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 March 31, 2017

We report the results of a large-scale (*N* = 1571) experiment to investigate whether spoken words can emerge from the process of repeated imitation. Participants played a version of the children's game "Telephone". The first generation was asked to imitate recognizable environmental sounds (e.g., glass breaking, water splashing); subsequent generations imitated the imitators for a total of 8 generations. We then examined whether the vocal imitations became more stable and word-like, retained a resemblance to the original sound, and became more suitable as learned category labels. The results showed the imitations became progressively more word-like, even after 8 generations, they could be matched above chance to the category of environmental sound that motivated them, and imitations from later generations were more effective as learned category labels. These results show how repeated imitation can create progressively more word-like forms while retaining a semblance of iconicity.

language evolution | 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. Some theories of language evolution have hypothesized that vocal imitation played an important role in generating the first words of spoken languages (e.g., 3–6); early humans may originally have referred to a predatory cat by imitating its roar, or to the discovery of a stream by imitating the sound of rushing water. Such vocal imitation might have served to clarify the referent of a vocalization and eventually establish a mutually understood word. In this study, we investigate the formation of onomatopoeic words: imitative words that resemble the sounds to which they refer. We ask whether onomatopoeic words can be formed gradually and without instruction simply from repeating the same imitation over generations of speakers.

Onomatopoeic words appear to be a universal lexical category found across the world's languages (7). Languages all have words for animal vocalizations and various environmental sounds that are conventional, but at the same time, exhibit an imitative quality. (8), for example, documented a repertoire of over 100 onomatopoeic words in English, which he notes exist along a continuum from "wild" to "tame". People often use more wild vocal imitations and other sound effects during demonstrative discourse, especially when producing quotations (9, 10). Wild words have a more imitative phonology whereas tame words take on more standard phonology of other words in the language. In some cases, words that begin as wild imitations of sounds become fully lexicalized and integrated into the broader linguistic system, when they behave like more "ordinary" words that can undergo typical morphological processes. Examples are English words like "crack" or the recently adapted "ping".

However, not all researchers agree that vocal imitation has any significant role in language. For instance, (11) suggested that, "Humans are not notably talented at vocal imitation in general, only at imitating speech sounds (and perhaps melodies). For example, 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." Nevertheless, experiments show that people can actually be quite effective at using vocal imitation. For example, (12) collected imitations and verbal descriptions of various mechanical and synthesized sounds. When participants listened to these and were asked to identify the source, they were more accurate with imitations than descriptions. A subsequent study found that vocal imitations tend to focus on a few salient features of the sound rather than a high fidelity representation, which aids identification of the source (13).

Thus humans can be effective at communicating with vocal imitation, it can play an important role in narration and discourse, and it appears to be the basis for substantial inventories of sound-imitative vocabulary across languages. But little is known about the process by which onomatopoeic words like "crack" and "ping" are actually formed and integrated into the vocabulary of a language. A basic question is whether word formation requires deliberation and an intention to create a new word, or whether words can originate from one-shot vocal imitations and repetition. Here we examine whether the repetition of imitations of environmental sounds is sufficient to create more word-like vocalizations, even without an intent to communicate.

To test this, we recruited participants to engage in a large scale 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 was

Significance Statement

Although many words spoken today appear to have imitative origins, the process by which nonverbal imitations might give rise to wordlike forms has not been documented. Here we report evidence of imitations gradually becoming more wordlike through a process of unguided repetition across different individuals. In addition to becoming more stable both in terms of acoustic and orthographic properties, the novel words created through this process retain an iconic resemblence to the originating sound. These results demonstrate a pervasive tendency for establishing convention in human vocal communication.

