Creating words from iterated vocal imitation

Pierce Edmiston (pedmiston@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Marcus Perlman (marcus.perlman@mpi.nl)
Max Planck Institute for Psycholinguistics
Nijmegen, 6500 AH The Netherlands

Gary Lupyan (lupyan@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Abstract

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, they 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 more word-like forms while retaining a semblance of iconicity.

Keywords: categorization; transmission chain; language evolution

People have long pondered the origins of languages, especially the words that compose them. For example, both Plato in his *Cratylus* dialogue (Plato and Reeve, 1999) and John Locke in his *Essay Concerning Human Understanding* (Locke, 1948) 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 et al., 1955; Donald, 2016; Imai and Kita, 2014; Perlman et al., 2015); 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 through repeated imitation.

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. Rhodes (1994), for example, documented a repertoire of over 100 onomatopoeic

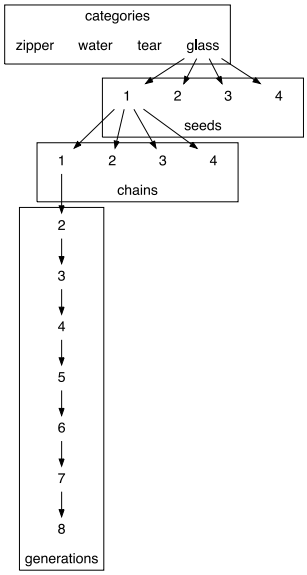

Figure 1: The design of the transmission chain experiment. 16 seed sounds were selected, four in each category of environmental sounds. Participants imitated each seed sound, and then the next generation of participants imitated the imitations and so on for 8 generations.

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 (Blackwell et al., 2015; Clark and Gerrig, 1990). Wild words have a more imitative phonology whereas tame words take on more standard phonology of other English words. 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, 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 be quite effective at using vocal imitation. For example, Lemaitre and Rocchesso (2014) collected imitations and verbal descriptions of 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 (Lemaitre et al., 2016).

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 the process by which onomatopoeic words like "meow", "ping" and "buzz" emerge from vocal imitations has yet to be observed. Here we examine whether simple repeated imitations of environmental sounds become more word-like even in the absence of explicit communication intent or the intent to create a word-like token. Alternatively, repeating imitations might never stabilize on a particular wordform, or the limited fidelity of human vocal imitation may simply restrict the formation of stable words through iterated imitations.

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 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 the 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.

General Methods

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

Table 1: Experiment sample sizes. Participants in Experiments 1-4 were recruited via Amazon Mechanical Turk and paid to participate in an online study. Participants in Experiment 5 were University of Wisconsin-Madison undergraduates who received course credit in exchange for participation.

#	Experiment	N
1	Collecting imitations	94
2	Matching imitations to seeds	752
3	Collecting transcriptions	218
4	Matching transcriptions to seeds	444
5	Category learning	63

Exp 1: Collecting imitations

In Experiment 1 we collected the iterated vocal imitations that served as the basis for the remaining experiments. Our hypothesis was that these vocal imitations would become more stable as they were repeated over generations of speakers.

Methods

We selected inanimate categories of sounds because they were less likely to have lexicalized onomatopoeic forms already in English, and they were less familiar and more difficult to imitate. Nonetheless, it is possible that lexical knowledge still influenced imitation fidelity–a possibility to be explored in future work. The sounds used here were selected using an odd-one-out norming procedure (N=105 participants) to reduce an initial set of 36 sounds in 6 categories to a final set of 16 "seed" sounds: 4 sounds in each of 4 categories. The four final categories included: water, glass, tear, zipper.

Participants 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. Participants listened to and imitated 4 sounds. Participants received 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. Imitations were monitored by an experimenter to catch any gross errors in recording before they were passed on to the next generation of imitators, including blocking sounds that violated the rules of the experiment, e.g., by saying something in English.

