoken discourse
70 [25,26]. This has led some to argue for the importance of imitation for understanding the
71 origin of spoken words [27–31]. In addition, counter to previous assumptions, people are
72 highly effective at using vocal imitations to refer to events such as coins dropping in a jar or
73 environmental sounds like scraping — even more effective in some cases than when using
74 conventional words [32]. These imitations are effective not because people can mimic
75 environmental sounds with high fidelity, but because people can capture with their
76 “imitations” salient features of the referent in ways that are understandable to listeners [33].
77 Similarly, the features of onomatopoeic words might highlight distinctive aspects of the
78 sounds they represent. For example, the initial voiced, plosive /b/ in “boom” represents an
79 abrupt, loud onset, the back vowel /u/ a low pitch, and the nasalized /m/ a slow, muffled
80 decay [34]. Such iconicity is not limited to imitations of sounds. People are able to create
81 novel imitative vocalizations for more abstract meanings (e.g. “slow”, “rough”, “good”,
82 “many”) such that the vocalizations are understandable to naïve listeners [31].

83 Thus, converging evidence suggests that people can use vocal imitation as an effective
84 means of communication. At the same time, vocal imitations are not words. If vocal
85 imitation played a role in the origin of some spoken words, then it is necessary to identify
86 circumstances in which vocal imitation may give rise to more word-like vocalizations that
87 can eventually be integrated into a vocabulary of a language. In the present set of studies we
88 ask whether vocal imitations can transition to more word-like forms through sheer repetition
89 — without an explicit intent to communicate. To answer this question, we recruited
90 participants to play an online version of the children’s game of “Telephone”. In our version of
91 the game the original message (the “seed”) was a recording of an environmental sound. The
92 initial group of participants imitated these seed sounds. The next generation imitated the
93 previous imitators, and so on for up to 8 generations.

94 Our approach uses a transmission chain methodology similar to that frequently used in
95 experimental studies of language evolution [35]. As with other transmission chain studies
96 (and iterated learning studies more generally), we sought to discover how various biases and
97 constraints of individuals changed the nature of a linguistic signal. While typical
98 transmission chain studies focus on the impact of learning biases [36], here we use iterated
99 reproduction which does not involve any learning. Participants simply attempt to imitate a
100 sound as best as they can.

101 After collecting the imitations, we conducted a series of analyses and additional
102 experiments to systematically answer the following questions: First, do imitations stabilize in
103 form and become more word-like as they are repeated? Second, do the imitations retain a
104 resemblance to the original environmental sound that inspired them? If so, it should be
105 possible for naïve participants to match the emergent words back to the original seed sounds.
106 Third, do the imitations become more suitable as categorical labels for the sounds that
107 motivated them? For example, does the imitation of a particular water-splashing sound
108 become, over generations of repeated imitation, a better label for the more general category
109 of water-splashing sounds?

Stabilization of imitations through repetition

In the first experiment, we collected the vocal imitations, and assessed the extent to which repeating imitations of environmental sounds results in progressive stabilization toward more word-like forms in three ways. First, we measured changes in the perception of acoustic similarity between subsequent generations of imitations. Second, we used algorithmic measures of acoustic similarity to assess the similarity of imitations sampled within and between transmission chains. Third, we obtained transcriptions of imitations, and measured the extent to which later generation imitations were transcribed with greater consistency and agreement. The results show that repeated imitation results in vocalizations that are easier to repeat with high fidelity and more consistently transcribed into English letters.

Methods

Selecting seed sounds. To avoid sounds with lexicalized or conventionalized onomatopoeic forms in English, we used inanimate categories of environmental sounds. We ensured that the sounds within each category were approximately equally distinguishable by using an odd-one-out norming procedure ($N=105$ participants; see Fig. S1), resulting in a final set of 16 sounds, 4 in each of 4 categories: glass (breaking), paper (tearing), water (splashing), zipper (moving).

Collecting vocal imitations. We recruited 94 participants from Amazon Mechanical Turk. Participants were instructed that they would hear some sound and their task was to reproduce it as accurately as possible using their computer microphone. Full instructions are provided in the Supplemental Materials.

Each participant listened to and imitated four sounds: one from each of the four categories. Sounds were assigned at random such that participants were unlikely to imitate the same person more than once. Participants were allowed to listen to each target sound as many times as they wished, but were only allowed a single recording in response. Recordings that were too quiet (less than -30 dBFS) were not accepted.

136 A total of 115 (24%) imitations were removed for being poor quality (e.g., loud
 137 background sounds) or for violating the rules of the experiment (e.g., an utterance in
 138 English). The final sample contained 365 imitations along 105 contiguous transmission
 139 chains (Fig. 1).

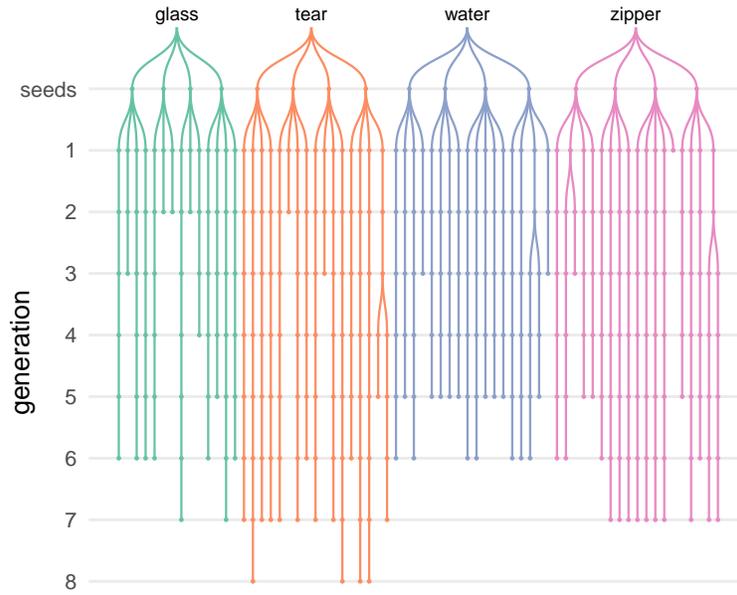


Figure 1. Vocal imitations collected in the transmission chain experiment. Seed sounds (16) were sampled from four categories of environmental sounds: glass, tear, water, zipper. Participants imitated each seed sound, and then the next generation of participants imitated the imitations, and so on, for up to 8 generations. Chains are unbalanced due to random assignment and the above-mentioned exclusion criteria.

