

Creating words from iterated vocal imitation

Pierce Edmiston^{a,1}, Marcus Perlman^b, and Gary Lupyan^{a,b}

^aDepartment of Psychology, University of Wisconsin-Madison, 1202 W. Johnson St., Madison, WI 53703; ^bMax Planck Institute for Psycholinguistics, Nijmegen, Netherlands

This manuscript was compiled on January 29, 2017

New words are constantly being invented, yet we have never witnessed the emergence of spoken words in the absence of prior convention. We report the results of a large-scale ($N = 1402$) experiment designed to investigate whether spoken words can emerge solely from the process of repeated imitation. Participants played a version of the children's game "Telephone" where a message is passed from person to person with the goal of keeping it unchanged. The first generation of players heard recognizable environmental sounds (e.g., glass breaking, water splashing) and had to imitate the sounds. The next generation imitated the imitations, and so on for a maximum of 8 generations ($N = 365$ imitations). We subsequently assessed whether the imitations (1) became more stable and word-like, (2) retained an iconic resemblance to the original sound relative to the other seed sounds, and (3) the extent to which the imitations from different generations were suitable as verbal category labels. The results showed that vocal imitations became progressively more word-like as measured by acoustic similarity and agreement of orthographic transcriptions, that even after 8 generations they could be matched above chance to the environmental sound that motivated them, and that imitations from later generations were more effective as learned category labels. These results show how repeated imitation, even in the absence of an explicit intention to communicate, can create progressively more word-like forms while retaining a semblance of iconicity.

Keyword 1 | Keyword 2 | Keyword 3 | ...

Over the ages, people have pondered the origins of language and especially the words that compose them (Plato, Bible). Where do words come from? Some theories of language evolution hypothesize that vocal imitation played an important role in generating the first words in the evolution of spoken languages (1–3). 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 (4). Languages all have conventional words for animal vocalizations and other environmental sounds. Such words appear to be especially prominent for very young language learners (5, 6). In some cases, words that begin as 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. (7) show this on a variable instance-by-instance basis rather than a monotonic historical process towards integration. (8) observed that onomatopoeic words exist along a continuum from "wild" to "tame" in which wild words have a more imitative phonology whereas tame words take on more standard phonology of other English words. In English, words like "crack" or the recently adapted "ping".

In addition to iconic words for sounds, many languages have a semantically broader system of ideophones. (9) found

that naïve listeners were better than chance at guessing the meanings of ideophones (sampled from five different languages) from five different semantic categories: color/visual, motion, shape, sound, and texture, but sound words were guessed more accurately than the rest. Experiments suggest that iconic vocalizations for non-sound concepts can also become more word-like over repeated use (3).

But the process by which onomatopoeic words may be born of nonverbal imitations has yet to be observed in a controlled lab environment. Do repeated nonverbal imitations become more wordlike even without any instruction to do so? Or alternatively does the limited fidelity of human vocal imitation restrict such stability from forming in the absence of a reason to communicate. To test this, we recruited participants to engage in a large scale online version of the children's game of "Telephone" in which an acoustic message is passed from one person to the next. After obtaining these imitations, we then investigated how the imitations changed over generations. We investigated the acoustic properties of the imitations as well as the orthographic properties of the imitations once transcribed into English words. Finally, we test how quickly these invented words are learned as category labels in a category learning experiment.

Results

Collecting iterated imitations. The design of the transmission chain experiment is shown in Fig. 1. We started with 16 seed sounds, four from each of four categories of environmental sounds (see Methods for seed selection procedures). Four participants imitated each environmental sound, starting unique transmission chains that were allowed to continue for a maximum of 8 generations. We removed all low quality recordings and recordings where the speaker violated the rules of the experiment, e.g., by recording something in English, result-

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 resemblance 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. carried out the research and analyzed the results. P.E., M.P., and G.L. wrote the manuscript.

The authors declare no conflicts of interest.

