**The emergence of words from vocal imitations**

We investigated how conventional spoken words might emerge from imitations of environmental sounds. Participants played a version of the children’s game “Telephone”. The first generation of particiapnts imitated recognizable environmental sounds (e.g., glass breaking, water splashing). Subsequent generations imitated the imitations of the prior generation for a maximum of 8 generations. The results showed that the imitations became more stable and word-like, and more easily learnable as category labels. At the same time, even after 8 generations, both spoken imitations and their written transcriptions could be matched above chance to the category of environmental sound that motivated them. These results show how repeated imitation can create progressively more word-like forms while retaining a semblance of iconicity with the original sound. The results speak to the possible role of human vocal imitation in explaining the origins of spoken words.

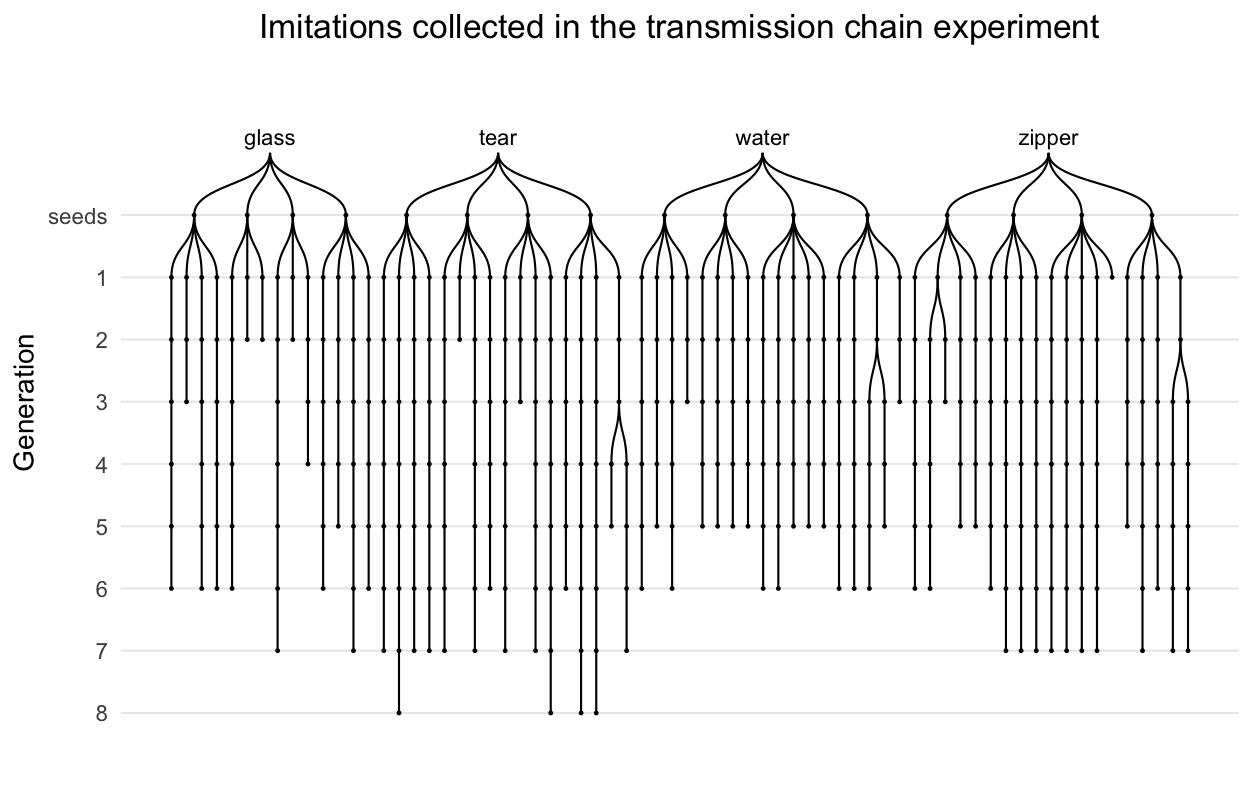
People have long pondered the origins of languages, especially the words that compose them. For example, both Plato in his *Cratylus* dialogue (Plato and Reeve 1999) and John Locke in his *Essay Concerning Human Understanding* (Locke 1948) examined the "naturalness" of words--whether they are somehow imitative of their meaning. Here, we investigated whether new words can be formed from the repetition of non-verbal imitations. Does the repetition of imitations over generations of speakers gradually give rise to novel word forms? In what ways do these words resemble the original sounds that motivated them? We report a large-scale experiment (N=1571) investigating how new words can form--gradually and without instruction--simply by repeating imitations of environmental sounds.