140 **Measuring acoustic similarity.** We obtained acoustic similarity judgments from
 141 five research assistants who listened to pairs of sounds (approx. 300 each) and rated their
 142 subjective similarity. On each trial, raters heard two sounds from subsequent generations
 143 played in random order, and indicated the similarity between the sounds on a 7- point Likert
 144 scale from *Entirely different and would never be confused* to *Nearly identical*. See
 145 Supplemental Materials for full instructions and inter-rater reliability measures.

146 We also obtained algorithmic measures of acoustic similarity using the acoustic
 147 distance functions from the Phonological Corpus Tools [37]. We computed Mel-frequency

148 cepstral coefficients (MFCCs) between pairs of imitations using 12 coefficients in order to
149 obtain speaker-independent estimates.

150 **Collecting transcriptions of imitations.** Transcriptions were obtained for the
151 first and last three generations of each transmission chain. We also transcribed the original
152 seed sounds(see Supplementary Materials, Fig. S6).

153 We recruited 216 additional participants from Amazon Mechanical Turk to listen to
154 the vocal imitations and write down what they heard as a single “word” so that the written
155 word would sound as much like the sound as possible. Participants were instructed to avoid
156 using English words in their transcriptions. Each participant completed 10 transcriptions.

157 Results

158 Imitations of environmental sounds became more stable over the course of being
159 repeated as revealed by increasing acoustic similarity judgments along individual
160 transmission chains. Acoustic similarity ratings were fit with a linear mixed-effects model
161 predicting perceived acoustic similarity from generation with random effects (intercepts and
162 slopes) for raters. To test whether the hypothesized increase in acoustic similarity was true
163 across all seed sounds and categories, we added random effects (intercepts and slopes) for
164 seed sounds nested within categories. The results showed that, across raters and seeds,
165 imitations from later generations were rated as sounding more similar to one another than
166 imitations from earlier generations, $b = 0.10$ (SE = 0.03), $t(11.9) = 3.03$, $p = 0.011$ (Fig. 2).
167 This result suggests that imitations became more stable (i.e., easier to imitate with high
168 fidelity) with each generation of repetition.

169 Although in some chains, imitations were repeated up to 8 times, an increase in
170 similarity between generations could be detected after about 5 generations. Imitations from
171 chains that did not reach 5 generations due to experimental constraints (see Fig. 1) were
172 included in all analyses, which included appropriate random effects to ensure that shorter
173 chains were weighed appropriately in the analyses. However, chains with fewer than 5

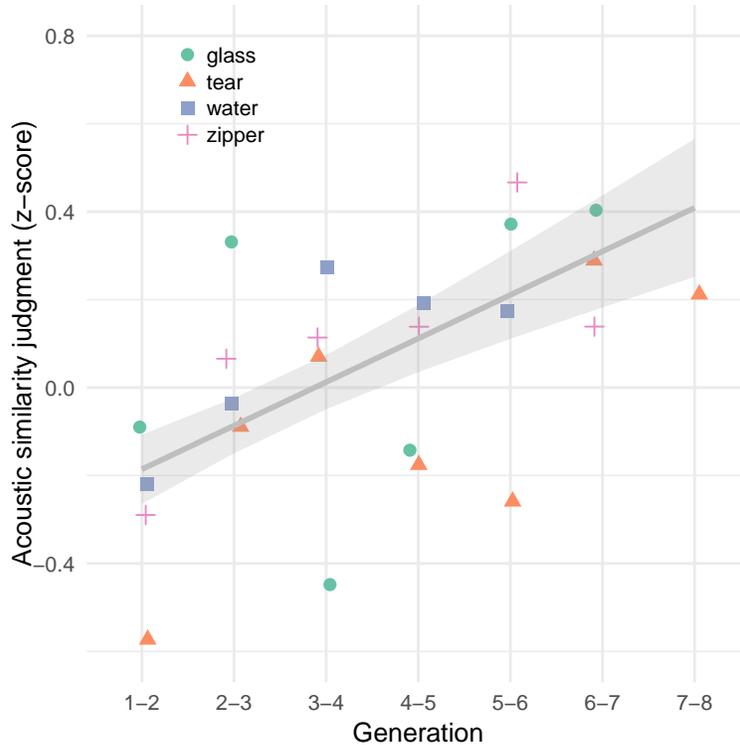


Figure 2. Change in perception of acoustic similarity over generations of iterated imitation. Points depict mean acoustic similarity ratings for pairs of imitations in each category. The predictions of the linear mixed-effects model are shown with ± 1 SE.

174 generations were excluded from analyses involving transcriptions of the first and last
 175 imitation in each chain because these analyses collapse across generation.

176 Increasing similarity along transmission chains could also reflect the uniform
 177 degradation of the signal due to repeated imitation, in which case acoustic similarity would
 178 increase both within as well as between chains. To test this, we calculated MFCCs for pairs
 179 of sounds sampled from within and between transmission chains across categories, and fit a
 180 linear model predicting acoustic similarity from the generation of sounds. We found that
 181 acoustic similarity increased within chains more than it increased between chains, $b = -0.07$
 182 (SE = 0.03), $t(6674.0) = -2.13$, $p = 0.033$ (Fig. S2), indicating that imitations were
 183 stabilizing on divergent acoustic forms as opposed to converging on similar forms through
 184 continuous degradation.

