Creating words from iterated vocal imitation

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 (1) the imitations became progressively more word-like, (2) even after 8 generations, could be matched above chance to the environmental sound that motivated them, and (3) imitations from later generations were more effective as learned category labels. These results show how repeated imitation can create progressively word-like forms while retaining a semblance of iconicity.

Over the ages, people have pondered the origins of languages and especially the words that compose them. For example, both Plato in his *Cratylus* dialogue (Plato & Reeve 1999) and John Locke in his *Essay Concerning Human Understanding* (Locke 1990) 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., Brown, Black, and Horowitz 1955; Donald 2016; Imai and Kita 2014; Perlman, Dale, and Lupyan 2015). For instance, early humans might 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. Thus 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 (Dingemanse 2012). Languages all have conventional words for animal vocalizations and various environmental sounds. , for example, a repertoire of over 100 onomatopoeic words in English, which he notes, . WIn 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". (Dingemanse and Akita 2016) found that the degree of expressiveness of an imitative word is inversely related to its degree of morphosyntactic integration.

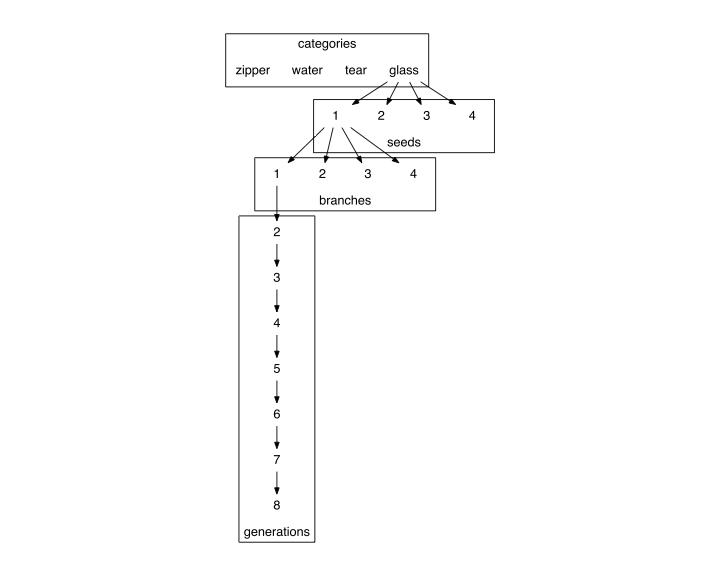
People often use more wild vocal imitations and other sound effects during demonstrative discourse, especially when producing quotations (Blackwell, Perlman, & Fox Tree 2015; Clark & Gerrig 1990). Vocal imitation might be especially prominent in some communities. For example, an ethnographic study of Mbendjele Pygmies living in the Congo found that Mbendjele speakers frequently use vocal imitation when narrating dramatic events (Lewis 2009). For example, when describing an event such as an encounter with a dangerous animal in the jungle, they use vocal imitation to emphasize acoustic details of the event, including the sound of the animal and also other inanimate details like the thrashing of trees.

However, not all researchers agree that vocal imitation has any significant role in language. For instance, Pinker and Jackendoff (2005) 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, (Lemaitre & Rocchesso 2014) 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 imitation 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 (Lemaitre et al. 2016).

Thus vocal imitation appears to play an important role in human communication, and it appears to be the basis for substantial inventories of sound-imitative vocabulary across languages. But the process by which onomatopoeic words may be born of nonverbal imitations has yet to be observed in a controlled lab environment. Here we examine whether simple repeated imitations become more wordlike even without any instruction to do so. Or alternatively does the limited fidelity of human vocal imitation restrict the formation of stable words 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 to determine whether they became more wordlike. 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.

# Methods

The design of the transmission chain experiment is shown in Fig. 1. We started with 16 seed sounds, four from each category of environmental sounds (seed selection procedures are described below). 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, resulting in a final sample of 365 imitations along 105 contiguous transmission chains. The remaining experiments (see Table 1) assessed the stability of these imitations in both acoustic and orthographic forms as well as measuring the extent to which imitations and transcriptions could be matched back to the original seed sounds.



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.

## Materials

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 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.

## Participants

In Experiments 1-4, participants were recruited via Amazon Mechanical Turk and paid to participate. Participants in Experiment 5 were University of Wisconsin-Madison undergraduates who received course credit in exchange for participation.