The importance of imitation and depiction in the origin of signs is clearly observable in the origin of words in signed languages (Klima Edward and Bellugi 1980; Goldin-Meadow 2016; Kendon 2014), but in considering the idea that vocal imitation may be key to understanding the origin of spoken words, many have argued that the human capacity for vocal imitation is far too limited to play a role (Arbib 2012; Tomasello 2010; Armstrong and Wilcox 2007; Corballis 2003; Hockett 1978; Hewes 1973). For example, Pinker and Jackendoff (Pinker and Jackendoff 2005) argued that, “most humans lack the ability … to convincingly reproduce environmental sounds … Thus ‘capacity for vocal imitation’ in humans might be better described as a capacity to learn to produce speech” (p. 209). Consequently, it is still widely assumed that vocal imitation--or more broadly, the use of any sort of resemblance between form and meaning--cannot be important to understanding the origin of spoken words.

Although most words of contemporary spoken languages are not clearly imitative in origin, there has been a growing recognition of the importance of imitative words in spoken languages (Dingemanse et al. 2015; Perniss, Thompson, and Vigliocco 2010), and the frequent use of vocal imitation and depiction in spoken discourse (Clark and Gerrig 1990; Lewis 2009). This has led some to argue for the importance of imitation for understanding the origin of spoken words (e.g., Brown, Black, and Horowitz 1955; Donald 2016; Imai and Kita 2014; Perlman, Dale, and Lupyan 2015; Dingemanse 2014). In addition, experiments show that counter to previous assumptions, people are highly effective at using vocal imitations in reference--in some cases, even more effective than with conventional words words (Lemaitre and Rocchesso 2014). Recent work has also shown that people are able to create novel imitative vocalizations for more abstract meanings (e.g. ‘slow’, ‘rough’, ‘good’, ‘many’) that are understandable to naïve listeners (Perlman, Dale, and Lupyan 2015). The effectiveness of these imitations arises not because people can mimic environmental sounds with high fidelity, but because they are able to produce imitations that capture the salient features of sounds in ways that are understandable to listeners (Lemaitre et al. 2016). Similarly, the features of onomatopoeic words might highlight distinctive aspects of the sounds they represent. For example, the initial voiced, plosive /b/ in “boom” represents an abrupt, loud onset, the back vowel /u/ a low pitch, and the nasalized /m/ a slow, muffled decay (Rhodes 1994).

Thus, converging evidence suggests that people can use vocal imitation as an effective means of communication. But how do vocal imitations become standardized words that are integrated into the vocabulary of a language? To investigate this question, we recruited participants to play an online version of the children's game of "Telephone". In the children’s game, a spoken message is whispered from one person to the next. In our version, the original message or seed sound was a recording of an environmental sound. The initial group (first generation) of participants imitated this seed sound, the next generation imitated the previous imitators, and so on for up to 8 generations (Fig. 1).

In subsequent experiments, we systematically answered the following questions about the form of the vocalizations and their potential to function as words. First, does iterated imitation drive the vocalizations to stabilize in form and become more word-like? Second, do the imitations become more suitable as labels for the category of sounds that motivated them? For example, does the imitation of a particular water-splashing sound become, over time, a better label for the more general category of water-splashing sounds? Third, do the imitations retain resemblance to the original environmental sounds that inspired them? If so, it should be possible for naïve participants to match the emergent imitative words back to the original sounds that motivated them.