185 As an additional test of stabilization we measured whether later generation imitations
 186 were transcribed more consistently than first generation imitations. We collected a total of
 187 2163 transcriptions — approximately 20 transcriptions per sound. Of these, 179
 188 transcriptions (8%) were removed because they contained English words. Some examples of
 189 the final transcriptions are presented in Table 1.

Table 1

Examples of words transcribed from imitations.

Category	First generation	Last generation
glass	dirrng	wayew
tear	feeshefee	cheecheea
water	boococucuwich	galong
zipper	bzzzzup	izzip

190 To measure the similarity among transcriptions for a given imitation, we calculated the
 191 average orthographic distance between the most frequent transcription and all other
 192 transcriptions of the same imitation. We then fit a hierarchical linear model predicting
 193 orthographic distance from the generation of the imitation (First generation, Last
 194 generation) with random effects (intercepts and slopes) for seed sound nested within
 195 category. The results showed that transcriptions of last generation imitations were more
 196 similar to one another than transcriptions of first generation imitations, $b = -0.12$ (SE =
 197 0.03), $t(3.0) = -3.62$, $p = 0.035$ (Fig. S3). The same result is reached through alternative
 198 measures of orthographic distance (Fig. S4). Differences between transcriptions of human
 199 vocalizations and transcriptions directly of environmental sound cues are reported in the
 200 Supplementary Materials (Fig. S6).

201 **Discussion**

202 Repeating imitations of environmental sounds over generations of imitators was
203 sufficient to create more word-like forms (defined here in terms of acoustic stability and
204 orthographic agreement), even without any explicit intent to communicate. With each
205 repetition, the acoustic forms of the imitations became more similar to one another,
206 indicating that it became easier to repeat them with greater consistency. The possibility that
207 this similarity was due to uniform degradation across all transmission chains was ruled out
208 by algorithmic analyses of acoustic similarity demonstrating that acoustic similarity
209 increased within chains but not between them. Further support for our hypothesis that
210 repeating imitations makes them more stable/word-like comes from the result showing that
211 later generation imitations were transcribed more consistently into English letters.

212 The results of Experiment 1 demonstrate the ease with which iterated imitation gives
213 rise to more stable forms. However, the results do not address how these emergent words
214 relate to the original sounds that were being imitated. As the imitations became more stable,
215 were they stabilizing on arbitrary acoustic and orthographic forms, or did they maintain
216 some resemblance to the environmental sounds that motivated them? The purpose of
217 Experiment 2 was to assess the extent to which repeated imitations and their transcriptions
218 maintained a resemblance to the original set of seed sounds.

219 **Resemblance of imitations to original seed sounds**

220 To assess the resemblance of repeated imitations to the original seed sounds, we
221 measured the ability of naïve participants to match imitations and their transcriptions back
222 to their original sound source relative to other seed sounds from either the same category or
223 from different categories (Fig. 3A). Using these match accuracies, we first asked whether and
224 for how many generations the imitations and their transcriptions could be matched back to
225 the original sounds and whether certain types of information were lost faster than other types.
226 Specifically, we tested the hypothesis that if imitations were becoming more word-like, then

227 they should also be interpreted more categorically, and thus we anticipated that imitations
228 would lose information identifying the specific source of an imitation more rapidly than
229 category information that identifies the category of environmental sound being imitated.

230 **Methods**

231 **Matching imitations to seed sounds.** Participants ($N=751$) recruited from
232 Amazon Mechanical Turk were paid to listen to imitations, one at a time, and for each one,
233 choose one of four possible sounds they thought the person was trying to imitate. The task
234 was not speeded and no feedback was provided. Participants completed 10 questions at a
235 time.

236 All imitations were tested in three question types (True seed, Category match, Specific
237 match) which differed in the relationship between the imitation and the four seed sounds
238 provided as the choices in the question (see Fig. 3A). The Question types were assigned
239 between-subject.

240 **Matching transcriptions to seed sounds.** We recruited $N=461$ participants
241 from Amazon Mechanical Turk to complete a modified version of the matching survey
242 described above. Instead of listening to imitations, participants now saw a transcription of
243 an imitation and were told that it was invented to describe one of the four presented sounds.
244 Of the unique transcriptions that were generated for each sound (imitations and seed
245 sounds), only the top four most frequent transcriptions were used in the matching
246 experiment. The distractors for all questions were between-category, i.e. true seed and
247 category match. Specific match questions were omitted.

248 **Results**

249 Response accuracies in matching imitations to seed sounds were fit by a generalized
250 linear mixed-effects model predicting match accuracy as different from chance (25%) based
251 on the type of question being answered (True seed, Category match, Specific match) and the
252 generation of the imitation. Question types were contrast coded using Category match

253 questions as the baseline condition in comparison to the other two question types, each
254 containing the actual seed that generated the imitation as one of the choices. The model
255 included random intercepts for participant, and random slopes and intercepts for seed sounds
256 nested within categories.

257 Accuracy in matching first generation imitations to seed sounds was above chance for
258 all question types, $b = 1.65$ (SE = 0.14) log-odds, odds = 0.50, $z = 11.58$, $p < 0.001$, and
259 decreased steadily over generations, $b = -0.16$ (SE = 0.04) log-odds, $z = -3.72$, $p < 0.001$.
260 After 8 generations, imitations were still recognizable, $b = 0.55$ (SE = 0.30) log-odds, odds =
261 -0.59 , $z = 1.87$, $p = 0.062$. We then tested whether this increase in difficulty was constant
262 across the three types of questions. The results are shown in Fig. 3B. Performance decreased
263 over generations more rapidly for specific match questions that required a within-category
264 distinction than for category match questions that required a between-category distinction, b
265 $= -0.08$ (SE = 0.03) log-odds, $z = -2.68$, $p = 0.007$. This suggests that the iconicity in
266 between-category information was more resistant to loss through repetition.