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

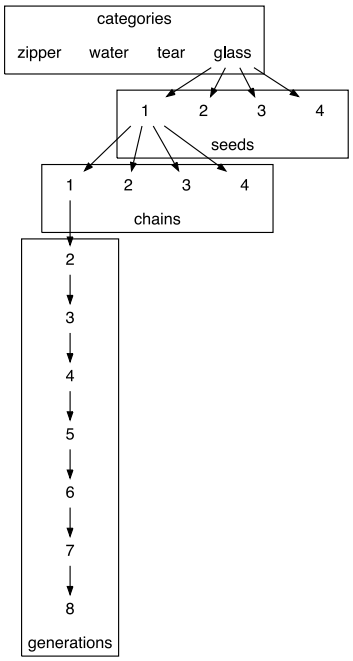

Fig. 1. The design of the transmission chain experiment. 16 seed sounds were selected, four in each category of environmental sound. Participants imitated each seed sound, and then the next generation of participants imitated the imitations and so on for 8 generations.

a recording of an environmental sound. The first generation participant imitated the sound, the next generation imitated the previous imitation, and so on for a maximum of 8 generations. After obtaining these imitations, we investigated how the imitations changed over generations to determine whether they became more word-like. We investigated the acoustic properties of the imitations as well as their orthographic properties once transcribed into English words. We find that by both measures the imitations become more stable through repetition. In addition to stability, we also find that the imitations can still be matched back to the original sounds at above chance levels for many generations. Finally, we measured how quickly the invented words are learned as category labels in a category learning experiment, and find that later generation imitations are easier to learn as category labels.

In Experiment 1 we collected iterated vocal imitations using the transmission chain design depicted in Fig. 1. In the remaining experiments we assessed changes in these imitations over generations. The extent to which each imitation could be matched back to its originating sound was measured in Experiment 2. Experiment 3 involved collecting transcriptions of imitations, and in Experiment 4 these transcriptions were matched back to the original sounds. In Experiment 5 we used transcriptions taken from first and last generation imitations as novel category labels in a simple category learning experiment.

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 repetition. As the imitations were repeated, they gradually lost their resemblance to the source sound. In particular, they lost information that distinguished the source sound from within-category competitors more readily than higher-level category information. This result suggests that through repetition and stabilization the imitations became better abstract category labels by virtue of cueing all category members equally as opposed to highlighting individual category members. We found support for this conclusion in the transcriptions of the imitations. Later generations of imitations were transcribed with better agreement, suggesting that imitations were indeed stabilizing on invented words that were increasingly distinctive and broadly recognizable. Still, these invented words retained some resemblance to the category of environmental sound that motivated them (at least relative to the other categories tested in this experiment). Participants were able to accurately match the transcriptions of final generation imitations in each transmission chain back to the category of environmental sounds that motivated them. Unlike the direct matching of imitations, the extent to which transcriptions were matched to individual source sounds as opposed to categories of sounds did not increase over generations. However, when transcriptions of first and last generation imitations were learned as novel labels of environmental sound categories, last generation transcriptions were easier to learn than those from the first generation. These results describe a process by which an imitation of an environmental sound may transition to a more word-like form through unguided repetition, and suggest that such a transition to more word-like forms might make them more effective as category labels.

Imitations stabilized over generations. We collected a total of 480 imitations from 94 participants in a study conducted online. 115 imitations were removed for bad audio quality or violating the rules of the experiment (e.g., saying something in English), leaving 365 imitations along 105 contiguous transmission chains for analysis.

Trained 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, $p < 0.001$. 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. 2). This result suggests that imitations may be stabilizing on particular acoustic forms through repetition.

We also calculated automated analyses of imitation fidelity using Mel Frequency Cepstral Coefficients (MFCCs) as a measure of acoustic distance. However, 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 that the automated analyses do 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. We therefore report the results of these automated analyses in the Supporting Information.

