Creating words from iterated vocal imitation

Pierce Edmiston[a, 1], **Marcus Perlman**[b], **and Gary Lupyan**[a]

[a]University of Wisconsin-Madison, Department of Psychology, 1202 W. Johnson St., Madison, WI 53703; [b]Max Planck Institute for Psycholinguistics, Nijmegen, 6500 AH The Netherlands

This manuscript was compiled on May 11, 2017

We investigated how conventional words might emerge from the unguided repetition of nonword imitations. Participants played a version of the children's game "Telephone". The first generation imitated recognizable environmental sounds (e.g., glass breaking, water splashing) and subsequent generations imitated the imitations of the prior generation for a maximum of 8 generations. We then examined whether the vocal imitations became more word-like, became more suitable as learned category labels, and retained a resemblance to the original sound. The results showed that the imitations became more stable and word-like, and more easily learnable as category labels. At the same time, even after as many as 8 generations, both spoken imitations and their written transcriptions could be matched above chance to the category of environmental sound that motivated them. These results show how repeated imitation can create progressively more word-like forms while retaining a semblance of iconicity with the original sound. We interpret these results as evidence for the role of human vocal imitation in explaining the origins of human language.

language evolution | iconicity | vocal imitation | transmission chain

People have long pondered the origins of languages, especially the words that compose them. For example, both Plato in his *Cratylus* dialogue (1) and John Locke in his *Essay Concerning Human Understanding* (2) examined the "naturalness" of words–whether they are somehow imitative of their meaning. Here, we investigated whether new words can be formed from the repetition of non-verbal imitations. Does the repetition of imitations over generations of speakers gradually give rise to novel word forms? In what ways do these words resemble the imitations that originated them? We report a large-scale experiment investigating how new words can form–gradually and without instruction–simply by repeating imitations of environmental sounds.

The importance of imitation and depiction in the origin of signs is clearly observable in signed languages (3), but in considering the idea that vocal imitation may be key to understanding the origin of spoken words, many have argued that the human capacity for vocal imitation is far too limited (4–8). For example, Pinker and Jackendoff (9) argued that, "most humans lack the ability (found in some birds) to convincingly reproduce environmental sounds ... Thus 'capacity for vocal imitation' in humans might be better described as a capacity to learn to produce speech" (p. 209). Consequently, it is still widely assumed that vocal imitation–or more broadly, the use of any sort of resemblance between form and meaning–cannot be important to understanding the origin of spoken words (10, 11).

But although most words of contemporary spoken languages are not clearly imitative in origin, there has been a growing recognition of the preponderance of imitative words in spoken language (12, 13) and the frequent use of vocal imitation and depiction in spoken discourse (14, 15), leading some to argue for the importance of imitation for understanding the origin of spoken words (e.g., 16–20). In addition, experiments show that people can, in fact, be highly effective at using vocal imitations to refer to different kinds of sounds–in some cases, even more effective than with the use of words (21). The effectiveness of these imitations arises not because people are able to mimic environmental sounds with high-fidelity, but because they are able to represent the salient features of sounds in ways that are understandable to listeners (22). Similarly, the features of onomatopoeic words might highlight distinctive aspects of the sound it represents. For example, the initial voiced, plosive /b/ in "boom" represents an abrupt, loud onset, the back vowel /u/ a low pitch, and the nasalized /m/ a slow, muffled decay (23). Recent work has also shows that people are able to create novel imitative vocalizations for more abstract meanings (e.g. 'slow', 'rough', 'good', 'many') in ways that are understandable to naïve listeners (19).

Thus, research shows that people can use vocal imitation as an effective means to communicate about the various sounds of their environment and even more abstract concepts. But how do vocal imitations become standardized words that are integrated into the vocabulary of a language? To investigate this question, we recruited participants to play an online version of the children's game of "Telephone". In the children's game, a spoken message is whispered from one person to the next. In our version, the original message or seed sound was a recording of an environmental sound. The first generation participant imitated this seed sound, the next generation imitated the previous imitation, and so on for up to 8 generations (Fig. 1).

In subsequent experiments, we systematically answered the following questions. First, does iterated imitation drive the vocalizations to stabilize in form and become more word-

Significance Statement

Although many words spoken today appear to have imitative origins (e.g., onomatopoeia), the process by which nonverbal imitations might give rise to wordlike forms has not been documented. Here we demonstrate the propensity of human vocal imitation to form more stable and repeatable utterances through unguided repetition. We also demonstrate potential cognitive benefits of forming more stable utterances: they are easier to learn and promote category inference. While becoming more conventional and wordlike, these invented words still retain a resemblance to the category of motivating sound, demonstrating how and why imitative words might pervade languages spoken today.