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

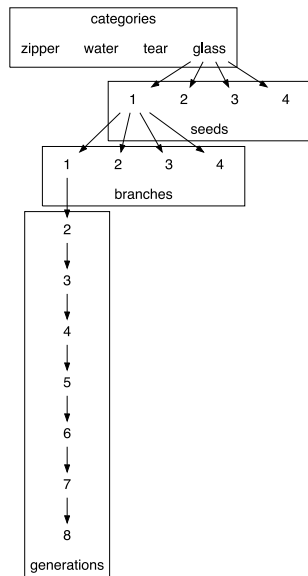


Fig. 1. The design of the transmission chain experiment. 16 seed sounds were used, 4 in each category of nonverbal environmental sounds. Each seed sound was imitated by 4 different participants, resulting in 4 branches off of each seed sound. Subsequent participants imitated the imitations and so on for a maximum of 8 generations.

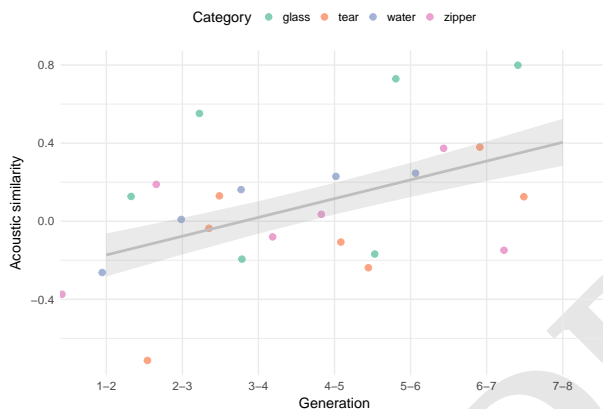


Fig. 2. Increase in acoustic similarity over generations.

ing in a final sample of 365 imitations along 105 contiguous transmission chains.

Imitations from later generations were rated as being more similar to one another than imitations from earlier generations, $b = 0.10$ (0.03), $t = 3.51$ (**Fig. 2**). Given the nonverbal nature of the imitations, large differences between speakers (e.g., speakers of different genders), and variability in recording equipment from conducting the experiment online, previously published techniques for calculating acoustic distance were inadequate (10). Instead, we obtained subjective measures of acoustic similarity using a controlled, randomized norming procedure (see Methods). We found that imitations from later generations were rated as being more similar to one another than imitations from earlier generations, suggesting that the acoustic signals were becoming more stable through repetition.

Matching imitations to original sounds. Next we measured the extent to which imitations taken from different generations could be matched back to their original source relative to the other sounds used as seeds. Participants ($N = 751$) completed

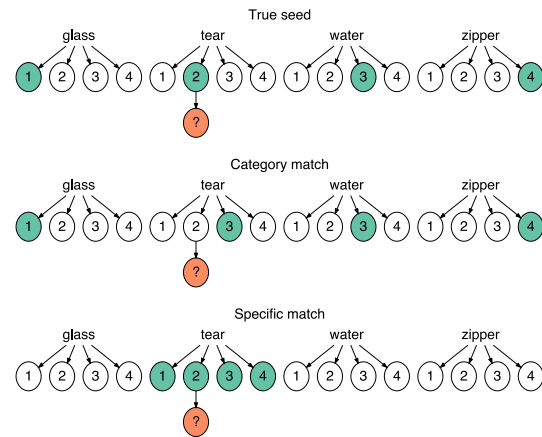


Fig. 3. Three types of matching questions depicted in relation to the original set of 16 seed sounds. For each question, participants listened to a sample imitation (orange circle) and had to guess which of 4 sound choices (green circles) they thought the person was trying to imitate. (A) True seed questions contained the actual seed that generated the imitation in the choices, and the distractor seeds were sampled from different categories. (B) Category match questions also used distractor sounds from different categories but the correct seed was not the actual seed, but a different sound within the same category. (C) Specific match questions pitted the actual seed against the other seeds within the same category.

a series of four alternative forced choice (4AFC) questions where they listened to an imitation drawn at random and 4 options strategically drawn from among the set of 16 seed messages (**Fig. 3**). We measured above chance matching performance in three distinct question types, all concerning the relationship between the imitation and the options in the question.