Imitations retained seed category information. Were the imitations stabilizing on arbitrary acoustic forms or were they maintaining some aspect of the original environmental sound? To test this, we measured the ability of participants naive to the design of the Telephone game to match each imitation back

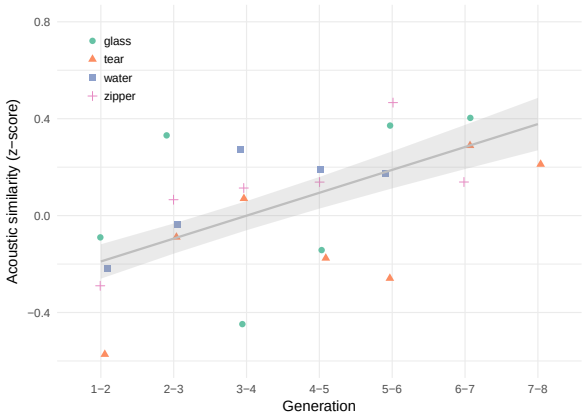

Fig. 2. Change in acoustic similarity over generations of repetition. Points show mean acoustic similarity ratings for imitations in each category of environmental sounds. The line shows the linear predictions of a hierarchical model with random effects for rater and category, with error bands designating +/- 1 standard error of the model predictions. The results show that acoustic similarity increases over generations, indicating that subsequent imitations become more similar to one another through repetition.

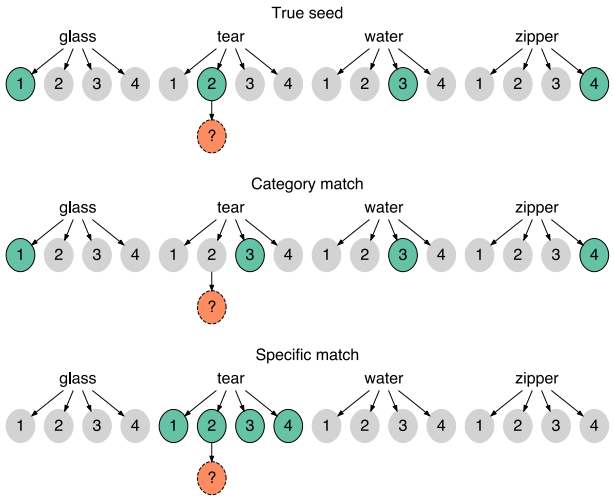

Fig. 3. Types of 4AFC matching questions depicted in relation to the original set of 16 seed sounds. For each question, participants listened to an imitation (orange dashed circle) and had to guess which of 4 sound choices (green solid circles) they thought the person was trying to imitate. (Top) True seed questions contained the actual sound that generated the imitation in the choices, and the three distractor sounds were sampled from different categories. (Middle) 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. (Bottom) Specific match questions pitted the actual seed against the other seeds within the same category.

to its original source relative to other seed sounds from either the same category or from different categories (Fig. 3). All 365 imitations were tested in the three conditions depicted in Fig. 3. 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. 4. 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. One 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. 3); the advantage of having the true seed among between-category distractors decreased over generations, b = -0.07 (0.02) log-odds, z = -2.77, $p = 0.006$.

These results indicate that as imitations are repeated they lose within-category information more rapidly than between-category information. Later generation imitations were just as likely to be recognized as identifiers of an entire category of environmental sounds as they were of particular sounds within the category.

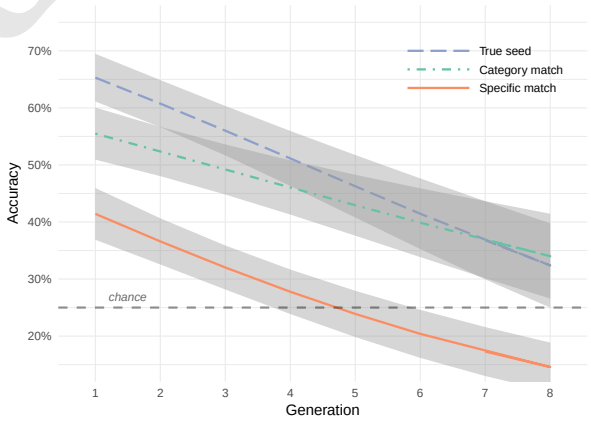

Fig. 4. Changes in matching accuracy over generations. Matching accuracy is the ability to guess the sound most likely to have generated the imitation relative to other seed sounds used in the experiment. Performance is separated by question type which describes the relationship between the imitation and the choices in the 4AFC task (see Fig. 3). Lines show predictions from a generalized linear mixed effects model along 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. These results suggest that imitions lose within-category information more rapidly than between-category information.