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

The authors declare no conflicts of interest.

[1]To whom correspondence should be addressed. E-mail: pedmiston@wisc.edu

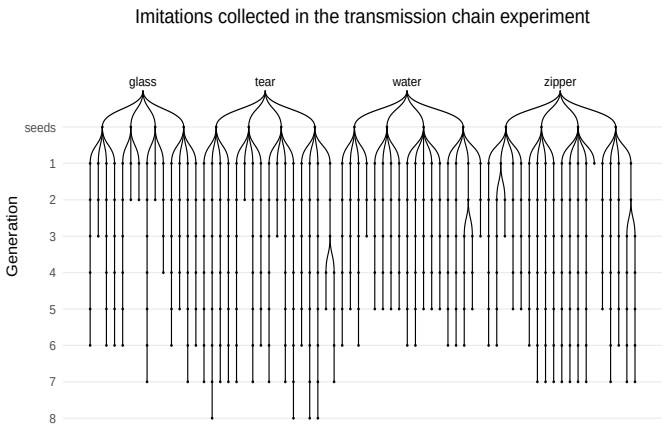

Fig. 1. The design of the transmission chain experiment. Seed sounds (16) were sampled from four categories of environmental sounds: glass, tear, water, zipper. Participants imitated each seed sound, and then the next generation of participants imitated the imitations and so on for a maximum of 8 generations.

like? Second, do the imitations become more suitable as labels for the category of sounds that motivated them? For example, does the imitation of a particular water-splashing sound become, over time, a better label for the more general category of water-splashing sounds? Third, do the imitations retain resemblance to the original environmental sounds that inspired them? If so, it should be possible for naïve participants to match the emergent imitative words back to the original sounds that motivated them.

Results

We begin with a summary of our main results. Measuring the acoustic similarity of repeated imitations revealed that imitations became more similar to one another through generations of repetition. In addition, later generations of imitations were transcribed into specific words (specific spellings) with greater agreement across different transcribers. These results suggest that imitations were stabilizing in acoustic and orthographic forms. Next, we investigated the advantages of this stabilization in terms of learnability. When transcriptions of first and last generation imitations were learned as novel labels for categories of environmental sounds, last generation transcriptions were easier to learn and faster to extend to new category members (new environmental sounds) than transcriptions of first generation imitations. While becoming more word-like, these invented words retained some resemblance to the category of environmental sound that motivated them–at least relative to the other categories tested. Participants were able to accurately match both the imitations and the transcriptions of imitations back to the category of environmental sounds that originally motivated them even after up to 8 generations of repetition. In sum, our results describe a process by which an imitation of an environmental sound may transition to a more word-like form through unguided repetition, suggest that such a transition to more word-like forms might make them more effective as category labels, and demonstrate that these created words are not entirely arbitrary and retain a resemblance to the category of environmental sounds that motivated them.

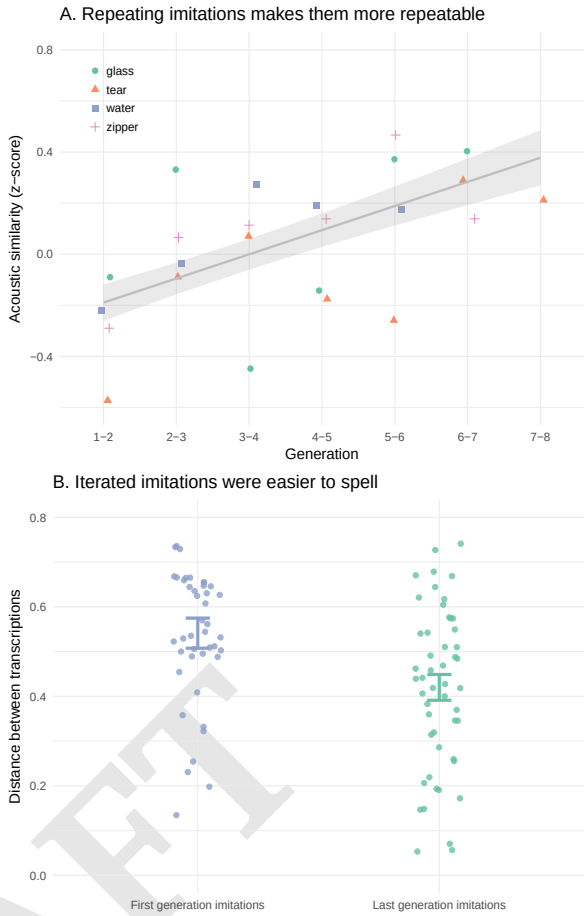

Fig. 2. Stabilization of imitations through iteration. A. Change in perception of acoustic similarity over generations of repetition. Predictions of a hierarchical linear model are shown with ±1 standard error of the model predictions. Acoustic similarity increases over generations, indicating that repeating imitations makes them more repeatable. B. Average orthographic distance among transcriptions of imitations taken from first and last generations. Each point shows the average orthographic distance (longest contiguous matching subsequence) between the most frequent transcription and all other transcriptions of a single imitation. Error bars are ±1 standard error of the hierarchical linear model predictions. Transcriptions of later generation imitations were more similar to one another than transcriptions of first generation imitations.