Given large differences in recording quality resulting from conducting the experiment online, we were unable to use previously published techniques for calculating acoustic distance (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 where a 1 meant the sounds could never be confused with one another and a 7 meant the sounds were nearly identical.

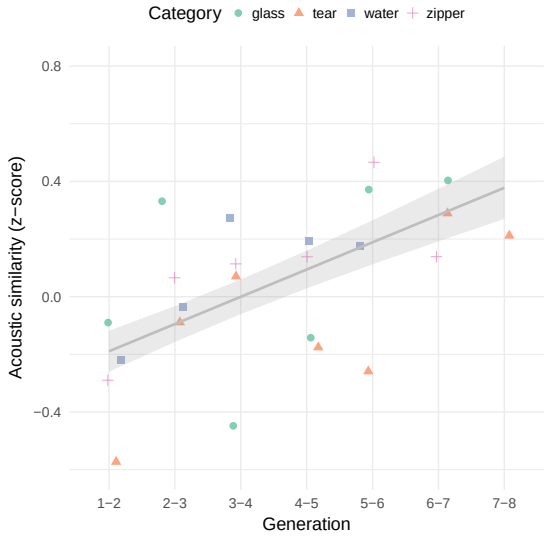

Figure 2: Increase in acoustic similarity over generations. Points depict mean acoustic similarity ratings for imitations in each category of environmental sounds. The predictions of the linear mixed effects model with random effects for rater and category are shown, with error bands denoting +/- 1 standard error of the model predictions.

Results

We collected a total of 480 imitations, of which 115 were removed, leaving 365 imitations along 105 contiguous transmission chains for analysis. Imitations from later generations were rated as being 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), suggesting that the imitations are stabilizing through repetition.

Exp 2: Matching imitations to seeds

Experiment 2 was conducted to determine if the imitations retained some resemblance to the original environmental sound that motivated it (i.e. the seed sound). Participants listened to imitations and guessed which seeds they came from. By varying the relationship between the imitation and the options presented to each participant, we were able to assess the extent to which the imitations retained categorical as opposed to specific, identifying information. On the view that repetition makes the imitations more word-like, we expected later imitations to be better indicators of categories of sounds as opposed to specific sounds within each category.

Methods

All 365 imitations collected in Experiment 1 were tested in each condition depicted in Fig. 3. On each trial participants listened to an imitation and selected among four possible options as to which option sounded the most like the imitation. 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. 3).

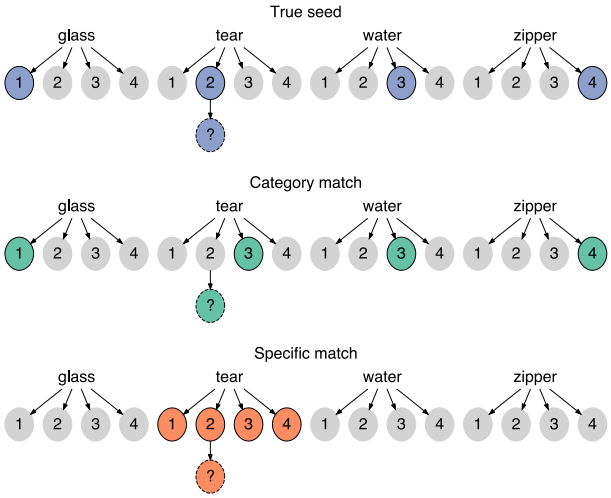

Figure 3: Types of matching questions depicted in relation to the 16 seed sounds. For each question, participants listened to an imitation (dashed circles) and had to guess which of 4 sound choices (solid circles) they thought the person was trying to imitate. (Top) True seed questions contained the actual seed that generated the imitation in the choices, and the distractor 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.

Results

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 question difficulty was constant across the three types of questions or if some question types became more difficult at later generations.