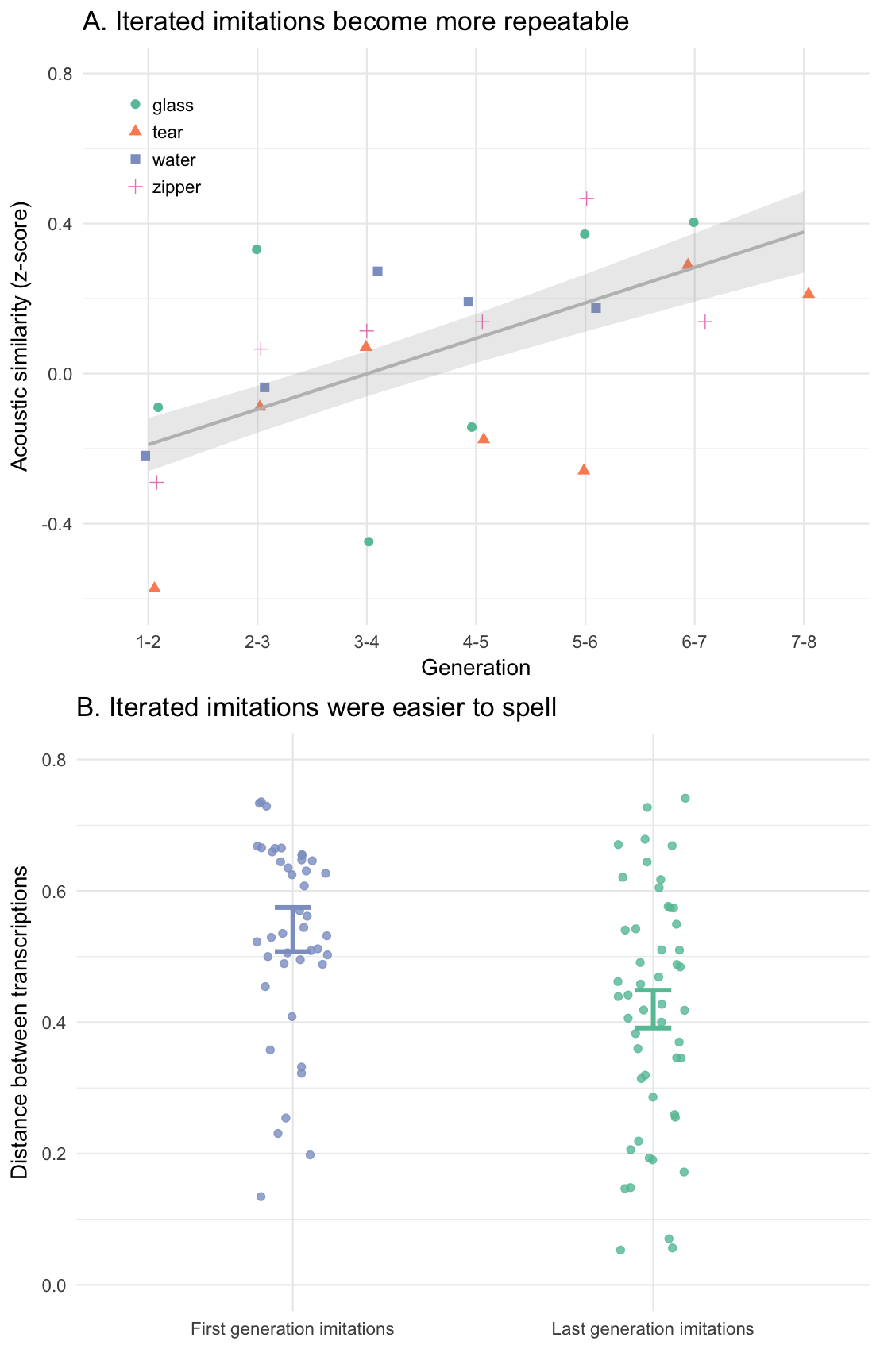
The design of the transmission chain experiment. Seed sounds (16) were sampled from four categories of environmental sounds: glass, tear, water, zipper. Participants imitated each seed sound, and then the next generation of participants imitated the imitations and so on for up to 8 generations.

# Results

We begin with a summary of our main results: (1) Imitations of environmental sounds became more stable over the course of being imitated as revealed by an increasing similarity between generation *n* and *n+*1. In addition, when, later generations of imitations had higher levels of agreement when transcribed into English orthography further suggesting an increase and stability and word-likeness. (2) When transcriptions of first and last generation imitations were used as novel labels for categories of environmental sounds, last generation transcriptions were learned faster and were generalized better to new category members than transcriptions of first generation imitations suggesting that imitating caused the forms to become better suited as category labels. (3) Even as the imitations became more word-like, they also retained resemblance to the category of environmental sound that motivated them as measured by the ability of naïve listeners to match both the auditory imitations and their written transcriptions to the original environmental sounds at levels above chance even after 8 generations of repetition. In sum, our results describe a process by which an imitation of an environmental sound may transition to a more word-like form through unguided repetition. They suggest that such a transition to more word-like forms might make them more effective as category labels. They also demonstrate that these created words are not entirely arbitrary, but instead retain a resemblance to the category of environmental sounds that motivated them.

## Iterated imitations became more stable and word-like

We collected a total of 480 imitations from 94 participants recruited on Amazon Mechanical Turk. The final set included 365 imitations along 105 contiguous transmission chains (Fig. 1; see Methods). Research assistants rated the acoustic similarity of pairs of imitations while blinded to all conditions and hypotheses (see Methods). Inter-rater reliability was high, ICC = 0.39, 95% CI [0.32, 0.47], F(170, 680) = 4.18, *p* < 0.001. Acoustic similarity ratings were fit with a hierarchical linear model predicting similarity from generation with random effects for rater and for category. Imitations from later generations were rated as sounding more similar to one another than imitations from earlier generations, *b* = 0.09 (0.02), *t*(4.5) = 4.42, *p* = 0.009 (Fig. 2A). This result suggests that imitations became more stable (i.e., easier to imitate with high fidelity) with each generation.



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

We also conducted automated analyses of acoustic similarity using Mel Frequency Cepstral Coefficients (MFCCs) as a measure of acoustic distance. By this measure, imitations from later generations were not significantly more similar to one another, *b* = 0.04 (0.03), *t*(357) = 1.18, *p* = 0.24. For our stimuli the correlation between automated analyses of acoustic similarity and rater judgments was low, *r* = 0.20, 95% CI [0.16, 0.25], suggesting that the automated analyses may not capture the acoustic features driving the perception of acoustic similarity of these stimuli. 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. Additional details of automated analyses are reported in the Supporting Information.