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

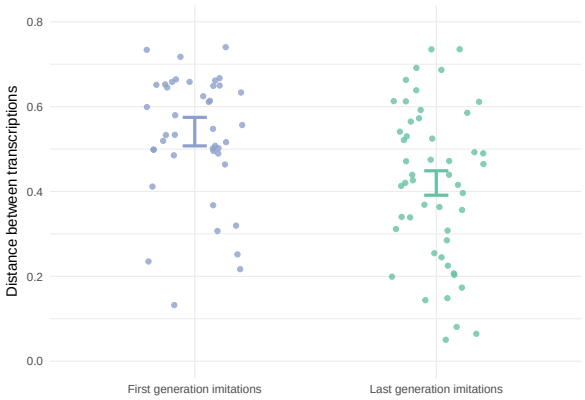

Fig. 5. Average orthographic distance among transcriptions of imitations taken from first and last generations. Each point shows the average orthographic distance between the most frequent transcription and all other transcriptions of a single imitation. Error bars are +/- 1 standard error of the linear mixed effects model predictions. Transcriptions of later generation imitations were more similar to one another than transcriptions of first generation imitations.

Transcription agreement increased over generations. We next tested whether the imitations became more clearly distinguishable as particular words, as opposed to non-linguistic, i.e., non-English sounds. We had English-speaking participants 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. 216 participants generated a total of 2163 or approximately 20 transcriptions per sound (imitation and seed sounds). Transcriptions containing actual English words and those from participants who failed a catch question were excluded from analysis (`n_transcriptions_dropped`).

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. Hierarchical linear models were fit 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.

Transcriptions of later generation imitations were more similar to one another in terms of orthographic distance than transcriptions from earlier generations, b = -0.12 (0.03), $t(3.0)$ = -3.62, p = 0.035 (Fig. 5). This result supports our hypothesis that unguided repetition drives imitations to become more distinctive as particular English words. The same conclusion was reached from alternative measures of orthographic distance, including exact string matches and excluding those imitations for which all transcriptions were unique.

Transcriptions retained seed category information. We previously demonstrated that people were able to accurately guess the source of an imitation after 8 repetitions, but what about the source of a transcription of an imitation? Do these invented words still resemble the category of sounds that was originally imitated? We tested the top 4 most frequent transcriptions for each imitation in a modified version of the "Guess the Seed" game (see Fig. 3). Participants were given a novel 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. 6). 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. This indicates that transcriptions of imitations may capture idiosyncratic elements of specific category members more than the imitations themselves. Possible reasons for this asymmetry between imitations and transcriptions are explored in the Discussion.

Repeated imitations were easier to learn as category labels. 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.

When participants learned some of the transcriptions as novel category labels for categories of environmental sounds, they were faster when the label came from transcriptions of later generation imitations than from transcriptions of first generation imitations, b = -114.13 (52.06), $t(39.9)$ = -2.19, p = 0.034 (Fig. 7A). 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), $t(1732.0)$ =