## Procedures

The procedures for each experiment listed in Table 1 are described below.

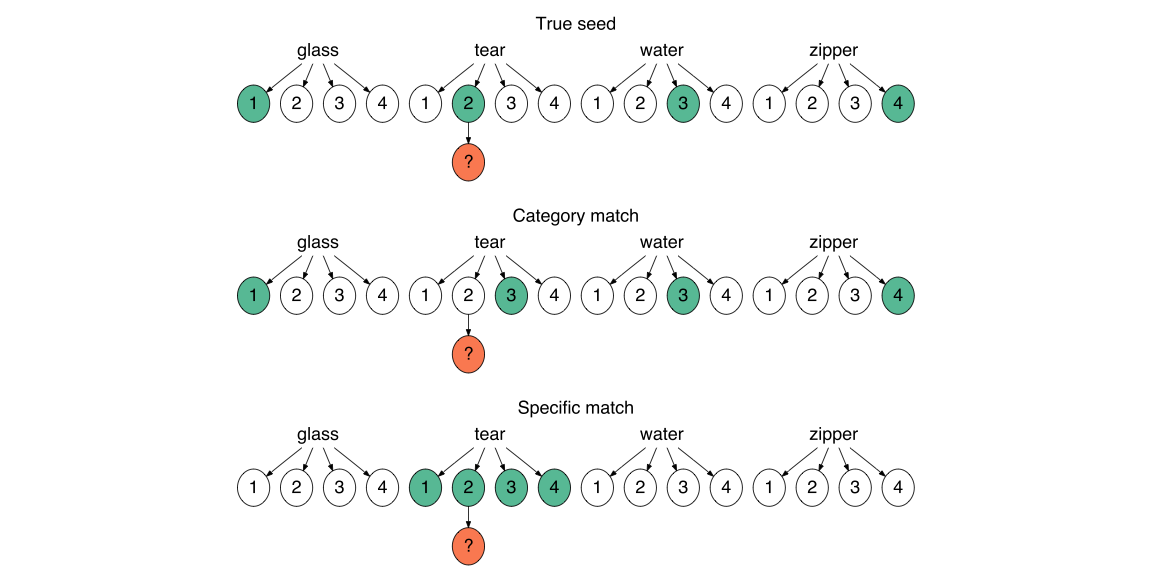
### Collecting vocal imitations

Participants were paid to participate in an online version of the children's game of "Telephone". The instructions stated that they would be imitating nonverbal sounds and that they should try to reproduce the message as accurately as possible. Participants in this online study listened to 4 messages and recorded 4 imitations from their homes or offices using their personal recording equipment (usually a built-in microphone). Participants received one message from each of the four categories of sounds drawn at random such that participants were unlikely to hear the same person more than once.

Imitations were monitored as they were received by an experimenter. The purpose of monitoring was to catch any gross errors in recording before they were passed on to the next generation of imitators. For example, most sounds required trimming the recording more closely to the vocalization--removing mouse clicks and other ambient sounds--before it was suitable for the next generation. Sounds were also prevented from being reproduced if the imitator violated the rules of the experiment, e.g., by saying something in English.

### Matching imitations to seeds

Each participant was assigned (between-subject) 4 different seed sounds as choices in a 4AFC task. On each trial, they listened to an imitation, and decided which of the 4 seed sounds they thought the imitation most closely resembled. They did not receive any feedback on their performance. We tested three types of matching questions that differed according to the relationship between the imitation and the four seed sounds serving as the options in the 4AFC task (Fig. 2).



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. (Top) True seed questions contained the actual seed that generated the imitation in the choices, and the distrator seeds were sampled from different categories. (Middle) 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. (Bottom) Specific match questions pitted the actual seed against the other seeds within the same category.

### Collecting transcriptions of imitations

Participants transcribed between 9 and 20 imitations each. They were instructed to write down what they heard as a word so that the written word would sound as much like the message as possible. We selected only the first and final three imitations in each transmission chain to be transcribed. Only the transcriptions from participants who passed a catch question were included. All transcriptions containing actual English words were excluded from analysis.