Iterated imitations became more stable and word-like. We collected a total of 480 imitations from 94 participants using Amazon Mechanical Turk. The final set included 365 imitations along 105 contiguous transmission chains (Fig. 1; see Methods). Research assistants coded these imitations for acoustic similarity using a blinded, pairwise comparison procedure (see Methods). Inter-rater reliability was high, ICC = 0.39, 95% CI [0.32, 0.47], $F(170, 680) = 4.18$, $p < 0.001$. Acoustic similarity ratings were fit with a hierarchical linear model predicting similarity from generation with random effects for rater and for category. Imitations from later generations were rated as sounding more similar to one another than imitations from earlier generations, $b = 0.09$ (0.02), $t(4.5) = 4.42$, $p = 0.009$ (Fig. 2A). This result suggests that the imitations become more repeatable through repetition.

We also conducted automated analyses of acoustic similarity using Mel Frequency Cepstral Coefficients (MFCCs) as a measure of acoustic distance. By this measure, imitations from later generations were not significantly more similar to

Table 1. Examples of invented words.

Category	Seed	First generation	Last generation
glass	1	tingtingting	dundunduh
glass	2	chirck	correcto
glass	3	dirrng	wayew
glass	4	boonk	baroke
tear	1	scheeept	cheecheea
tear	2	feeshefee	cheeoooo
tear	3	hhhweerrr	chhhhhhewwwe
tear	4	ccccchhhhyeaahh	shhhhh
water	1	boococucuwich	eeverlusha
water	2	chwoochwooochwooo	cheiopshpshcheiopsh
water	3	atoadelchoo	mowah
water	4	awakawush	galonggalong
zipper	1	euah	izoo
zipper	2	zoop	veeeep
zipper	3	arrgt	owww
zipper	4	bzzzzup	izzip

one another, b = 0.04 (0.03), $t(357.0) = 1.18$, $p = 0.24$. For our stimuli the correlation between automated analyses of acoustic similarity and rater judgments was low, r = 0.20, 95% CI [0.16, 0.25], suggesting to us that the automated analyses may not capture the acoustic features driving the perception of acoustic similarity. This is possibly due to the non-verbal nature of the imitations as well as variation in recording quality between participants in the online study. Further results of these automated analyses are reported in the Supporting Information.

We next tested whether the imitations became more distinguishable as particular words, as opposed to stabilizing on non-linguistic, i.e., non-English sounds. We had English-speaking participants recruited via Amazon Mechanical Turk transcribe the imitations into English orthography, and then we measured whether transcription agreement increased over generations. We selected the first and final three imitations in each transmission chain to be transcribed. As a control, we also obtained "transcriptions" of the seed sounds themselves, the results of which are reported in the Supporting Information. A total of 2163 were collected, or approximately 20 transcriptions per sound (imitation and seed sounds). Examples of the transcribed words are presented in Table 1.

To measure transcription agreement we took the average orthographic distance (longest contiguous matching subsequence) between the most frequent transcription and all other transcriptions of a given imitation. A hierarchical linear model predicting orthographic distance from the type of imitation being transcribed (First generation imitations, Last 3 generation imitations) with random effects for transmission chains nested within categories of environmental sounds. Transcriptions of later generation imitations were more similar to one another in orthographic distance than transcriptions from earlier generations, b = -0.12 (0.03), $t(3.0) = -3.62$, $p = 0.035$ (Fig. 2B). This result supports our hypothesis that unguided repetition drives imitations to become more distinctive as particular English words. The same conclusion was drawn from alternative measures of orthographic distance such as exact string matching, and when excluding imitations for which all transcriptions were unique (see Supporting Information).

Iterated imitations were easier to learn. Our hypothesis was that repetition of imitations would result in increasingly word-like forms, but what are the consequences of this transition for the language user? To examine this question, we tested whether the words created through repetition were easier to learn as category labels. We selected a sample of transcriptions taken from first and last generation imitations to use as novel labels for categories of environmental sounds. As before, a sample of transcriptions generated directly from the seed sounds was used as a control. The procedure for selecting otherwise-equal transcriptions in each of the three categories of transcriptions (first generation, last generation, seed sound control) is detailed in the Supporting Information. Here we focus on the results of the comparison between first and last generation transcriptions in the category learning experiment.