The results are shown in Fig. 4. Performance decreased over generations more rapidly for specific match questions than for category match questions, $b = -0.05$ (0.02) log-odds, $z = -2.53$, $p = 0.012$, suggesting that category information was more resistant 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, $z = -2.83$, $p = 0.005$. 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.

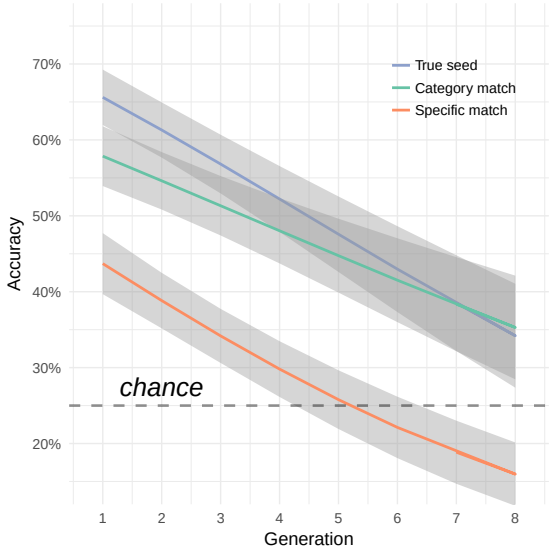

Figure 4: Accuracy in matching imitations back to seed sounds. Performance is separated by question type based on the relationship between the imitation and the options in the question (see Fig. 3). Lines depict predictions from the generalized linear mixed effects model along with +/- 1 standard error of the model predictions.

Exp 3: Collecting transcriptions of imitations

In addition to assessing stability in the acoustic properties of the imitations, we also measured orthographic agreement. If imitations are becoming more wordlike we would expect orthographic agreement to increase over generations.

Methods

We selected the first and final three imitations in each transmission chain to be transcribed into English orthography. Participants were instructed to write down what they heard as a word so that the written word would sound like the message.

Results

We collected a total of 2182 or roughly 21 transcriptions per imitation. All transcriptions containing actual English words were excluded from analysis. Orthographic agreement was measured as the longest contiguous substring match between the most frequent transcription of an imitation and all other transcriptions. 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), $t(3.0)$ = -3.62, p = 0.035 (Fig. 5). This result supports our hypothesis that the imitations were becoming more stable in both acoustic and orthographic forms.

Exp 4: Matching transcriptions to seeds

Experiment 4 tested whether the transcriptions could be matched back to the original seed sounds.

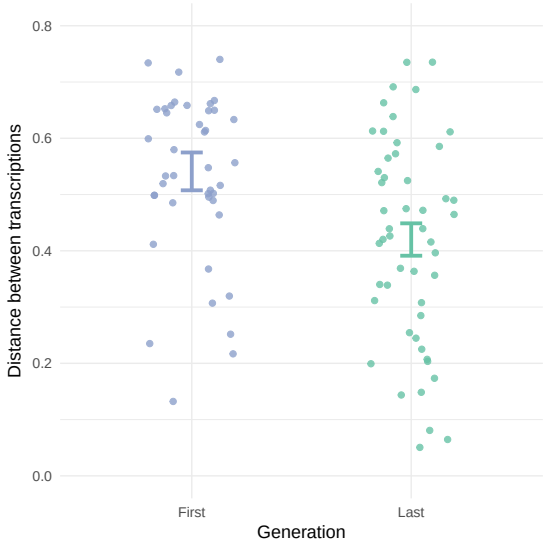

Figure 5: Average orthographic distance among transcriptions of imitations taken from first and last generations. Each point represents the average distance among all transcriptions for a single imitation. Error bars are +/- 1 standard error of the linear mixed effects model predictions.

Methods

The top 4 most frequent transcriptions for each imitation transcribed in Experiment 3 were tested in Experiment 4. Participants completed a modified version of the 4AFC described in Experiment 2. 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. Specific match questions (see Fig. 3) were not collected for transcriptions.