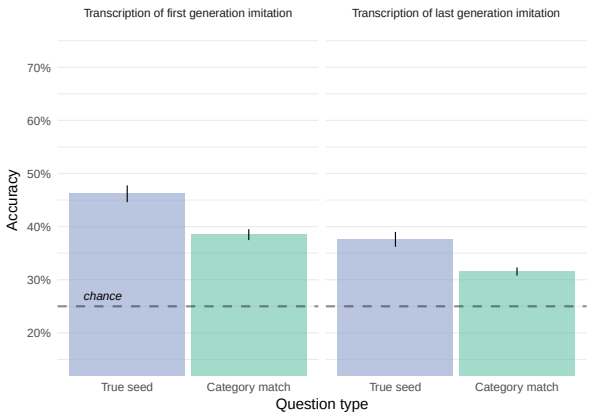

Fig. 6. Matching accuracy for transcriptions of imitations taken from first and last generations. Bars represent the predicted means taken from the generalized linear mixed effects model with +/- 1 standard error of the model predictions. True seed questions contained transcriptions of the actual seed generating the transcribed word. Category match questions contained transcriptions of imitations of other seeds from the same category. See Fig. 3 for a more detailed description of these question types. The results show that even after 8 generations of repetition, imitations can be transcribed into words and matched back to the category of sounds motivating the original imitation.

-2.30, $p = 0.022$ (Fig. 7B). This result suggests that learning transcriptions from later generation imitations were easier to generalize to new category members.

Discussion

Imitative words for sounds, i.e., onomatopoeia, are found across the languages of the world (7). However, little is known about how these words are formed and incorporated into the lexicon of a language. To examine this process, we conducted a large-scale, online, iterated vocal imitation experiment–essentially a version of the children's game of "Telephone". The first generation of participants imitated environmental sounds, and then the next generation of participants imitated these imitations, and so on. Our results show that through simple repetition, that is, without any intention to communicate, the imitations gradually became more word-like. Over generations, they became more stable in sound, and also more easily transcribed into specific English orthographic forms. However, at the same time, the imitations maintained an onomatopoeic quality: listeners were able to match the vocalizations to both their original sound, and to the sound category, even after eight generations. Even when the imitations were transcribed into English, participants were still able to guess the categorical origin of the word above chance relative to the other categories tested in this experiment.

The imitations also become more word-like in that they served as more effective category labels. Information that distinguished an imitation from other sound categories was more resilient to transmission decay than exemplar information within a category. Previous research has found that words, as opposed to more veridical cues, make categorization easier (14, 15). Similarly, we found that 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 evidence completes the transition from vocal imitation to abstract word and demonstrates the impact of this transition on communication.

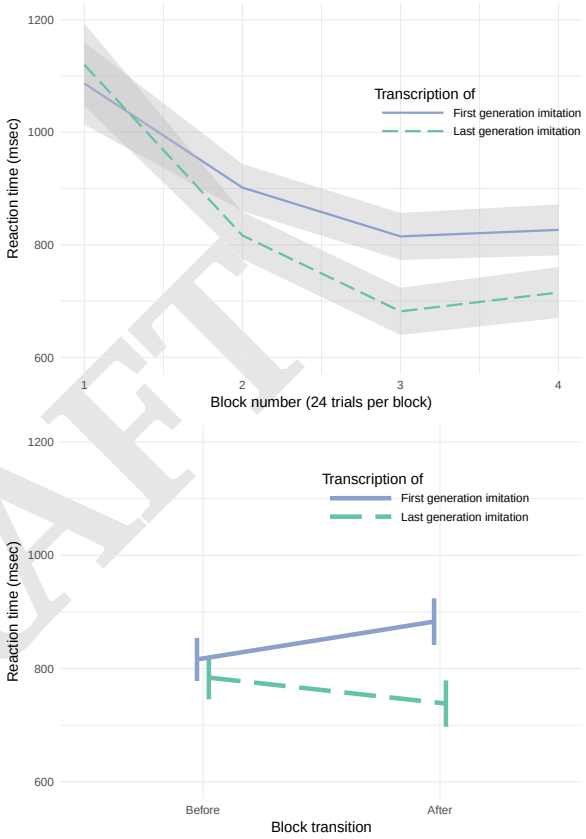

Fig. 7. (Top) RTs on correct trials by block, showing faster responses when learning category labels transcribed from last generation imitations. (Bottom) RTs on trials leading up to and immediately following the block transition where new category members are introduced.