When participants learned novel labels for categories of environmental sounds, they were faster when learning a label from a last generation imitation than from a first generation imitation, b = -114.13 (52.06), $t(39.9) = -2.19$, $p = 0.034$ (Fig. 3A). In addition to becoming more stable both in terms of acoustic and orthographic properties, imitations that have been more repeated were also easier to learn as category labels. The effect can be further localized within each block. Comparing RTs on the trials leading up to a block transition and the trials immediately after the block transition revealed a reliable interaction between block transition and the generation of the transcribed label, b = -112.50 (48.96), $t(1732.0) = -2.30$, $p = 0.022$ (Fig. 3B). The same result is not found when all trials in the block are considered. This result suggests that learning transcriptions from later generation imitations were easier to initially generalize to new category members, although further investigation with a more difficult category learning experiment is required.

Iterated imitations retained seed category information. Were the imitations stabilizing on arbitrary acoustic forms or were they maintaining some resemblance to the original environmental sound? To test this, we measured the ability of participants naïve to the design of the experiment to match imitations back to the original source relative to other seed sounds from either the same category or from different categories (Fig. 4A). All 365 imitations were tested in the three conditions depicted in Fig. 4A. These conditions differed in the relationship between the imitation and the four seed sounds serving as the choices in the 4 alternative forced choice (4AFC) task. Responses were fit by hierarchical generalized linear models predicting match accuracy as different from chance (25%) based on the type of question being answered (True seed, Category match, Specific match) and the generation of the imitation.

Matching accuracy for all question types started above chance for the first generation of imitations, $b = 1.65$ (0.14) log-odds, odds = 0.50, $z = 11.58$, $p < 0.001$, and decreased steadily over generations, b = -0.16 (0.04) log-odds, $z = -3.72$, $p < 0.001$. We tested whether this increase in matching difficulty was constant across the three types of questions or if some question types became more difficult at later generations than others. The results are shown in Fig. 4B. Performance decreased over generations more rapidly for questions requiring a within-category distinction than for between-category questions, b = -0.08 (0.03) log-odds, z = -2.69, p = 0.007, suggesting that between-category information was more resistant to loss through transmission. We call this the category advantage.

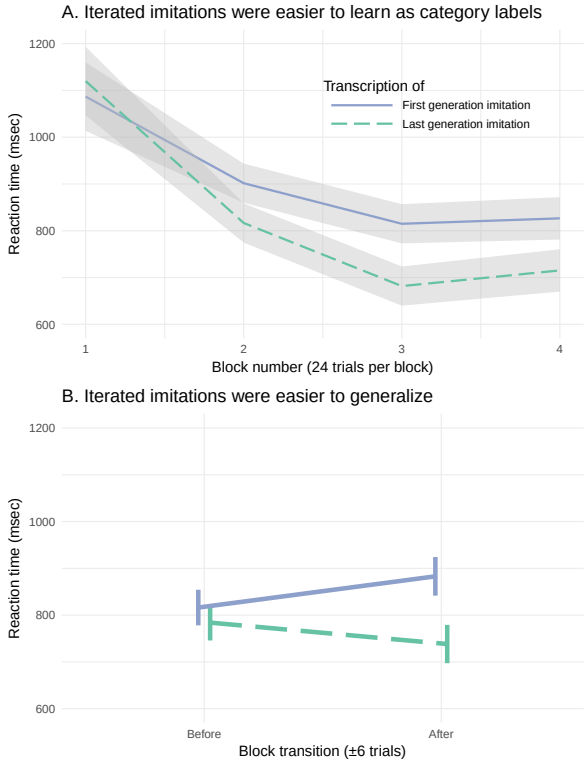

Fig. 3. (Top) RTs on correct trials by block, showing faster responses when learning category labels transcribed from last generation imitations. (Bottom) RTs on trials leading up to and immediately following the block transition where new category members are introduced.

An alternative explanation for this result is that the within-category match questions are simply more difficult* because the sounds are more acoustically similar to one another than the between-category questions and therefore performance might be expected to drop off more rapidly with repeated imitations. However, performance also decreased for the easiest type of question where the correct answer was the actual seed generating the imitation (True seed questions; see Fig. 4A); the advantage of having the true seed among between-category distractors decreased over generations, b = -0.07 (0.02) log-odds, z = -2.77, p = 0.006. Combined, the observed increase in the "category advantage" (advantage of having between-category distractors), along with a decrease in the "true seed advantage" (advantage of having the actual seed among the choices), supports our hypothesis that repetition makes imitations relatively more categorical in interpretation.