Results

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, b = 0.75 (0.15) log-odds, odds = -0.28, z = 4.87, $p < 0.001$, and for category match questions, b = 1.02 (0.16) log-odds, odds = 0.02, z = 6.39, $p < 0.001$. The effect of generation did not vary across these question types, b = 0.05 (0.10) log-odds, z = 0.47, p = 0.637.

Exp 5: Transcriptions as category labels

In Experiment 5 we examined whether there was a learning advantage to the more word-like imitations emerging through iterated repetition as compared to direct imitations of the source of the sound. We hypothesized that transcriptions of the more word-like forms emerging through repeated imitation should be easier to generalize to new category members than transcriptions from direct imitations.

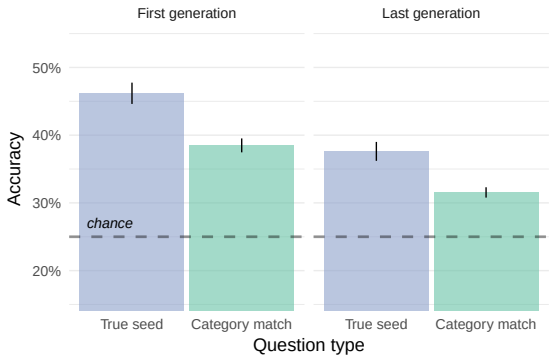

Figure 6: Matching accuracy for transcriptions of imitations taken from first and last generations. True seed questions contained transcriptions of the actual seed generating the transcribed word. Category match questions contained transcriptions of imitations of other seeds from the same category.

Methods

To determine which transcriptions to test as category labels, we selected transcriptions which were matched above chance in Exp. 2. Of these, transcriptions with fewer than two unique characters or more than 10 characters in length were excluded. The final set comprised first and last generation imitations sampled to control for overall matching accuracy.

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 saw 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. They received accuracy feedback after each trial.

The experiment was divided into blocks so that participants had repeated exposure to each sound and the novel labels multiple times within a block. At the start of a new block, participants received four new sounds from the same four categories (e.g., a new zipping sound, a new water-splash sound, etc.) that they had not heard before, and had to associate these sounds with the same novel labels from the previous blocks. The extent to which their performance declined at the start of each block serves as a measure of how well the label they associated with the sound worked as a label for the category.

Results

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), $t(39.9)$ = -2.19, p = 0.034 (Fig. 7A). Accuracy improved over generations but did not significantly differ between groups, $p > 0.05$. 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 (6 trials) revealed a reliable interaction between block transition and the generation of the transcribed label, b = -146.75 (65.47), $t(1869.7)$ = -2.24, p = 0.025 (Fig. 7B). 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 faster to learn as category labels.

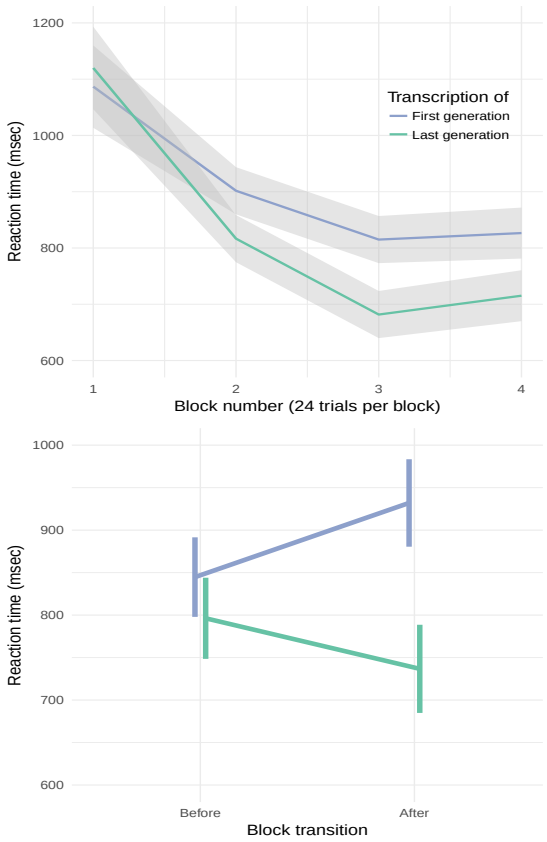

Figure 7: (Top) RTs on correct trials by block, showing faster responses when learning category labels transcribed from last generation imitations. (Bottom) RTs on trials leading up to and immediately following the block transition where new category members are introduced.