267 An alternative explanation of the relatively greater decrease in accuracy for specific
268 match questions is that they are simply more difficult than the category-match questions
269 because the sounds presented as choices are more acoustically similar to one another.
270 However, performance also decreased relative to the category match questions for the easiest
271 type of question where the correct answer was the actual seed generating the imitation (True
272 seed questions; see Fig. 3A). That is, the advantage of having the true seed among
273 between-category distractors decreased over generations, $b = -0.07$ (SE = 0.02) log-odds, $z =$
274 -2.77 , $p = 0.006$. Together, the observed decrease in the “true seed advantage” (the
275 advantage of having the actual seed among the choices) and the increase in the “category
276 advantage” (the advantage of having between-category distractors) shows that the changes
277 induced by repeated imitation caused the imitations to lose some of properties that linked
278 the earlier imitations to the specific sound that motivated them, while nevertheless
279 preserving a more abstract category-based resemblance.

280 We next report the results of matching the written transcriptions of the auditory
281 sounds back to the original environmental sounds. Remarkably, participants were able to
282 guess the correct meaning of a word that was transcribed from an imitation that had been
283 repeated up to 8 times, $b = 0.83$ (SE = 0.13) log-odds, odds = -0.18, $z = 6.46$, $p < 0.001$
284 (Fig. 3C) both for True seed questions containing the actual seed generating the transcribed
285 imitation, $b = 0.75$ (SE = 0.15) log-odds, $z = 4.87$, $p < 0.001$, and for Category match
286 questions where participants had to associate transcriptions with a particular category of
287 environmental sounds, $b = 1.02$ (SE = 0.16) log-odds, $z = 6.39$, $p < 0.001$. The effect of
288 generation did not vary across these question types, $b = 0.05$ (SE = 0.10) log-odds, $z = 0.47$,
289 $p = 0.638$. The results of matching “transcriptions” directly of the environmental sounds are
290 shown in Fig. S6.

291 Discussion

292 Even after being repeated up to 8 times across 8 different individuals, vocalizations
293 retained a resemblance to the environmental sound that motivated them. This resemblance
294 remained even after the vocalizations were transcribed into orthographic forms. For vocal
295 imitations, but not for transcriptions, this resemblance was stronger for the category of
296 environmental sound than the specific seed sound, suggesting that iterated imitation
297 produces vocalizations that are interpreted by naïve listeners in a more categorical way.
298 Iterated imitation appears to strip the vocalizations of some of the characteristics that
299 individuate each particular sound while maintaining some category-based resemblance. This
300 happened even though participants were never informed about the meaning of the
301 vocalizations and were not trying to communicate.

302 Transcriptions of the vocalizations, like the vocalizations themselves, were able to be
303 matched to the original environmental sounds at levels above chance. Unlike vocalizations,
304 the transcriptions continued to be matched more accurately to the true seed compared to the
305 general category; transcription appeared to impact specific and category-level information

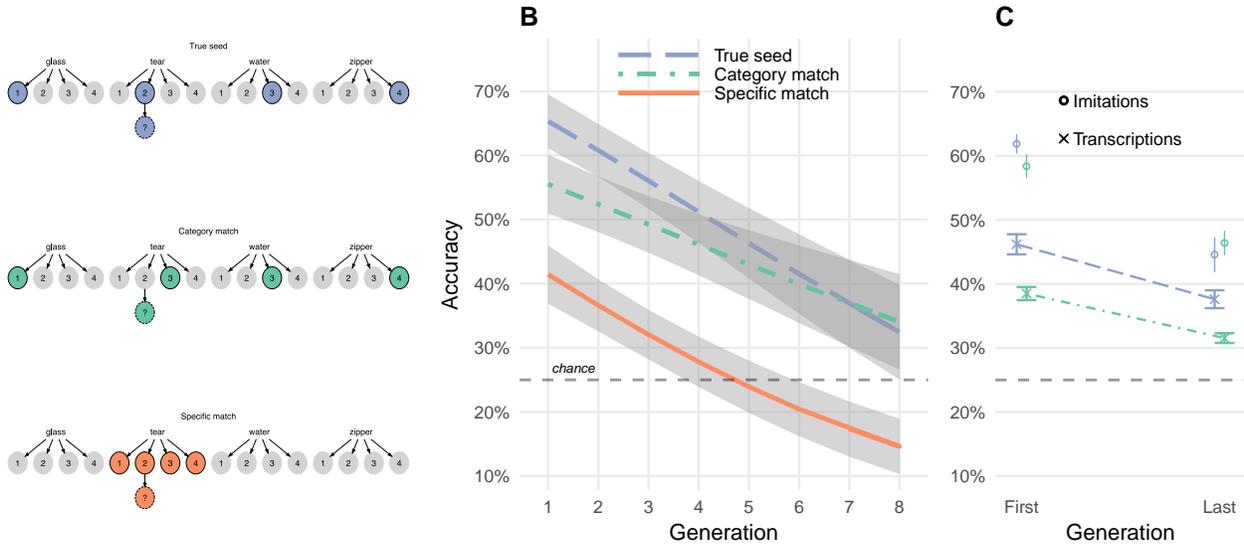


Figure 3. Repeated imitations retained category resemblance. A. Three types of matching questions. True seed and category match questions contained choices from different sound categories. Specific match questions pitted the actual seed against the other seeds within the same category. B. Accuracy in matching vocal imitations to original seed sounds. Curves show predictions of the generalized linear mixed effects models with ± 1 SE of the model predictions. C. Accuracy in matching transcriptions of the imitations to original seed sounds (e.g., “boococuwich” to a water splashing sound). Circles show mean matching accuracy for the vocal imitations that were transcribed for comparison.