One result that did not fit squarely with imitations becoming more word-like is that with transcriptions, individuating information was retained over generations more so than category information. If the results of matching transcriptions back to seed sounds would have perfectly mirrored the results of matching imitations back to seed sounds we would have expected the difference between True seed questions and Category match questions to decrease over generations for transcriptions as for imitations. Instead we found a main effect of question type. Although participants were able to match transcriptions to categories of sounds even after 8 generations of repetition, it was still easier for them to match a transcription to the actual seed that generated the transcription.

Our study focused on the formation of onomatopoeia–sound-imitative words–but in addition to onomatopoeia, many languages have semantically rich systems of ideophones. These words comprise a grammatically and phonologically distinct class of words that are used to express a variety of sensory-rich meanings (7, 16). Notably, these words are often recognized by native speakers to be imitative of their meaning to some degree. For example, in Japanese, the word 'koron' with a voiceless [k] refers to a light object rolling once, the reduplicated 'korokoro' to a light object rolling repeatedly, and 'gorogoro' with a voiced [g] to a heavy object rolling repeatedly (5). The iconicity of ideophones was verified by results showing that naïve listeners were able to guess the meanings of words sampled from five different languages (17). Although words for sounds were guessed more accurately than the rest, listeners were better than chance at guessing the synonyms of ideophones that expressed meanings from all five semantic categories tested, which also included color/visual, motion, shape, and texture. In addition, laboratory experiments show that people are able to generate imitative vocalizations for a variety of non-sound concepts, and that these are also understandable to naïve listeners (6). Thus vocal imitation has the potential to play a role in word formation that extends beyond just the imitation of sounds.

Our findings from an online game of Telephone suggest that the formation of words from vocal imitation can be a simple process. The results show how repeated imitation can create progressively more word-like forms while retaining a resemblance to the original sound that motivated it. This raises the possibility that onomatopoeic words can be created from the repetition of one-shot 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 of36 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. 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 played in succession. 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).

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.

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.

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.

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 script that sampled the words in this experiment is linked in the Supporting Information.

Participants ($N = 67$) were randomly assigned four novel names 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. 63 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. At the start of a new block of trials, participants heard four new sounds they hadn't heard before, and had to learn to associate these new sounds with the words they had learned in the previous blocks.

ACKNOWLEDGMENTS.

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. Brown RW, Black AH, Horowitz AE (1955) Phonetic symbolism in natural languages. *Journal of abnormal psychology* 50(3):388–393.
4. Donald M (2016) Key cognitive preconditions for the evolution of language. *Psychonomic Bulletin & Review* pp. 1–5.
5. 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.
6. Perlman M, Dale R, Lupyan G (2015) Iconicity can ground the creation of vocal symbols. *Royal Society Open Science* 2(8):150152–16.
7. Dingemanse M (2012) Advances in the Cross-Linguistic Study of Ideophones. *Language and Linguistics Compass* 6(10):654–672.
8. Rhodes R (1994) Aural images. *Sound symbolism* pp. 276–292.
9. Blackwell NL, Perlman M, Tree JEF (2015) Quotation as a multimodal construction. *Journal of Pragmatics* 81:1–7.
10. Clark HH, Gerrig RJ (1990) Quotations as Demonstrations. *Language, Cognition, and Neuroscience* 66(4):764–805.
11. Pinker S, Jackendoff R (2005) The faculty of language: what's special about it? *Cognition* 95(2):201–236.
12. 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.
13. Lemaitre G, Houix O, Voisin F, Misdariis N, Susini P (2016) Vocal Imitations of Non-Vocal Sounds. *PloS one* 11(12):e0168167–28.
14. 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.
15. Edmiston P, Lupyan G (2015) What makes words special? Words as unmotivated cues. *Cognition* 143(C):93–100.
16. Voeltz FE, Kilian-Hatz C (2001) *Ideophones*. (John Benjamins Publishing) Vol. 44.
17. Dingemanse M, Schuerman W, Reinisch E (2016) What sound symbolism can and cannot do: testing the iconicity of ideophones from five languages. *Science*.