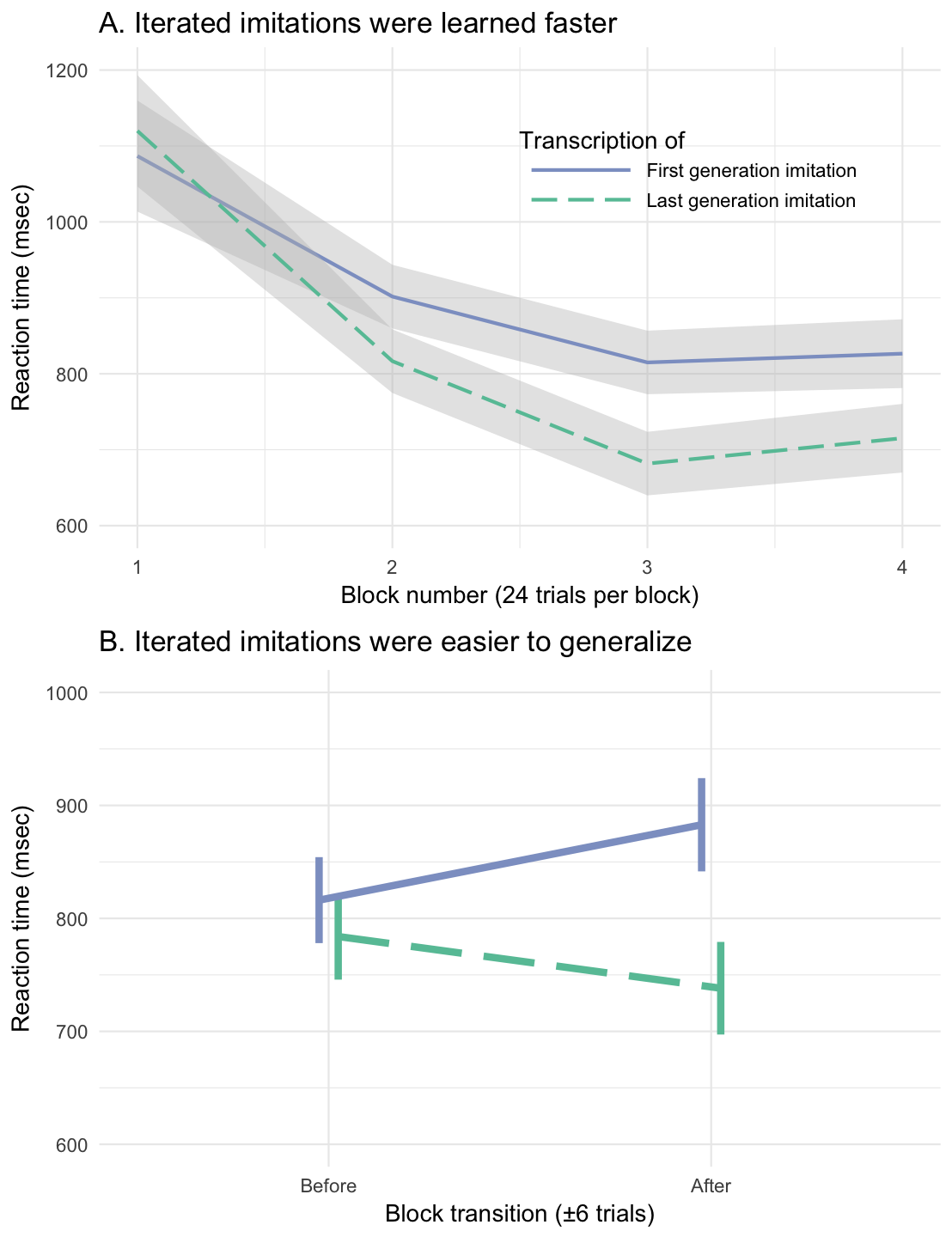
As an additional test of stabilization, we had English-speaking participants recruited via Amazon Mechanical Turk transcribe the vocal imitations into English orthography, and then measured whether transcription agreement and transcription similarity increased over generations. We selected the first and final three imitations in each transmission chain to be transcribed. As a control, we also obtained "transcriptions" of the seed sounds themselves, the results of which are reported in the Supporting Information. We collected a total of 2163 transcriptions ---approximately 20 transcriptions per sound. Some examples of the transcriptions are presented in Table 1.

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

## Iterated imitations make for better category labels

One consequence of imitations becoming more word-like is that they may make for better category labels. For example, an imitation from a later generation, by virtue of having a more word-like form, may be easier to learn to associate with the category of sounds that motivated it than an earlier imitation. To the extent that the later generation has abstracted away from the idiosycracies of a particular category member, it may also be easier to generalize it to novel category members. We tested these predictions using label-referent matching task wherein participantes had to learn to associate the original environmental sounds to transcriptions either from the first generation or the last generation. As before, a sample of transcriptions generated directly from the seed sounds was used as a control. The procedure for selecting otherwise-equal transcriptions in each of the three categories of transcriptions (first generation, last generation, seed sound control) is detailed in the Supporting Information. Here we focus on the results of the comparison between first and last generation transcriptions in the category learning experiment.