306 equally. One possible explanation of the difference between the acoustic and orthographic
 307 forms of this task is that the process of transcribing a non-linguistic vocalization into a
 308 written word encourages transcribers to emphasize individuating information about the
 309 vocalization. However, this does not provide a complete explanation of our results: the fact
 310 that transcriptions of imitations can be matched back to other category members (Category
 311 match questions) suggests that transcriptions still do carry some category information.
 312 Another possibility is that by selecting only the most frequent transcriptions, we
 313 unintentionally excluded less frequent transcriptions that were more diagnostic of category
 314 information.

315 Experiments 1 and 2 document a process of gradual change from an imitation of an
316 environmental sound to a more word-like form. But do these emergent words function like
317 other words in a language? In Experiment 3, we test the suitability of imitations taken from
318 the beginning and end of transmission chains in serving as category labels in a category
319 learning task.

320 **Suitability of created words as category labels**

321 If, as we claim, repeated imitation leads to more word-like forms, they should make for
322 better category labels. For example, an imitation from a later generation may be easier to
323 learn as a label for the category of sounds that motivated it than an earlier imitation, which
324 is more closely yoked to a particular environmental sound. To the extent that repeating
325 imitations abstract away the idiosyncrasies of a particular category member [38,39], it may
326 also be easier to generalize later imitations to new category members. We tested these
327 predictions using a category learning task in which participants learned novel labels for the
328 categories of environmental sounds. The novel labels were transcriptions of either first or last
329 generation imitations gathered in Experiment 1.

330 **Methods**

331 **Selecting words to learn as category labels.** Of the unique words created
332 through the transmission chain and transcription procedures, we sampled 56 words
333 transcribed from first and last generation imitations that were equated in terms of length and
334 match accuracy to the original sounds (see Supplementary Materials for additional details).

335 **Procedure.** Participants ($N=67$) were University of Wisconsin undergraduates.
336 Participants were tasked with learning to associate novel labels (transcriptions of seed
337 sounds) with the original seed sounds. Full instructions are provided in the Supplementary
338 Materials. Participants were assigned between-subject to learn labels of either first or last
339 generation imitations. On each trial, participants heard one of the 16 seed sounds. After a 1s
340 delay, participants saw a label (one of the transcribed imitations) and responded *yes* or *no*

341 using a gamepad controller depending on whether the sound and the word went together.
342 Participants received accuracy feedback (a bell sound and a green checkmark if correct; a
343 buzzing sound and a red “X” if incorrect). Four outlier participants were excluded due to
344 high error rates and slow RTs.

345 Participants categorized all 16 seed sounds over the course of the experiment, but they
346 learned them in blocks of 4 sounds at a time. Within each block of 24 trials, participants
347 heard the same four sounds and the same four words multiple times, with a 50% probability
348 of the sound matching the word on any given trial. At the start of a new block of trials,
349 participants heard four new sounds they had not heard before, and had to learn to associate
350 these new sounds with the words they had learned in the previous blocks.

351 Results

352 Participants began by learning through trial-and-error to associate four written labels
353 with four categories of environmental sounds. The small number of categories made this an
354 easy task (mean accuracy after the first block of 24 trials was 81%; Fig. S5). Participants
355 learning transcriptions of first or last generation imitations did not differ in overall accuracy,
356 $p = 0.887$, or reaction time, $p = 0.616$.

357 After this initial learning phase (i.e. after the first block of trials), accuracy
358 performance quickly reached ceiling and did not differ between groups $p = 0.775$. However,
359 the response times of participants learning last generation transcriptions declined more
360 rapidly with practice than participants learning first generation transcriptions, $b = -114.13$
361 (SE = 52.06), $t(39.9) = -2.19$, $p = 0.034$ (Fig. 4A). These faster responses suggest that, in
362 addition to becoming more stable both in terms of acoustic and orthographic properties,
363 repeated imitations become easier to process as category labels. We predict that a harder
364 task (i.e. more than four categories and 16 exemplars) would also yield differences in initial
365 learning rates.

366 Next, we examined specifically whether transcriptions from last generation imitations

367 were easier to generalize to novel category exemplars by comparing RTs on trials
 368 immediately prior to the introduction of novel sounds (new category members) and the first
 369 trials after the block transition (± 6 trials). The results revealed a reliable interaction
 370 between the generation of the transcribed imitation and the block transition, $b = -110.77$
 371 ($SE = 52.84$), $t(39.7) = -2.10$, $p = 0.042$ (Fig. 4B). This result suggests that transcriptions
 372 from later generation imitations were easier to generalize to new category members.