### Matching transcriptions to seeds

Participants completed a modified version of the 4AFC described in **Matching imitations to seeds**, above. Instead of listening to imitations, participants now read 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 four presented sounds, and they had to guess which one. We contained the size of the experiment by only using the 4 most frequent transcriptions for each imitation.

### Using transcriptions as category labels

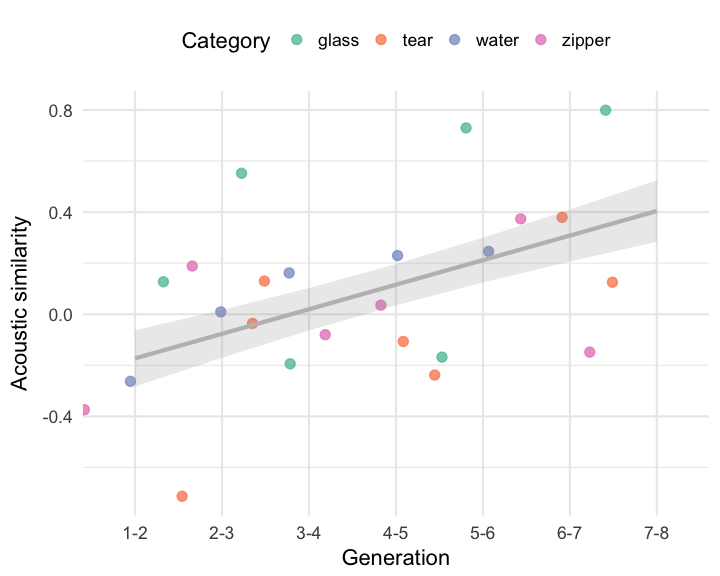
Participants learned, through trial-and-error, the names for four different categories of sounds. On each trial participants listened to one of the 16 environmental sounds used as seeds and then read a novel word--a transcription of one of the imitations. Participants responded by pressing a green button if the label was the correct label and a red button otherwise. The experiment was divided into blocks so that participants had repeated exposure to each sound and the labels multiple times within a block. At the transition to a new block, participants received four new sounds that they had not heard before, and had to associate these sounds with the same novel labels from the previous blocks.

To determine which transcriptions to test in the experiment, we first selected only those transcriptions which had above chance matching performance when matching back to the original seeds. Then we excluded transcriptions that had less than 2 unique characters, and transcriptions that were over 10 characters long. From this set of eligible transcriptions, we sampled from both first and last generation imitations to reach a final set that controlled for overall matching accuracy.

# Results

## Changes in acoustic similarity

We found that imitations from later generations were more similar to one another than imitations from earlier generations (Fig. 2). Given large differences in recording quality resulting from conducting the experiment online, previously published techniques for calculating acoustic distance were inadequate (cf. Lemaitre et al. 2016). Instead, we obtained subjective measures of acoustic similarity using a controlled, randomized norming procedure completed by research assistants. Five RAs listened to pairs of imitations while blind to generation and rated their similarity on a 7-point scale. We found that 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.



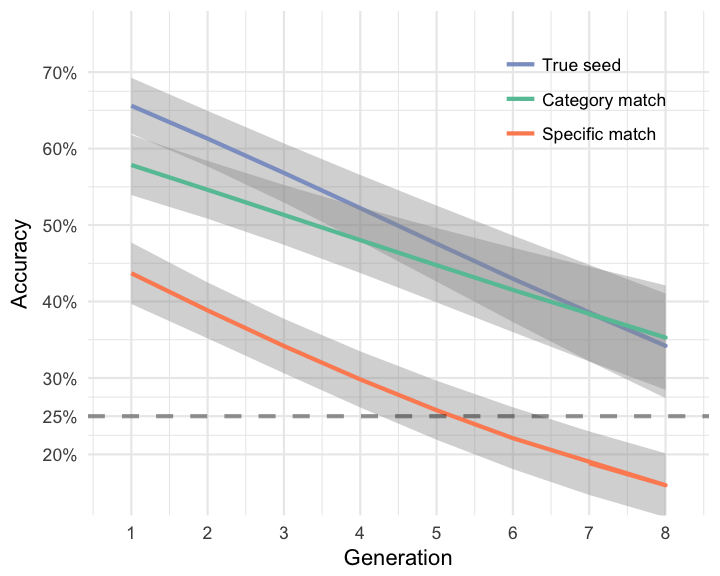
Increase in acoustic similarity over generations.