Matching accuracy for all question types was above chance for first generation imitations, $b = 1.65$ (0.14) log-odds, odds = 0.50, $z = 11.58$, $p = 0.00$, and decreased over generations, $b = -0.16$ (0.04) log-odds, $z = -3.72$, $p = 0.00$. We tested whether this decay in match accuracy was constant across the three question types or if some question types were more or less resistant to generational decay than others. In particular we hypothesized that if the imitations were becoming more word-like and therefore categorical as they were repeated, then performance on questions where category information enabled a correct response will be more resilient to generational decay.

The results are shown in **Fig. 4**. The first evidence in support of our hypothesis comes in comparing performance on questions requiring a category match to performance on questions where guessing correctly required distinguishing the true seed from other sounds within the same category. Performance decreased over generations more rapidly for these specific match questions than for category match questions, $b = -0.05$ (0.02) log-odds, $z = -2.53$, $p = 0.01$, suggesting that category information was more resistant to information loss through transmission. A simpler explanation for this result is that the specific match questions are simply harder than the category match questions. However, performance also decreased more rapidly for the easiest type of question where the correct answer was the actual seed generating the imitation. The advantage for having the correct seed among the options decreased over generations, $b = -0.07$ (0.02) log-odds, $z = -2.83$, $p = 0.00$. These results indicate that later generation imitations were more likely to be recognized as identifiers of a particular category than they were of particular exemplars

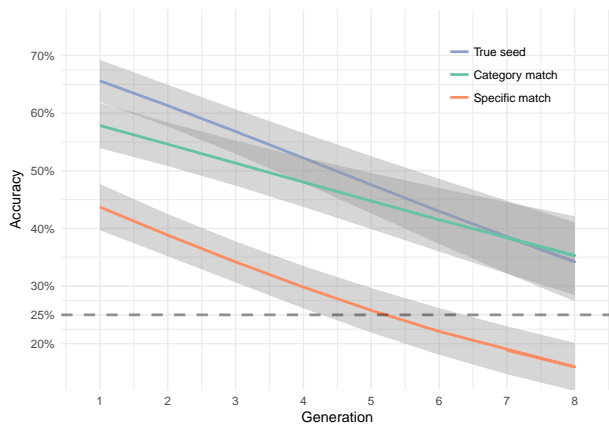


Fig. 4. Matching accuracy for guessing the seed from imitations taken from different generations of the transmission chain experiment. Lines denote different question types concerning the relationship between the imitation and the options in the question (see Fig. 3). The advantage of having a True seed among the options (True seed versus Category match) decreased over generations, and the advantage of having distractors from different categories (Category match versus Specific match) increased over generations. These results suggest that category information was more resilient to generational decay than specific information about the particular exemplar being imitated.

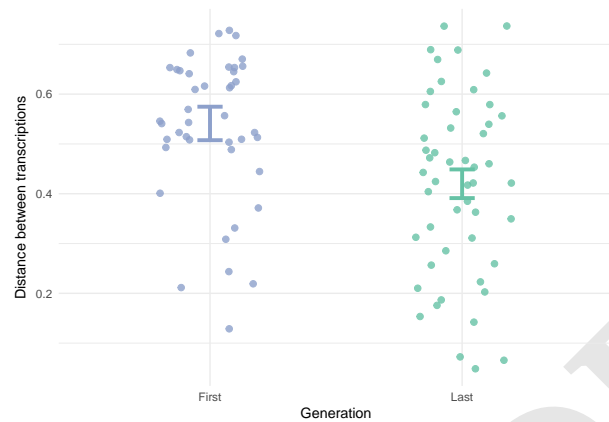


Fig. 5. Average orthographic distance among transcriptions of imitations taken from first and last generations.

within each category.

Matching transcriptions of imitations. An additional measure of wordlike-ness is the stability of the auditory signal once transcribed into an orthographic form. We next investigated how the orthographic forms of the imitations changed over generations by having participants ($N = 218$) transcribe a set of first and last generation imitations from each transmission chain. A total of 106 imitations were transcribed, with approximately 21 transcriptions per imitation.