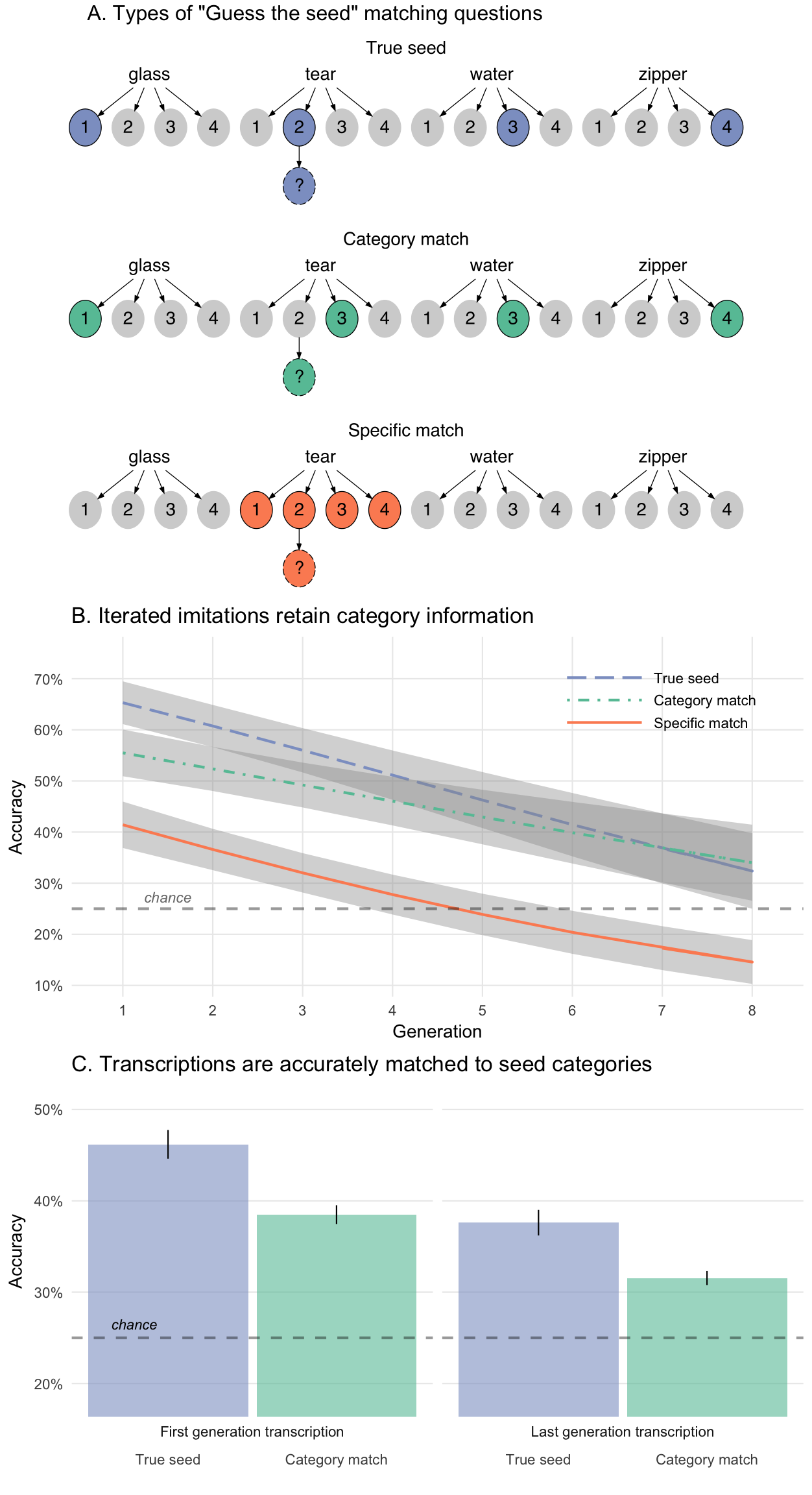
When participants learned novel labels for categories of environmental sounds, they were faster when learning a label from a last generation imitation than from a first generation imitation, *b* = -114.13 (52.06), *t*(39.9) = -2.19, *p* = 0.034 (Fig. 3A). And so in addition to becoming more stable both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



A. Participants achieved faster reaction times in matching transcribed labels to environmental sounds for labels transcribed from later compared to earlier generations. B. There was a generalization cost for first-generation labels, but not last generation labels.

## Iterated imitations retained resemblance to the sounds that motivated them.

As the imitations became more word-like, were they stabilizing on arbitrary acoustic forms, or did they maintain some resemblance to the original environmental sound that motivated them? To test this, we measured the ability of participants naïve to the design of the experiment to match imitations back to their original source relative to other seed sounds from either the same category or from different categories (Fig. 4A). All 365 imitations were tested in the three conditions depicted in Fig. 4A. These conditions differed in the relationship between the imitation and the four seed sounds serving as the choices in a 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.