## Matching imitations to seeds

Matching accuracy for all question types (see Fig. 2) 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 increase in question difficulty was constant across the three types of questions or if some question types became more difficult at later generations. In particular we hypothesized that if the imitations were becoming more like category labels as they were repeated, then performance on questions where category information enabled a correct response would 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 resistent to loss through transmission.

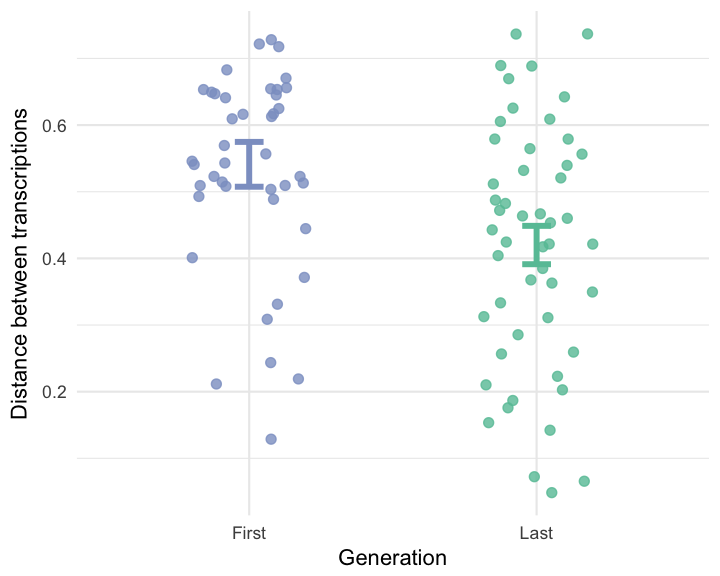
One explanation for this result is that the specific match questions are simply harder than the category match questions. However, performance also decreased more rapidly for the easiest type of question where the correct answer was the actual seed generating the imitation. The advantage for having the true 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 within each category.



Accuracy in matching imitations back to seed sounds. Lines denote different question types concerning the relationship between the imitation and the options in the question (see Fig. 2). 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.

## Changes in orthographic agreement

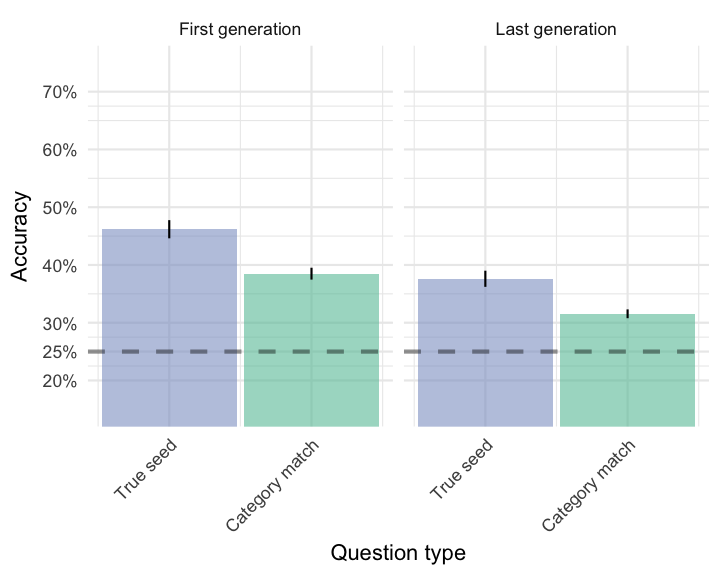
We next investigated how agreement among orthographic forms of the imitations changed over generations. A total of 106 imitations were transcribed, with approximately 21 transcriptions per imitation. Analyzing changes in orthographic agreement over generations paralleled what was observed in the analysis 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 our hypothesis that the imitations were becoming more stable in both acoustic and orthographic forms.



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

## Matching transcriptions to seeds

To investigate whether the invented words retained any resemblance to the original seed sounds, we selected the top 4 most frequent transcriptions for each imitation and presented them in a modified version of the matching game. 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.