Analyzing changes in orthographic agreement over generations paralleled what was observed in the analysis of acoustic similarity: Transcriptions from later generation imitations were more similar to one another in terms of orthographic distance than transcriptions from earlier generations, $b = -0.12$ (0.03), $t = -3.62$ (Fig. 5). This result supports the conclusion that the imitations were becoming more stable in both acoustic and orthographic forms.

To investigate whether these invented words retained any iconic resemblance to the original seed sounds, we selected

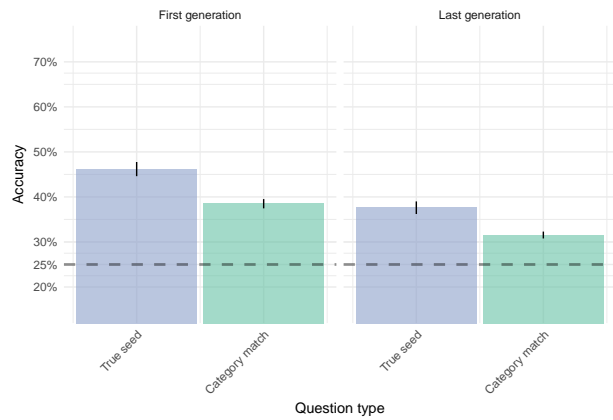


Fig. 6. Matching accuracy for transcriptions of imitations taken from first and last generations. True seed questions contained transcriptions of the actual seed generating the transcribed word. Category match questions contained transcriptions of imitations of other seeds from the same category.

the top 4 most frequent transcriptions for each imitation that was transcribed and presented them in a modified version of the “Guess the seed” matching game ($N = 461$). On each trial, participants were presented with a random transcription generated from one of the imitations, and tasked with guessing the meaning of the invented word from among 4 possible sounds. Participants were able to guess the correct meaning of the transcribed word above chance even after between 5 and 8 generations of repetition, $b = 0.83$ (0.13) log-odds, odds = -0.18 , $z = 6.46$, $p = 0.00$ (Fig. 6). This was true both for true seed questions, $b = 0.75$ (0.15) log-odds, odds = -0.28 , $z = 4.87$, $p = 0.00$, and for category match questions, $b = 1.02$ (0.16) log-odds, odds = 0.02 , $z = 6.39$, $p = 0.00$. The effect of generation did not vary across these question types, $b = 0.05$ (0.10) log-odds, $z = 0.47$, $p = 0.64$.

The result that matching accuracy was above chance even for category match questions is especially informative. This finding indicates that participants were able to match a transcription back to a seed sound from the same category as the one that eventually generated the transcription, but not from the exact chain of imitations leading to the transcription. If transcriptions only captured idiosyncratic elements of the imitations, then the only way participants would be able to guess the “meaning” of the transcription at above chance levels would be through its surface resemblance to the specific sound being transcribed, in which case performance in category match questions would be indistinguishable from chance. Instead we found that the transcription—the result of an imitation passed through generations of different speakers—still retained some iconic resemblance to the category of sounds being imitated, and thus participants were able to use this category information to match the transcribed word to the category of sounds at above chance levels. This result supports our hypothesis that through repetition, imitations are becoming more wordlike.

Using transcriptions as novel category labels. Although we found evidence that the imitations were becoming more stable over generations and that they still retained some iconic resemblance to the original sound, we have not directly observed an advantage to the more wordlike imitations emerging through iterated repetition compared to direct imitations of the source

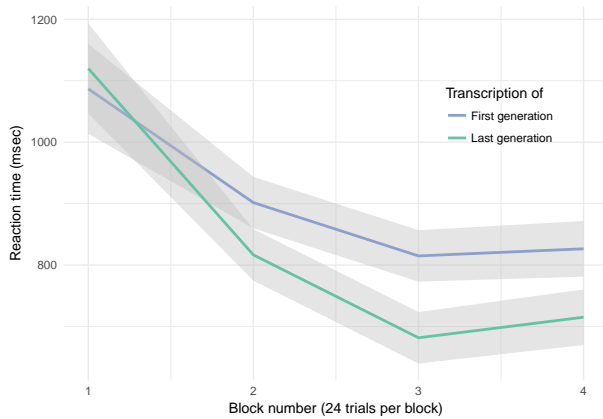


Fig. 7. Response times across blocks in the category learning experiment. Each new block introduced four new sounds into the categories participants were learning the names of. Participants who learned names that were actually transcriptions from the end of the transmission chain branches were faster to generalize these labels to new category members than participants who learned names transcribed directly from imitations of the source sounds.