A. Three types of matching questions used to assess the resemblance between the imitation (and transcriptions of imitations) and the original seed sounds. For each question, participants listened to an imitation (dashed circles) and had to guess which of 4 sound choices (solid circles) they thought the person was trying to imitate. True seed questions contained the actual sound that generated the imitation as one of the choices (correct response). The remaining sounds were sampled from different categories. Category match questions replaced the original seed sound with another sound from the same category. Specific match questions pitted the actual seed against the other seeds within the same category. B. Change in matching accuracy over generations of imitations, shown as predictions of the generalized linear models with ±1 SE of the model predictions. The "category advantage" (Category match vs. Specific match) increased over generations, while the "true seed advantage" (True seed v. Category match) decreased (see main text), suggesting that imitations lose within-category information more rapidly than between-category information. C. Change in matching accuracy over generations of imitations transcribed into English-sounding words. Imitations and transcriptions of imitations could still be matched back to the category of sound that motivated the original imitation even after 8 generations.

Matching accuracy for all question types was 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 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 matching the imitations to the targets was more difficult for later imitations for some trial-types compared to others. The results are shown in Fig. 4B. Performance decreased over generations more rapidly for questions requiring a within-category distinction than for between-category questions, *b* = -0.08 (0.03) log-odds, *z* = -2.69, *p* = 0.007, suggesting that between-category information was more resistant to loss through transmission. An alternative explanation for this result is that the within-category match questions are simply more difficult[[1]](#footnote-1) because the sounds are more acoustically similar to one another than the between-category questions, and therefore, performance might be expected to drop off more rapidly with repeated imitations. However, performance also decreased for the easiest type of question where the correct answer was the actual seed generating the imitation (True seed questions; see Fig. 4A); the advantage of having the true seed among between-category distractors decreased over generations, *b* = -0.07 (0.02) log-odds, *z* = -2.77, *p* = 0.006. The observed increase in the "category advantage" (the advantage of having between-category distractors) combined with a decrease in the "true seed advantage" (the advantage of having the actual seed among the choices) shows that the changes induced by repeated imitation caused the imitations to lose some of properties that linked the earlier imitations to the specific sound that motivated them, while nevertheless preserving a more abstract category-based resembance.

We next tested whether the same was true for *transcriptions* of imitations: Is it possible to match the *written transcription* of the auditory sounds to the original environmental sounds? To test this, we selected a sample of the most frequent transcriptions of first and last generation imitations to use in a modified version of the "Guess the seed" game. Participants were given a novel word and had to guess the sound that was represented by the invented word. The distractors for all questions were between-category, i.e. True seed and Category match. Specific match questions were omitted.

Remarkably, participants were able to guess the correct meaning of a word that was transcribed from an imitation that had been repeated up to 8 times, *b* = 0.83 (0.13) log-odds, odds = -0.18, *z* = 6.46, *p* < 0.001 (Fig. 4C). This was true 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. We hypothesize that later generation transcriptions should be less likely to be distinguished from other sounds within the same category. However, the True seed advantage persisted, indicating that transcriptions of imitations may capture idiosyncratic elements of specific category members more than the acoustic imitations themselves. An alternative reason for not observing a decrease in the True seed advantage for transcriptions is that the sample of transcriptions tested in the Guess the seed game was too small to adequately model changes happening on a per-chain level. Futher reasons for this discrepancy are explored in the Discussion.

# Discussion

Imitative (or “iconic”) words are found across the spoken languages of the world (Dingemanse et al. 2015; Imai and Kita 2014; Perniss, Thompson, and Vigliocco 2010). Counter to past assumptions about the limitations of human vocal imitation, people are surprisingly effective at using vocal imitation to represent and communicate about the sounds in their environment (Lemaitre et al. 2016) and more abstract meanings (Perlman, Dale, and Lupyan 2015), making the hypothesis that early spoken words originated from imitations, a plausible one. We examined whether simply repeating an imitation of an environmental sound--with no intention to create a new word or even to communicate--produce from imitations of environmental sounds more word-like forms.

Our results show that through simple repetition, imitative vocalizations became more word-like both in form and function. In form, the vocalizations gradually stabilized over generations, becoming more similar from imitation to imitation. They also became increasingly standardized according to the phonology of English, as later generations were more consistently transcribed into English orthography. In function, the increasingly word-like forms became more effective as category labels. In a category learning experiment, naïve participants were faster to learn category labels derived from transcriptions of later-generation imitations than those derived from direct imitations of the environmental sound. This fits with previous research showing that the relatively arbitrary forms that are typical of words (e.g. “dog”) makes them better suited to function as category labels compared to direct auditory cues (e.g. the sound of a dog bark; Lupyan and Thompson-Schill 2012; Edmiston and Lupyan 2015; Boutonnet and Lupyan 2015).