After demonstrating that repeated imitations retain a resemblance to the motivating category of environmental sounds up to 8 repetitions, we next tested whether the same was true for transcriptions of imitations. Do the imitations, once transcribed, still retain a resemblance to the category of environment sounds that motivated them? To test this, we selected a sample of the most frequent transcriptions to first and last generation imitations to use in a modified version of the "Guess the seed" game. Participants were given a novel

*We observed that performance on some Specific match questions dropped *below* chance for later generations. We were unable to account for this seeming aversion to the correct answer. For example, it was not the case that responses on low performance questions indicated that people were converging on a single (but ultimately incorrect) answer. Performance on specific match questions dropped to chance levels the most quickly, but we are unable to account for why performance might fall below chance after that.

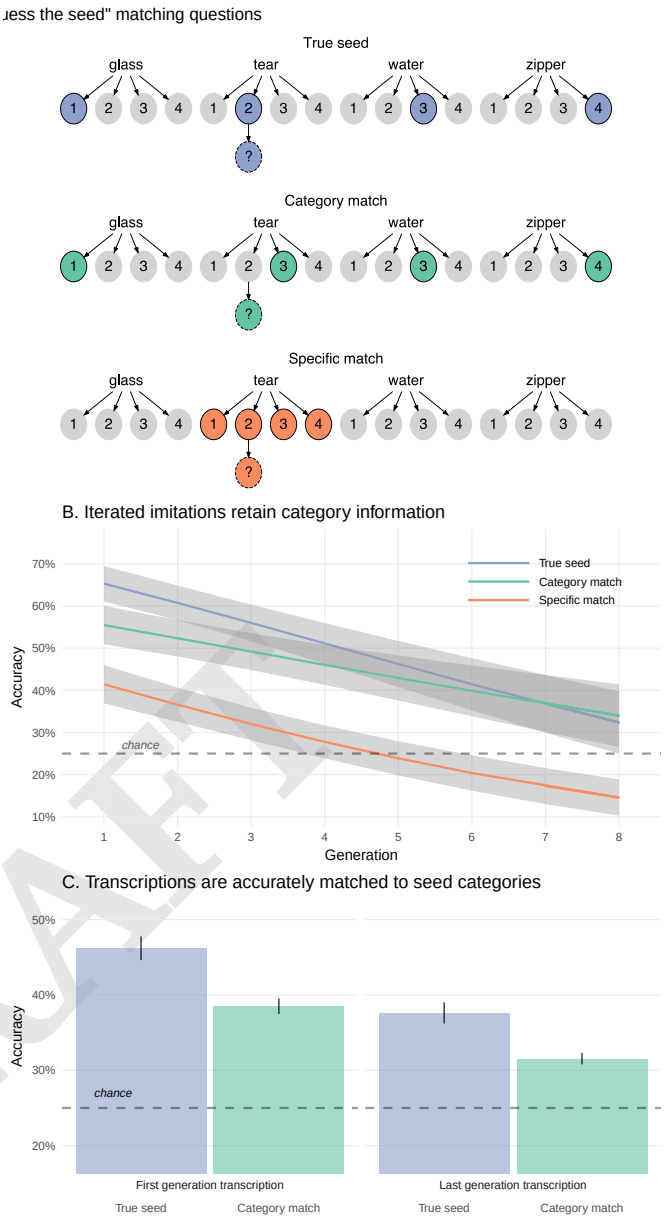

Fig. 4. Imitations retained resemblance to seed category. A. Types of 4AFC matching questions used to assess resemblance between imitations (and transcriptions of imitations) and original seed sounds. For each question, participants listened to an imitation (dashed circles) and had to guess which of 4 sound choices (solid circles) they thought the person was trying to imitate. True seed questions contained the actual sound that generated the imitation in the choices, and the three distractor sounds were sampled from different categories. Category match questions also used distractor sounds from different categories but the "correct" sound was not the actual seed, but a different sound within the same category. Specific match questions pitted the actual seed against the other seeds within the same category. B. Change in matching accuracy over generations of imitations, shown as predictions of the generalized linear models with ±1 standard error of the model predictions. The "category advantage" (Category match v. Specific match) increased over generations, while the "true seed advantage" (True seed v. Category match) decreased (see main text), suggesting that imitations lose within-category information more rapidly than between-category information. C. Change in matching accuracy over generations of imitations transcribed into English-sounding words. Specific match question accuracy was not collected for transcriptions. After up to 8 generations of repetition, imitations and transcriptions of imitations could still be matched back to the category of sound that motivated the original imitation.

word and had to guess which sound they thought the person who invented the word was talking about. The distractors for all questions were between-category, i.e. Specific match questions were not tested with transcriptions.