of the sound. The advantage of word cues as opposed to environmental cues to a particular category is that words activate more categorical representations useful in generalizing across idiosyncratic differences between category members (11). On the view that repeated imitation creates more wordlike forms, we predicted that the words emerging through repeated imitation should be easier to learn as category labels requiring generalizing to new category members than words that were transcribed from direct imitations.

To test this, we recruited participants ($N = 63$) to learn a novel category label for each of the 4 categories of environmental sounds used in the transmission chain experiment. When participants had to generalize the meaning of the novel label to new category members (new sounds), they were faster when the label came from later generations than when the words were transcribed from first generation imitations, $b = -114.13$ (52.06), $t = -2.19$. This suggests that in addition to becoming more stable both in terms of acoustic and orthographic properties, imitations that have been more repeated are also easier to learn as category labels, completing the transition from nonverbal imitation to a fully lexicalized word form and demonstrating the impact of this transition on communication.

Discussion

We show that repeated imitation of an originally imitative vocalization works to integrate it into the ambient language.

Consequently, it becomes more abstract in its resemblance to its initial source sound, but maintains some resemblance to the more abstract category of sounds. The vocalizations become more learnable as linguistic labels for categories.

Repeated drawings from memory (without delay) result in drawings that become more conventional, moving towards cultural attractors such as letters (12). Also Bartlett.

More broadly, imitative words are hypothesized to play an important role in the evolution of language (1–3).

Materials and Methods

Selecting seed sounds. We initially selected a set of 36 sounds in 6 different categories of environmental sounds. Inanimate categories of sounds were selected because they were less likely to have lexicalized onomatopoeic forms already in English, and they were assumed to be less familiar and harder to imitate. We reduced this initial set to the final sample of 16 seeds using an odd-one-out norming procedure ($N = 105$ MTurk participants) designed to remove any sounds that were outliers within each category. The results of this norming procedure and selection criteria are provided in the Supplemental Materials.

Collecting imitations. We collected the imitations online.

ACKNOWLEDGMENTS. Please include your acknowledgments here, set in a single paragraph. Please do not include any acknowledgments in the Supporting Information, or anywhere else in the manuscript.

1. Brown RW, Black AH, Horowitz AE (1955) Phonetic symbolism in natural languages. *Journal of abnormal psychology* 50(3):388–393.
2. 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.
3. Perlman M, Dale R, Lupyan G (2015) Iconicity can ground the creation of vocal symbols. *Royal Society Open Science* 2(8):150152–16.
4. Dingemans M (2012) Advances in the Cross-Linguistic Study of Ideophones. *Language and Linguistics Compass* 6(10):654–672.
5. Laing CE, Vihman M, Keren-Portnoy T (2016) How salient are onomatopoeia in the early input? A prosodic analysis of infant-directed speech. *Journal of child language* pp. 1–23.
6. Tardif T, et al. (2008) Baby's first 10 words. *Developmental Psychology* 44(4):929–938.
7. Dingemans M, Akita K (2016) An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics* pp. 1–32.
8. Rhodes R (1994) Aural images. *Sound symbolism* pp. 276–292.
9. Dingemans M, Schuerman W, Reinisch E (2016) What sound symbolism can and cannot do: testing the iconicity of ideophones from five languages. *Science*.
10. Lemaitre G, Houix O, Voisin F, Misdariis N, Susini P (2016) Vocal Imitations of Non-Vocal Sounds. *PloS one* 11(12):e0168167–28.
11. Edmiston P, Lupyan G (2015) What makes words special? Words as unmotivated cues. *Cognition* 143(C):93–100.
12. Tamariz M, Kirby S (2014) Culture: Copying, Compression, and Conventionality. *Cognitive Science* 39(1):171–183.