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

Our study focused on the process by which words are formed from vocal imitation, and future research remains to determine the full scope of vocal imitation as a source of vocabulary in spoken languages. Although some have estimated the number of imitative words to be small (Crystal 1987; Newmeyer 1992), increasing evidence from across disparate languages shows that vocal imitation is, in fact, a widespread source of vocabulary. Cross-linguistic surveys indicate that onomatopoeia--imitative words used to represent sounds--are a universal lexical category found across the world's languages (Dingemanse 2012). Even English, a language that has been characterized as relatively limited in iconic vocabulary (G. Vigliocco, Perniss, and Vinson 2014), is documented to have hundreds of words for human and animal vocalizations and various kinds of environmental sounds (Rhodes 1994; Sobkowiak 1990). In addition to words that are directly imitative of sounds, many languages also contain semantically broader inventories of ideophones. These words comprise a grammatically and phonologically distinct class of words that are used to express various sensory-rich meanings, such as qualities related to manner of motion, visual properties, textures and touch, inner feelings and cognitive states (Dingemanse 2012; Nuckolls 1999; Voeltz and Kilian-Hatz 2001). Notably, these words are often recognized by native speakers to bear a degree of resemblance to their meaning, an intuition that is confirmed by experiments with naïve listeners (Dingemanse, Schuerman, and Reinisch 2016).

Therefore, if we are to understand the ongoing evolution of spoken languages, it is critical to examine how words are formed from vocal imitation. Here we show that the transition from imitation to word does not require deliberation or guidance: the mere act of repeated imitation can drive vocalizations to become more word-like in both form and function. Notably, as onomatopoeia and ideophones of natural languages maintain a resemblance to the quality they represent, so did our vocal imitations retain a resemblance to the original sound that inspired them. Altogether, our findings suggest how words might be created simply by repeating an imitation of an environmental sound.

# Methods

## Selecting seed sounds

To avoid sounds having lexicalized or conventionalized onomatopoeic forms in English, we used inanimate categories of environmental sounds. Using an odd-one-out norming procedure (*N*=105 participants; see Supporting Information), an initial set of 36 sounds in 6 categories was reduced to a final set of 16 "seed" sounds: 4 sounds in each of 4 categories. The four final categories were: water, glass, tear, zipper.

## Collecting imitations

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

## Measuring acoustic similarity

Acoustic similarity was measured by having research assistants listen to pairs of sounds and rate their subjective similarity. On each trial, raters heard two sounds from subsequent generations were played in succession, but in random order. They then indicated the similarity between the sounds on a 7-point Likert scale from “Entirely different and would never be confused” to “Nearly identical”. 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 (z-scored) by rater prior to analysis.

## 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. Participants also provided transcriptions of the original environmental seed sounds.

Transcriptions from participants who failed a catch trial were excluded (*N*=2), leaving 2163 transcriptions for analysis. Of these, 179 transcriptions (8%) were removed because they contained English words, which was a violation of the instructions of the experiment.

## Learning transcriptions as category labels

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

Participants (67 University of Wisconsin undergraduates) were randomly assigned four novel labels to learn for four categories of environmental sounds. Participants were assigned between-subject to learn labels (transcriptions) of the first or last generation imitations, as well as labels from transcriptions of seed sounds as a control. On each trial, participants heard one of the 16 seed sounds. After a x ms delay , participants saw a label--one of the transcribed imitations--and responded *yes* or *no* using a gamepad controller depending on whether the sound and the word went together. Participants received accuracy feedback (a bell if correct; a buzzing sound if incorrect ) Four participants were excluded from the final sample due to exceedingly high error rates (z>3) and slow reaction times.

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

## Matching imitations to seed sounds.

Participants (*N*=751, recruited from Amazon Mechanical Turk) were asked to listen to imitations, one at a time, imitation and choose one of four possible sounds they thought the person was trying to imitate. The task was unspeeded and no feedback was provided.

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

## Matching transcriptions to seed sounds.

Participants (*N*=468, recruited from Amazon Mechanical Turk) 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.

# References

Arbib, Michael A. 2012. *How the brain got language: The mirror system hypothesis*. Vol. 16. Oxford University Press.

Armstrong, David F, and Sherman Wilcox. 2007. *The gestural origin of language*. Oxford University Press.

Boutonnet, B, and Gary Lupyan. 2015. “Words Jump-Start Vision: A Label Advantage in Object Recognition.” *Journal of Neuroscience* 35 (25): 9329–35. doi:[10.1523/JNEUROSCI.5111-14.2015](https://doi.org/10.1523/JNEUROSCI.5111-14.2015).

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, H H, and R J Gerrig. 1990. “Quotations as demonstrations.” *Language* 66: 764–805. doi:[10.2307/414729?ref=search-gateway:df58c19715a2e512d551c6fd62e27164](https://doi.org/10.2307/414729?ref=search-gateway:df58c19715a2e512d551c6fd62e27164).

Corballis, Michael C. 2003. *From hand to mouth: The origins of language*. Princeton University Press.

Crystal, David. 1987. *The Cambridge Encyclopedia of Language*. Vol. 2. Cambridge Univ Press.

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).

———. 2014. “Making new ideophones in Siwu: Creative depiction in conversation.” *Pragmatics and Society*. doi:[10.1075/ps.5.3.04din](https://doi.org/10.1075/ps.5.3.04din).