Discussion

We show that repeated imitation of an originally imitative vocalization gradually becomes more word-like as it is transmitted along the chain of a "Telephone" game. The first evidence provided showed that imitations became more stable over generations of repetition, both in terms of acoustic similarity as well as in orthographic agreement. But more than just becoming more stable over generations, the imitations also become more word-like in that they served as more effective category labels. Category information was more resilient to transmission decay than specific information identifying a particular exemplar within a category. This category information remained even when the imitations were transcribed into lexical forms, as participants were able to guess the categorical meaning of the word at above chance levels even after 8 generations of repetition. One such consequence of having words is that they make categorization easier. In support of this conclusion, we found that participants were faster and

not less accurate in learning category labels that had emerged through repeated imitation than those who learned from transcriptions of direct imitations of the environmental sounds, completing the transition from nonverbal imitation to a fully lexicalized word form and demonstrating the impact of this transition on communication.

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. If the results of matching transcriptions back to seed sounds would have mirrored the results of matching imitations we would have expected the difference between True seed questions and Category match questions to decrease over generations, but it did not. Although participants were able to match transcriptions to categories of sounds after 8 generations of repetition, it was easier for them to match a transcription to the actual seed that generated the transcription, meaning that individuating information was retained 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. To test this we need to collect match accuracy for transcriptions on Specific match questions to see if transcriptions are able to be matched within-category even when the imitations that generated those transcriptions are not.

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 (Dingemanse, 2012; Voeltz and Kilian-Hatz, 2001). Notably, these words are often recognized by native speakers to be somehow imitative of their meaning. 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 (Imai and Kita, 2014). The iconicity of ideophones was verified by an experiment that tested the ability of nave listeners to guess the meanings of words sampled from five different languages (Dingemanse et al., 2016). 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 – color/visual, motion, shape, sound, 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 nave listeners (Perlman et al., 2015). 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 simply through repeated imitation.

References

Blackwell, N. L., Perlman, M., and Tree, J. E. F. (2015). Quotation as a multimodal construction. *Journal of Pragmatics*, 81:1–7.

Brown, R. W., Black, A. H., and Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *Journal of abnormal psychology*, 50(3):388–393.

Clark, H. H. and Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, 66:764–805.

Dingemanse, M. (2012). Advances in the Cross-Linguistic Study of Ideophones. *Language and Linguistics Compass*, 6(10):654–672.

Dingemanse, M., Schuerman, W., and Reinisch, E. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, 92.

Donald, M. (2016). Key cognitive preconditions for the evolution of language. *Psychonomic Bulletin & Review*, pages 1–5.

Imai, M. and 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).

Lemaitre, G., Houix, O., Voisin, F., Misdariis, N., and Susini, P. (2016). Vocal Imitations of Non-Vocal Sounds. *PloS one*, 11(12):e0168167–28.

Lemaitre, G. and 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.

Locke, J. (1948). An essay concerning human understanding. In Dennis, W., editor, *Readings in the history of psychology*. Norwalk, CT.

Perlman, M., Dale, R., and Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, 2(8):150152–16.

Pinker, S. and Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, 95(2):201–236.

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

Rhodes, R. (1994). Aural images. *Sound symbolism*, pages 276–292.

Voeltz, F. E. and Kilian-Hatz, C. (2001). *Ideophones*, volume 44. John Benjamins Publishing.