Participants were able to guess the correct meaning of the transcribed word above chance even after 8 generations of repetition, $b = 0.83$ (0.13) log-odds, odds = -0.18, $z = 6.46$, $p < 0.001$ (Fig. 4C). This was true both for "True seed" questions containing the actual seed generating the transcribed imitation, $b = 0.75$ (0.15) log-odds, odds = -0.28, $z = 4.87$, $p < 0.001$, and for "Category match" questions where participants had to associate transcriptions with a particular category of environmental sounds, $b = 1.02$ (0.16) log-odds, odds = 0.02, $z = 6.39$, $p < 0.001$.

Interestingly, the effect of generation did not vary across these question types, $b = 0.05$ (0.10) log-odds, $z = 0.47$, $p = 0.637$. Our theory predicts that later generation transcriptions should be less likely to be distinguished from other sounds within the same category. However, the True seed advantage persisted, indicating that transcriptions of imitations may capture idiosyncratic elements of specific category members more than the acoustic imitations themselves. An alternative reason for not observing a decrease in the True seed advantage for transcriptions is that the sample of transcriptions tested in the Guess the seed game was too small to adequately model changes happening on a per-chain level. Futher reasons for this discrepancy are explored in the Discussion.

Discussion

People can be effective at using vocal imitation to represent and communicate about the sounds in their environment, as well as more abstract concepts. Moreover, imitative (or "iconic") words are found across the spoken languages of the world (12, 13). However, little is known about the process by which vocal imitations can develop into standardized words. Must new words be deliberately invented as such, or can words form simply by repeating an imitation of a sound–even when there is no intention to communicate. To examine this question, we conducted a large-scale, online, iterated vocal imitation experiment.

Our results show that through simple repetition, imitative vocalizations became more word-like both in form and function. In form, the vocalizations gradually stabilized over generations, becoming more similar from imitation to imitation. They also became increasingly standardized according to the phonology of English, as later generations were more consistently transcribed into English orthography. In function, the imitations became more word-like in a relative increase in effectiveness as category labels. In a category learning experiment, naïve participants were faster to learn category labels derived from transcriptions of later-generation imitations than those derived from direct imitations of the environmental sound. This fits with previous research showing that the relatively arbitrary forms that are typical of words (e.g. "dog") makes them better suited to function as category labels compared to direct auditory cues (e.g. the sound of a dog bark; 24–26).

However, at the same time as the vocalizations became more word-like, they nevertheless maintained an imitative quality. Interestingly, while after eight generations they could no longer be matched to the particular sound from which they originated, the imitations could still be matched to the general category of the sound. Thus, information that distinguished an imitation from other sound categories was more resilient to transmission decay than exemplar information within a category. Even after the vocalizations were transcribed into English, participants were able to guess their original category from the written "word". However, unlike with the vocalizations, participants continued to be more accurate at matching late generation transcriptions back to their particular source sound relative to other exemplars of the category. With transcriptions, individuating information was retained over generations over and above category information. One possible explanation for this is that by converting the imitations into orthographic representations of phonemes, idiosyncratic features of the sound could become rendered as categorical phonological features. This process could exaggerate the features and facilitate identification of the source.

Our study focused on the process by which words are formed from vocal imitation, and future research remains to determine the full scope of vocal imitation as a source of vocabulary in spoken languages. Although some have estimated the number of imitative words to be small (27, 28), increasing evidence from across disparate languages shows that vocal imitation is, in fact, a widespread source of vocabulary. Cross-linguistic surveys indicate that onomatopoeia–imitative words used to represent sounds–are a universal lexical category found across the world's languages (29). Even English, a language that has been characterized as relatively limited in iconic vocabulary (30), is documented to have hundreds of words for human and animal vocalizations and various kinds of environmental sounds (23, 31). In addition to words that are directly imitative of sounds, many languages also contain semantically broader inventories of ideophones. These words comprise a grammatically and phonologically distinct class of words that are used to express various sensory-rich meanings, such as qualities related to manner of motion, visual properties, textures and touch, inner feelings and cognitive states (29, 32, 33). Notably, these words are often recognized by native speakers to bear a degree of resemblance to their meaning, an intuition that is confirmed by experiments with naïve listeners (34).