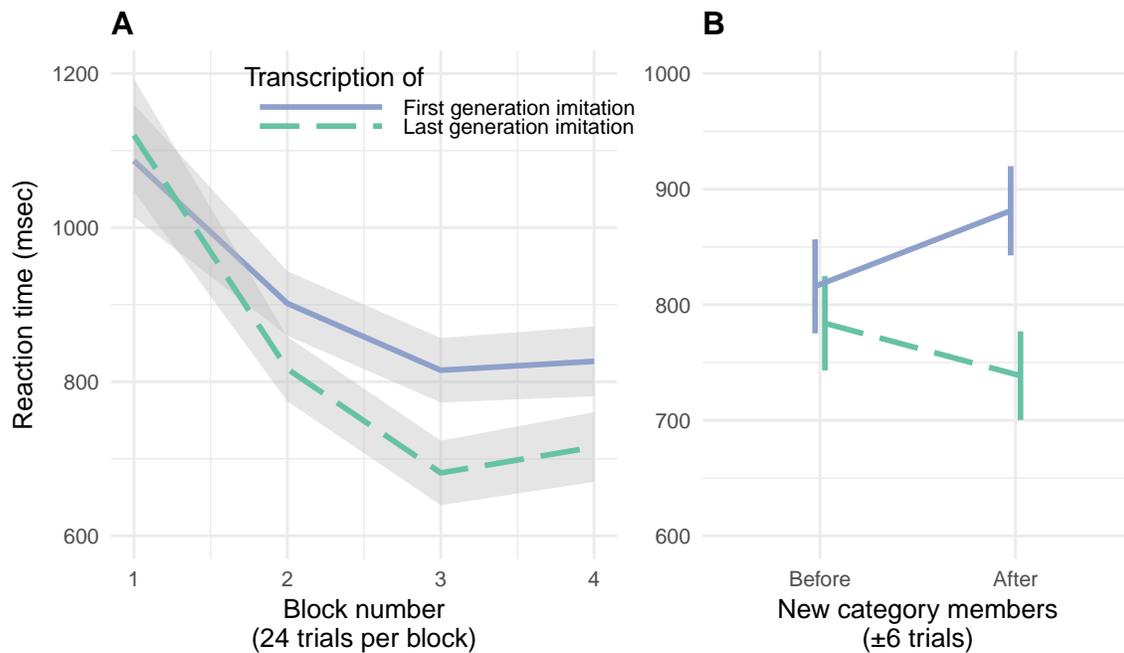


Figure 4. Repeated imitations made for better category labels. A. Mean RTs for correct responses in the category learning experiment with ± 1 SE. B. Cost of generalizing to new category members with ± 1 SE.

373 Discussion

374 Transcriptions of vocal imitations that have undergone more repetitions were processed
 375 more quickly, and easier to generalize to new category members. These results show how
 376 repeated imitation may lead to more stable forms that are in turn easier to integrate into the
 377 language as category labels.

General Discussion

378

379 Accumulating evidence shows that iconic words are prevalent across the spoken
380 languages of the world [23,24,30]. Counter to past assumptions about the limitations of
381 human vocal imitation, people are surprisingly effective at using vocal imitation to represent
382 and communicate about the sounds in their environment [33] and more abstract meanings
383 [31]. These findings raise the possibility that early spoken words originated from vocal
384 imitations, perhaps comparable to the way that many of the signs of signed languages
385 appear to be formed originally from pantomimes [31,40]. Here, we examined whether simply
386 repeating an imitation of an environmental sound — with no intention to create a new word
387 or even to communicate — produces more word-like forms.

388 Our results show that through unguided repetition, imitative vocalizations became
389 more word-like both in form and function. In form, the vocalizations gradually stabilized
390 over generations, becoming more similar from imitation to imitation. The standardization
391 was also found when the vocalizations were transcribed into English letters. Even as the
392 vocalizations became more word-like, they maintained a resemblance to the original
393 environmental sounds that motivated them. Notably, this resemblance appeared more
394 resilient with respect to the category of sound (e.g., water-splashing sounds), rather than to
395 the specific exemplar (a particular water-splashing sound). After eight generations the
396 vocalizations could no longer be matched to the specific sound from which they originated
397 any more accurately than they could be matched to the general category of environmental
398 sound. Thus, information that distinguished an imitation from other sound categories was
399 more resistant to transmission decay than exemplar information within a category. The
400 resemblance to the original sounds was maintained even when the vocalizations were
401 transcribed into a written form: participants were able to match the transcribed
402 vocalizations to the original sound category at levels above chance.

403 We further tested the hypothesis that repeated imitation led to vocalizations becoming
404 more word-like by testing the ease with which people learned the (transcribed) vocalizations

405 as category labels (e.g., “pshfft” from generation 1 vs. “shewp” from generation 8 as labels
406 for tearing sounds) (Exp. 3). Labels from the last generation were responded to more
407 quickly than labels from the first generation. More importantly the labels from the last
408 generation generalized better to novel category members. This fits with previous research
409 showing that the relatively arbitrary forms that are typical of words (e.g. “dog”) makes
410 them better suited to function as category labels compared to direct auditory cues (e.g., the
411 sound of a dog bark) [38,39,41].

412 Compared to the large number of iconic signs in signed languages [8], the number of
413 iconic words in spoken languages may appear to be very small [42,43]. However, increasing
414 evidence from disparate language suggests that vocal imitation is, in fact, a widespread
415 source of vocabulary. Cross-linguistic surveys indicate that onomatopoeia—iconic words used
416 to represent sounds—are a universal lexical category found across the world’s languages [44].
417 Even English, a language that has been characterized as relatively limited in iconic
418 vocabulary [45], is documented as having hundreds of onomatopoeic words not only for
419 animal and human vocalizations (“meow”, “tweet”, “slurp”, “babble”, murmur”), but also for
420 a variety of environmental sounds (e.g., “ping”, “click”, “plop”) [34,46]. Besides words that
421 directly resemble sounds — the focus of the present study — many languages contain
422 semantically broader inventories of ideophones. These words comprise a grammatically and
423 phonologically distinct class of words that are used to express various sensory-rich meanings,
424 such as qualities related to manner of motion, visual properties, textures and touch, inner
425 feelings and cognitive states [44,47,48]. As with onomatopoeia, ideophones are often
426 recognized by naïve listeners as bearing a degree of resemblance to their meaning [49].

427 Our study focused on imitations of environmental sounds as a source domain of
428 meaning. Additional work is required to determine the extent to which vocal imitation can
429 ground *de novo* vocabulary in other semantic domains [31,50]. Our hypothesis that vocal
430 imitation may have played a role in the origin of some of the first spoken words does not
431 preclude that gesture played an equal or more important role in establishing the first

432 linguistic conventions [8,9,51]. In addition, the present studies—like nearly all experimental
433 investigations of the evolution of language—are limited in their inferential power by the use
434 of participants who already speak at least one language. It may turn out that the ability to
435 repeat vocal imitations and converge on more word-like forms only arises in people who
436 already know and use a full linguistic system, which would limit the relevance of our findings
437 for the origins of spoken words.