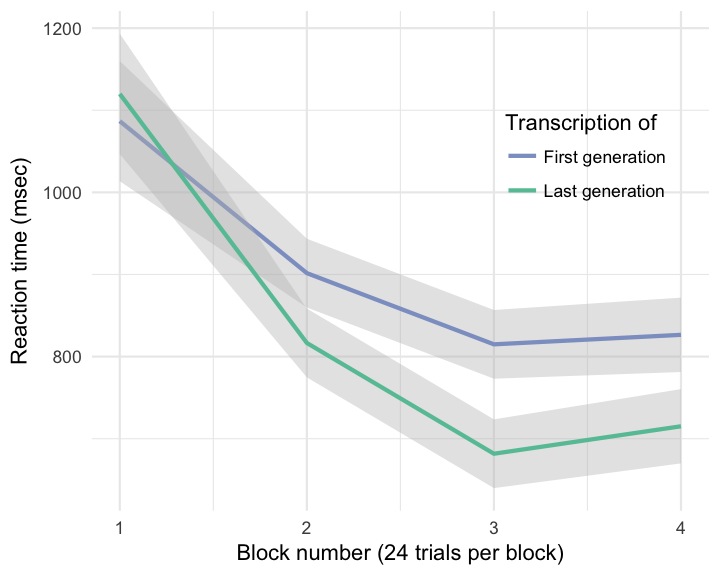


Matching accuracy for transcriptions of imitations taken from first and last generations. True seed questions contained transcriptions of the actual seed generating the transcribed word. Category match questions contained transcriptions of imitations of other seeds from the same category.

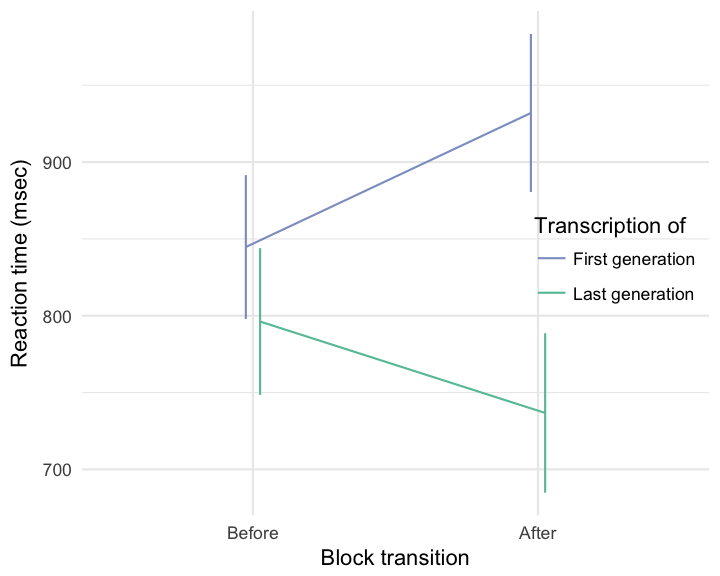
## Using transcriptions as novel category labels

Last, we examined whether there was a learning advantage to the more wordlike imitations emerging through iterated repetition compared to direct imitations of the source of the sound. An 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 (Edmiston and Lupyan 2015). Therefore, we hypothesized that transcriptions of the more wordlike forms emerging through repeated imitation should be easier to generalize to new category members than transcriptions from direct imitations.

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 transcriptions of later generation imitations than from transcriptions of first generation imitations, *b* = -114.13 (52.06), *t* = -2.19. The effect can be further localizing 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* = -146.75 (65.47), *t* = -2.24. This suggests that in addition to becoming more stable both in terms of acoustic and orthographic properties, imitations that have been more repeated may also be easier to learn as category labels.



Response times across blocks in the category learning experiment. Participants learned the labels for sounds in each category. Each new block introduced four new sounds into the categories. Participants who learned labels that were transcriptions from later generation imitation in the transmission chain were faster to generalize these labels to new category members than participants who learned labels transcribed directly from imitations of the source sounds.



Cost to generalizing to new category members at block transition.

# 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.

Need to summarize the specific findings?