Therefore, if we are to understand the ongoing evolution of spoken languages, it is critical to examine how words are formed from vocal imitation. Here we show that the transition from imitation to word can be a simple process: the mere act of repeated imitation can drive vocalizations to become more word-like in both form and function. Notably, as onomatopoeia and ideophones of natural languages maintain a resemblance to the quality they represent, so did our vocal imitations retain a resemblance to the original sound that inspired them. Altogether, our findings show how words might be created from the repetition of vocal imitations of an original sound.

Materials and Methods

Selecting seed sounds. We selected inanimate categories of sounds because they were less likely to have lexicalized onomatopoeic forms already in English, and they were assumed to be less familiar and more difficult to imitate. Using an odd-one-out norming procedure (N=105 participants), an initial set of 36 sounds in 6 categories was reduced to a final set of 16 "seed" sounds: 4 sounds in each of 4 categories. The four final categories included: water, glass, tear, zipper. The results of the norming procedure are presented in the Supporting Information.

Collecting imitations. Participants (N=94) were paid to participate in an online version of the children's game of "Telephone". The instructions informed participants that they would hear some sound and their task is to reproduce it as accurately as possible using their computer microphone. Full instructions are provided in the Supporting Information. Participants listened to and imitated 4 sounds, receiving one sound from each of the four categories of sounds drawn at random such that participants were unlikely to hear the same person more than once. Recordings that were too quiet (less than –30 dBFS) were not allowed. Imitations were monitored by an experimenter to catch any gross errors in recording before they were heard by the next generation of imitators. For example, recordings were trimmed to the length of the imitation, and recordings with loud sounds in the background were removed. The experimenter also blocked sounds that violated the rules of the experiment, e.g., by saying something in English. A total of 115 imitations were removed.

Measuring acoustic similarity. Acoustic similarity was measured by having research assistants listen to pairs of sounds and rate their subjective similarity. On each trial, raters heard two sounds from subsequent generations were played in succession but in random order. Then they rated the similarity between the sounds on a 7-point scale. They were instructed that a 7 on this scale meant the sounds were nearly identical, whereas a 1 meant the sounds were entirely different and would never be confused. Raters were encouraged to use as much of the scale as they could while maximizing the likelihood that, if they did this procedure again, they would reach the same judgments. Full instructions are provided in the Supporting Information. Ratings were normalized prior to analysis (z-scores).

Collecting transcriptions of imitations. Participants (N=216) were paid to transcribe sounds into words in an online survey. They listened to imitations and were instructed to write down what they heard as a single word so that the written word would sound as much like the message as possible. Instructions are provided in the Supporting Information.

Imitations were drawn at random from the first and last three generations of all imitations collected in the Telephone game. As a control, we also had participants "transcribe" words directly from listening to the environmental seed sounds. Transcriptions from participants who failed a catch trial were excluded (N=2), leaving 2163 transcriptions for analysis. Of these, 179 transcriptions were removed because they contained English words, which was a violation of the instructions of the experiment.

Learning transcriptions as category labels. Our transmission chain design and subsequent transcription procedure created 2110 novel words. From these, we sampled words transcribed from first and last generation imitations as well as from seed sounds that were equated in length and overall matching accuracy. Specifically, we removed transcriptions that contained less than 3 unique characters and transcriptions that were over 10 characters long. Of the remaining transcriptions, a sample of 56 were selected to have approximately equal means and variances of overall matching accuracy. The procedure for sampling the words in this experiment is linked in the Supporting Information.

Participants (N=67) were randomly assigned four novel names to learn for four categories of environmental sounds. Participants were assigned between-subject to learn words from first or last generation imitations, as well as words from transcriptions of seed sounds as a control. They learned the referents for these names in a trial-and-error category learning experiment. On each trial, participants heard one of the 16 seed sounds and then saw a word–one of the transcriptions of the imitations. They responded yes or no using a gamepad as to whether the sound and the word went together. Initially they were forced to guess, but because they received feedback on their performance, over trials they learned the names of the categories. 4 outlier participants were excluded from the final sample due to high error rates and slow reaction times.

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

Matching imitations to seeds. Participants (N=751) were paid to complete an online survey containing 4AFC questions. For each question in the survey, participants listened to an imitation and guessed which of four possible sounds they thought the person was trying to imitate. No feedback was provided.

Question types (True seed, Category match, Specific match) were assigned between-subject. Participants in the True seed and Category match conditions were provided four seed sounds from different categories as choices in each question. Participants in the Specific match condition were provided four seed sounds from the same category. All 365 imitations were tested in each of the three conditions.