438 Although our results show that repeated imitations lead to increases in stability of
439 spoken (as well as transcribed) forms, we recognize that there are additional requirements for
440 the vocalizations to be incorporated into a linguistic system. One of these may be familiarity
441 with the referents that are being imitated. The extent to which our results depend on prior
442 familiarity with the referents can be measured by extending our procedure to less familiar
443 referential domains. Another design limitation is the use of auditory referents that can be
444 imitated (environmental sounds). But although vocal imitation may seem to be restricted to
445 auditory referents, prior results indicate that people show considerable agreement on how to
446 vocally “imitate” non-auditory and even somewhat abstract meanings [31,50].

447 Among the qualities that distinguish natural language from other communication
448 systems is the extreme diversity of signals (e.g. words) that individuals learn and use, and
449 the speed with which these signals change over generations of speakers. As a consequence,
450 the origins of most spoken words are opaque, making it difficult to investigate the process by
451 which they were formed. Our experimental results show that the transition from vocal
452 imitation to more word-like signals can, in some cases, be a rapid and simple process. The
453 mere act of repeated imitation can drive vocalizations to become more word-like in both
454 form and function with the vocalizations nevertheless still retaining some resemblance to
455 their real-world referents. These findings suggest that repeated vocal imitation may
456 constitute a significant mechanism for the origin of new words. It remains for future work to
457 determine the extent to which the functioning of this process depends on the linguistic
458 competencies of modern humans.

Ethics

This was approved by the University of Wisconsin-Madison's Educational and Social/Behavioral Sciences IRB and conducted in accordance with the principles expressed in the Declaration of Helsinki. Informed consent was obtained for all participants.

Data, code, and materials

Our data, methods, materials, and analysis scripts, are available at osf.io/3navm.

Competing interests

We have no competing interests.

Authors' contributions

P.E., M.P., and G.L. designed the research. P.E. conducted the research and analyzed the data. P.E., M.P., and G.L. wrote the manuscript.

Funding

This research was supported by NSF 1344279 awarded to G.L.

References

1. Seyfarth RM, Cheney DL. 1986 Vocal development in vervet monkeys. *Animal Behaviour* **34**, 1640–1658.
2. Brysbaert M, Stevens M, Mandera P, Keuleers E. 2016 How Many Words Do We Know? Practical Estimates of Vocabulary Size Dependent on Word Definition, the Degree of Language Input and the Participant's Age. *Frontiers in Psychology* **7**, 55–11.
3. Wierzbicka A. 1996 *Semantics: Primes and universals: Primes and universals*. Oxford University Press, UK.
4. Evans N, Levinson SC. 2009 The myth of language universals: Language diversity and its importance for cognitive science. *Brain and Behavioral Sciences* **32**, 429–492.

- 482 5. Pagel M, Atkinson QD, Meade A. 2007 Frequency of word-use predicts rates of lexical
483 evolution throughout Indo-European history. *Nature* **449**, 717–720.
- 484 6. Sapir E. 1921 *Language: An introduction to the study of speech*. New York: Harcourt,
485 Brace; Company.
- 486 7. Labov W. 1972 *Sociolinguistic patterns*. University of Pennsylvania Press.
- 487 8. Goldin-Meadow S. 2016 What the hands can tell us about language emergence.
488 *Psychonomic Bulletin & Review* **24**, 1–6.
- 489 9. Kendon A. 2014 Semiotic diversity in utterance production and the concept of 'language'.
490 *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**,
491 20130293–20130293.
- 492 10. Klima ES, Bellugi U. 1980 *The signs of language*. Harvard University Press.
- 493 11. Frishberg N. 1975 Arbitrariness and Iconicity: Historical Change in American Sign
494 Language. *Language* **51**, 696–719.
- 495 12. Stokoe W. 1965 *Dictionary of the American Sign Language based on scientific principles*.
496 Gallaudet College Press, Washington.
- 497 13. Wescott RW. 1971 Linguistic iconism. *Linguistic Society of America* **47**, 416–428.
- 498 14. Goldin-Meadow S, Feldman H. 1977 The development of language-like communication
499 without a language model. *Science* **197**, 401–403.
- 500 15. Fay N, Lister CJ, Ellison TM, Goldin-Meadow S. 2014 Creating a communication system
501 from scratch: Gesture beats vocalization hands down. *Frontiers in Psychology* **5**, 663.
- 502 16. Arbib MA. 2012 *How the brain got language: The mirror system hypothesis*. Oxford
503 University Press.
- 504 17. Armstrong DF, Wilcox S. 2007 *The gestural origin of language*. Oxford University Press.
- 505 18. Corballis MC. 2003 *From hand to mouth: The origins of language*. Princeton University
506 Press.
- 507 19. Hewes GW. 1973 Primate Communication and the Gestural Origin of Language.
508 *Current Anthropology* **14**, 5–24.