Dingemanse, Mark, Damián E Blasi, Gary Lupyan, Morten H Christiansen, and Padraic Monaghan. 2015. “Arbitrariness, Iconicity, and Systematicity in Language.” *Trends in Cognitive Sciences* 19 (10). Elsevier Ltd: 603–15. doi:[10.1016/j.tics.2015.07.013](https://doi.org/10.1016/j.tics.2015.07.013).

Dingemanse, Mark, W Schuerman, and E Reinisch. 2016. “What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages.” *Language* 92. <https://muse.jhu.edu/article/621185/summary>.

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).

Goldin-Meadow, Susan. 2016. “What the hands can tell us about language emergence.” *Psychonomic Bulletin & Review* 24 (1). Psychonomic Bulletin & Review: 1–6. doi:[10.3758/s13423-016-1074-x](https://doi.org/10.3758/s13423-016-1074-x).

Hall, Kathleen Currie, Blake Allen, Michael Fry, Scott Mackie, and Michael McAuliffe. n.d. “Phonological CorpusTools.” *14th Conference for Laboratory Phonology*. Tokyo, Japan.

Hewes, Gordon W. 1973. “Primate Communication and the Gestural Origin of Language.” *Current Anthropology* 14 (1/2). The University of Chicago Press: 5–24. doi:[10.2307/2741093?ref=search-gateway:aea2a42e8a28f3866483ca8d1f65eb5b](https://doi.org/10.2307/2741093?ref=search-gateway:aea2a42e8a28f3866483ca8d1f65eb5b).

Hockett, Charles F. 1978. “In search of Jove’s brow.” *American Speech* 53 (4). JSTOR: 243–313.

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). doi:[10.1098/rstb.2013.0298](https://doi.org/10.1098/rstb.2013.0298).

Kendon, A. 2014. “Semiotic diversity in utterance production and the concept of ’language’.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1651): 20130293–3. doi:[10.1098/rstb.2013.0293](https://doi.org/10.1098/rstb.2013.0293).

Klima Edward, S, and Ursula Bellugi. 1980. *The signs of language*. Harvard University Press.

Lemaitre, Guillaume, and Davide Rocchesso. 2014. “On the effectiveness of vocal imitations and verbal descriptions of sounds.” *The Journal of the Acoustical Society of America* 135 (2): 862–73. doi:[10.1121/1.4861245](https://doi.org/10.1121/1.4861245).

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).

Lewis, J. 2009. “As well as words: Congo Pygmy hunting, mimicry, and play.” In *The Cradle of Language*. The cradle of language. <http://books.google.com/books?hl=en&lr=&id=IVRkzK1VX1oC&oi=fnd&pg=PA236&dq=As+well+as+words+Congo+Pygmy+hunting+mimicry+and+play&ots=vCI59rdgYY&sig=d2svCjyUidFhnlfHmje9SNjgs9s>.

Locke, John. 1948. “An essay concerning human understanding.” In *Readings in the History of Psychology*, edited by Wayne Dennis. Norwalk, CT.

Lupyan, Gary, and Sharon L Thompson-Schill. 2012. “The evocative power of words: Activation of concepts by verbal and nonverbal means.” *Journal of Experimental Psychology: General* 141 (1): 170–86. doi:[10.1037/a0024904](https://doi.org/10.1037/a0024904).

Newmeyer, F J. 1992. “Iconicity and generative grammar.” *Language*. <http://www.jstor.org/stable/416852>.

Nuckolls, Janis B. 1999. “The case for sound symbolism.” *Annual Review of Anthropology* 28 (1). Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA: 225–52.

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).

Perniss, Pamela, Robin L Thompson, and Gabriella Vigliocco. 2010. “Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages.” *Frontiers in Psychology* 1. doi:[10.3389/fpsyg.2010.00227](https://doi.org/10.3389/fpsyg.2010.00227).

Pinker, Steven, and Ray Jackendoff. 2005. “The faculty of language: what’s special about it?” *Cognition* 95 (2): 201–36. doi:[10.1016/j.cognition.2004.08.004](https://doi.org/10.1016/j.cognition.2004.08.004).

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

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

Sobkowiak, Wlodzimierz. 1990. “On the phonostatistics of English onomatopoeia.” *Studia Anglica Posnaniensia* 23: 15–30.

Tomasello, Michael. 2010. *Origins of human communication*. MIT press.

Vigliocco, G, Pamela Perniss, and D Vinson. 2014. “Language as a multimodal phenomenon: implications for language learning, processing and evolution.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1651): 20130292–2. doi:[10.1098/rstb.2013.0292](https://doi.org/10.1098/rstb.2013.0292).

Voeltz, FK Erhard, and Christa Kilian-Hatz. 2001. *Ideophones*. Vol. 44. John Benjamins Publishing.

1. We observed that performance on some Specific match questions dropped below chance for later generations because on some matching trials participants had an apparent aversion to the nominally correct answer. Additional analyses showed that participants were not converging on a single incorrect response. The reason for this pattern is at present unclear. Removing these trials from the analysis does not substantively change the conclusions. [↑](#footnote-ref-1)