One result that did not fit squarely with imitations becoming more word-like is that with transcriptions, there was no difference over generations between question types.

completing the transition from nonverbal imitation to a fully lexicalized word form and demonstrating the impact of this transition on communication

Our study focused on the formation of onomatopoeia -- sound-imitative words – but studies suggest that many words outside of the domain of sound have an imitative or *iconic* element. For example, many languages of the world have large inventories of ideophones. In addition to iconic words for sounds, many languages have a semantically broader system of ideophones. These words comprise a distinctively iconic, grammatical class of words that are used to express a variety of sensory-rich meanings (Dingemanse, 2012; Voeltz & Kilian-Hatz, 2001). (Dingemanse, Schuerman, and Reinisch 2016) 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 (Perlman, Dale, and Lupyan 2015).

More broadly, imitative words are hypothesized to play an important role in the evolution of language (Brown, Black, and Horowitz 1955; Imai and Kita 2014; Perlman, Dale, and Lupyan 2015).

# References

Blackwell, Natalia L., Perlman, Marcus, and Fox Tree, Jean E. 2015. Quotation as a multimodal construction. *Journal of Pragmatics, 81,* 1-7.

Brown, R W, A H Black, and A E Horowitz. 1955. “Phonetic symbolism in natural languages.” *Journal of Abnormal Psychology* 50 (3): 388–93. <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=14381156&retmode=ref&cmd=prlinks>.

Clark, Herbert H. and Gerrig, Richard J. 1990. Quotations as demonstrations. *Language, 66,* 764-805.

Dingemanse, Mark. 2012. “Advances in the Cross-Linguistic Study of Ideophones.” *Language and Linguistics Compass* 6 (10): 654–72. doi:[10.1002/lnc3.361](https://doi.org/10.1002/lnc3.361).

Dingemanse, Mark, and Kimi Akita. 2016. “An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese.” *Journal of Linguistics*, October. University of Wisconsin-Madison Libraries, 1–32. doi:[10.1017/S002222671600030X](https://doi.org/10.1017/S002222671600030X).

Dingemanse, Mark, W Schuerman, and E Reinisch. 2016. “What sound symbolism can and cannot do: testing the iconicity of ideophones from five languages.” *Science*. <http://pubman.mpdl.mpg.de/pubman/faces/viewItemOverviewPage.jsp?itemId=escidoc:2286810>.

Donald, Merlin. 2016. “Key cognitive preconditions for the evolution of language.” *Psychonomic Bulletin & Review*, June. Psychonomic Bulletin & Review, 1–5. doi:[10.3758/s13423-016-1102-x](https://doi.org/10.3758/s13423-016-1102-x).

Edmiston, Pierce, and Gary Lupyan. 2015. “What makes words special? Words as unmotivated cues.” *Cognition* 143 (C). Elsevier B.V.: 93–100. doi:[10.1016/j.cognition.2015.06.008](https://doi.org/10.1016/j.cognition.2015.06.008).

Imai, M, and S Kita. 2014. “The sound symbolism bootstrapping hypothesis for language acquisition and language evolution.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1651): 20130298–8. doi:[10.1098/rstb.2013.0298](https://doi.org/10.1098/rstb.2013.0298).

Lemaitre, Guillaume, Olivier Houix, Frédéric Voisin, Nicolas Misdariis, and Patrick Susini. 2016. “Vocal Imitations of Non-Vocal Sounds.” *PloS One* 11 (12): e0168167–28. doi:[10.1371/journal.pone.0168167](https://doi.org/10.1371/journal.pone.0168167).

Lematire, Guillaume and Rocchesso, Davide. On the effectiveness of vocal imitations and verbal descriptions of sounds. *Journal of the Acoustical Society of America, 135,* 862-873.

Lewis, Jerome. 2009. As well as words: Congo Pygmy hunting, mimicry, and play. In Rudolf Botha & Chris Knight (Eds.), *The cradle of language* (pp. 236-256). Oxford, UK: Oxford University Press.

Perlman, Marcus, R Dale, and Gary Lupyan. 2015. “Iconicity can ground the creation of vocal symbols.” *Royal Society Open Science* 2 (8): 150152–16. doi:[10.1098/rsos.150152](https://doi.org/10.1098/rsos.150152).

Pinker, Steven and Jackendoff, Ray. 2005. What’s special about the human language faculty? *Cognition, 95,* 201-236.

Plato, and Reeve, C D C 1999. *Cratylus*. Indianapolis: Hackett.

Rhodes, Richard. 1994. “Aural images.” *Sound Symbolism*. Cambridge University Press: Cambridge, UK, 276–92.