- 509 20. Hockett CF. 1978 In search of Jove's brow. *American speech* **53**, 243–313.
- 510 21. Tomasello M. 2010 *Origins of human communication*. MIT press.
- 511 22. Pinker S, Jackendoff R. 2005 The faculty of language: what's special about it? *Cognition*
512 **95**, 201–236.
- 513 23. Dingemanse M, Blasi DE, Lupyan G, Christiansen MH, Monaghan P. 2015 Arbitrariness,
514 Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences* **19**, 603–615.
- 515 24. Perniss P, Thompson RL, Vigliocco G. 2010 Iconicity as a General Property of Language:
516 Evidence from Spoken and Signed Languages. *Frontiers in Psychology* **1**.
- 517 25. Clark HH, Gerrig RJ. 1990 Quotations as demonstrations. *Language* **66**, 764–805.
- 518 26. Lewis J. 2009 As well as words: Congo Pygmy hunting, mimicry, and play. In *The cradle*
519 *of language*, The cradle of language.
- 520 27. Brown RW, Black AH, Horowitz AE. 1955 Phonetic symbolism in natural languages.
521 *Journal of abnormal psychology* **50**, 388–393.
- 522 28. Dingemanse M. 2014 Making new ideophones in Siwu: Creative depiction in
523 conversation. *Pragmatics and Society*
- 524 29. Donald M. 2016 Key cognitive preconditions for the evolution of language. *Psychonomic*
525 *Bulletin & Review*, 1–5.
- 526 30. Imai M, Kita S. 2014 The sound symbolism bootstrapping hypothesis for language
527 acquisition and language evolution. *Philosophical Transactions of the Royal Society*
528 *B: Biological Sciences* **369**.
- 529 31. Perlman M, Dale R, Lupyan G. 2015 Iconicity can ground the creation of vocal symbols.
530 *Royal Society Open Science* **2**, 150152–16.
- 531 32. Lemaitre G, Rocchesso D. 2014 On the effectiveness of vocal imitations and verbal
532 descriptions of sounds. *The Journal of the Acoustical Society of America* **135**,
533 862–873.
- 534 33. Lemaitre G, Houix O, Voisin F, Misdariis N, Susini P. 2016 Vocal Imitations of
535 Non-Vocal Sounds. *PloS one* **11**, e0168167–28.

- 536 34. Rhodes R. 1994 Aural images. *Sound symbolism*, 276–292.
- 537 35. Tamariz M. 2017 Experimental Studies on the Cultural Evolution of Language. *Annual*
538 *Review of Linguistics* **3**, 389–407.
- 539 36. Kirby S, Cornish H, Smith K. 2008 Cumulative cultural evolution in the laboratory: an
540 experimental approach to the origins of structure in human language. *Proceedings of*
541 *the National Academy of Sciences* **105**, 10681–10686.
- 542 37. Hall KC, Allen B, Fry M, Mackie S, McAuliffe M. 2016 Phonological CorpusTools. *14th*
543 *Conference for Laboratory Phonology*
- 544 38. Edmiston P, Lupyan G. 2015 What makes words special? Words as unmotivated cues.
545 *Cognition* **143**, 93–100.
- 546 39. Lupyan G, Thompson-Schill SL. 2012 The evocative power of words: Activation of
547 concepts by verbal and nonverbal means. *Journal of Experimental Psychology:*
548 *General* **141**, 170–186.
- 549 40. Fay N, Ellison TM, Garrod S. 2014 Iconicity: From sign to system in human
550 communication and language. *Pragmatics and Cognition* **22**, 244–263.
- 551 41. Boutonnet B, Lupyan G. 2015 Words Jump-Start Vision: A Label Advantage in Object
552 Recognition. *Journal of Neuroscience* **35**, 9329–9335.
- 553 42. Crystal D. 1987 *The Cambridge Encyclopedia of Language*. Cambridge Univ Press.
- 554 43. Newmeyer FJ. 1992 Iconicity and generative grammar. *Language*
- 555 44. Dingemanse M. 2012 Advances in the Cross-Linguistic Study of Ideophones. *Language*
556 *and Linguistics Compass* **6**, 654–672.
- 557 45. Vigliocco G, Perniss P, Vinson D. 2014 Language as a multimodal phenomenon:
558 implications for language learning, processing and evolution. *Philosophical*
559 *Transactions of the Royal Society B: Biological Sciences* **369**, 20130292–20130292.
- 560 46. Sobkowiak W. 1990 On the phonostatistics of English onomatopoeia. *Studia Anglica*
561 *Posnaniensia* **23**, 15–30.
- 562 47. Nuckolls JB. 1999 The case for sound symbolism. *Annual Review of Anthropology* **28**,

- 563 225–252.
- 564 48. Voeltz FE, Kilian-Hatz C. 2001 *Ideophones*. John Benjamins Publishing.
- 565 49. Dingemanse M, Schuerman W, Reinisch E. 2016 What sound symbolism can and cannot
566 do: Testing the iconicity of ideophones from five languages. *Language* **92**.
- 567 50. Perlman M, Lupyán G. 2018 People Can Create Iconic Vocalizations to Communicate
568 Various Meanings to Naïve Listeners. *Scientific Reports*
- 569 51. Fay N, Arbib MA, Garrod S. 2013 How to Bootstrap a Human Communication System.
570 *Cognitive Science* **37**, 1356–1367.

571

Table captions

572 *Table 1.* Examples of words transcribed from imitations.

Figure captions

573

574 *Figure 1.* Vocal imitations collected in the transmission chain experiment. Seed
575 sounds (16) were sampled from four categories of environmental sounds:
576 glass, tear, water, zipper. Participants imitated each seed sound, and
577 then the next generation of participants imitated the imitations, and
578 so on, for up to 8 generations. Chains are unbalanced due to random
579 assignment and the above-mentioned exclusion criteria.

580 *Figure 2.* Change in perception of acoustic similarity over generations of iterated
581 imitation. Points depict mean acoustic similarity ratings for pairs of
582 imitations in each category. The predictions of the linear mixed-effects
583 model are shown with ± 1 SE.

584 *Figure 3.* Repeated imitations retained category resemblance. A. Three types of
585 matching questions. True seed and category match questions contained
586 choices from different sound categories. Specific match questions pitted
587 the actual seed against the other seeds within the same category. B.
588 Accuracy in matching vocal imitations to original seed sounds. Curves
589 show predictions of the generalized linear mixed effects models with ± 1
590 SE of the model predictions. C. Accuracy in matching transcriptions of
591 the imitations to original seed sounds (e.g., “boococucuwich” to a water
592 splashing sound). Circles show mean matching accuracy for the vocal
593 imitations that were transcribed for comparison.

594 *Figure 4.* Repeated imitations made for better category labels. A. Mean RTs for
595 correct responses in the category learning experiment with ± 1 SE. B.
596 Cost of generalizing to new category members with ± 1 SE.