Matching transcriptions to seeds. Participants (N=468) completed a modified version of the "Guess the seed" game. Instead of listening to imitations, participants now read a word (a transcription of an imitation), which they were told was an invented word. They were instructed that the word was invented to describe one of the four presented sounds, and they had to guess which one. Of all the unique transcriptions that were collected for each sound (imitations and seed sounds), only the top four most frequent transcriptions were used in the matching experiment. 6 participants failed a catch trial and were excluded, leaving 461 participants in the final sample.

ACKNOWLEDGMENTS. The authors acknowledge Jesse Reid, Hailey Schiedermayer, Zoe Hansen, Maggie Parker, and Yacong Wu for conducting the acoustic similarity analysis.

1. Plato, Reeve CDC (1999) *Cratylus*. (Hackett, Indianapolis).
2. Locke J (1948) An essay concerning human understanding in *Readings in the history of psychology*, ed. Dennis W. (Norwalk, CT).
3. Klima Edward S, Bellugi U (1980) *The signs of language*. (Harvard University Press).
4. Arbib MA (2012) *How the brain got language: The mirror system hypothesis*. (Oxford University Press) Vol. 16.
5. Armstrong DF, Stokoe WC, Wilcox SE (1995) *Gesture and the nature of language*. (Cambridge University Press).
6. Corballis MC (2003) *From hand to mouth: The origins of language*. (Princeton University Press).
7. Hockett CF (1978) In search of Jove's brow. *American speech* 53(4):243–313.
8. Tomasello M (2010) *Origins of human communication*. (MIT press).
9. Pinker S, Jackendoff R (2005) The faculty of language: what's special about it? *Cognition* 95(2):201–236.
10. Goldin-Meadow S (2016) What the hands can tell us about language emergence. *Psychonomic Bulletin & Review* pp. 1–6.
11. Kendon A (2014) Semiotic diversity in utterance production and the concept of 'language'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651):20130293–20130293.
12. Dingemanse M, Blasi DE, Lupyan G, Christiansen MH, Monaghan P (2015) Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences* 19(10):603–615.
13. Perniss P, Thompson RL, Vigliocco G (2010) Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology* 1.
14. Clark HH, Gerrig RJ (1990) Quotations as demonstrations. *Language*.
15. Lewis J (2009) As well as words: Congo Pygmy hunting, mimicry, and play in *The cradle of language*. (The cradle of language).
16. Brown RW, Black AH, Horowitz AE (1955) Phonetic symbolism in natural languages. *Journal of abnormal psychology* 50(3):388–393.
17. Donald M (2016) Key cognitive preconditions for the evolution of language. *Psychonomic Bulletin & Review* pp. 1–5.
18. Imai M, Kita S (2014) The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651):20130298–20130298.
19. Perlman M, Dale R, Lupyan G (2015) Iconicity can ground the creation of vocal symbols. *Royal Society Open Science* 2(8):150152–16.
20. Dingemanse M (2014) Making new ideophones in Siwu: Creative depiction in conversation. *Pragmatics and Society*.
21. Lemaitre G, Rocchesso D (2014) On the effectiveness of vocal imitations and verbal descriptions of sounds. *The Journal of the Acoustical Society of America* 135(2):862–873.
22. Lemaitre G, Houix O, Voisin F, Misdariis N, Susini P (2016) Vocal Imitations of Non-Vocal Sounds. *PloS one* 11(12):e0168167–28.
23. Rhodes R (1994) Aural images. *Sound symbolism* pp. 276–292.

24. Lupyan G, Thompson-Schill SL (2012) The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General* 141(1):170–186.
25. Edmiston P, Lupyan G (2015) What makes words special? Words as unmotivated cues. *Cognition* 143(C):93–100.
26. Boutonnet B, Lupyan G (2015) Words Jump-Start Vision: A Label Advantage in Object Recognition. *Journal of Neuroscience* 35(25):9329–9335.
27. Crystal D (year?) *The Cambridge Encyclopedia of Language*. (Cambridge Univ Press) Vol. 2.
28. Newmeyer FJ (1992) Iconicity and generative grammar. *Language*.
29. Dingemanse M (2012) Advances in the Cross-Linguistic Study of Ideophones. *Language and Linguistics Compass* 6(10):654–672.
30. Vigliocco G, Perniss P, Vinson D (2014) Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651):20130292–20130292.
31. Sobkowiak W (1990) On the phonostatistics of English onomatopoeia. *Studia Anglica Posnaniensia* 23:15–30.
32. Nuckolls JB (1999) The case for sound symbolism. *Annual Review of Anthropology* 28(1):225–252.
33. Voeltz FE, Kilian-Hatz C (2001) *Ideophones*. (John Benjamins Publishing) Vol. 44.
34. Dingemanse M, Schuerman W, Reinisch E (2016) What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*.