Creating words from iterated vocal imitation

We report the results of a large-scale (*N* = 1571) experiment to investigate how spoken words might emerge from the imitation of environmental sounds. Participants played a version of the children’s game “Telephone”. The first generation imitated recognizable environmental sounds (e.g., glass breaking, water splashing) and subsequent generations imitated the imitations of the prior generation for a total of 8 generations. We then examined whether the vocal imitations became more word-like, became more suitable as learned category labels, and retained a resemblance to the original sound. The results showed that the imitations became more stable in form, became more word-like, and more easily learned 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.

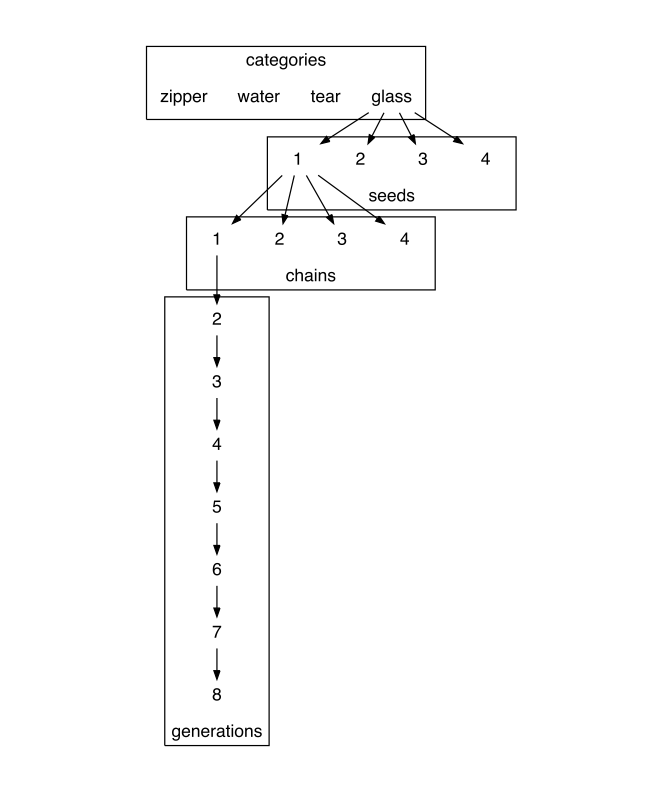
People have long pondered the origins of languages, especially the words that compose them. For example, both Plato in his Cratylus dialogue, and John Locke in his Essay Concerning Human Understanding (1689/1948), examined the “naturalness” of words—whether words are in some way imitative of their meaning. Here, we report a large-scale experiment investigating whether new words can form – gradually and without instruction – simply by iterated imitations of environmental sounds.

The importance of imitation and depiction in the origin of signs is clearly observable in signed languages (Klima & Bellugi 1979), 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 (Arbib 2012; Armstrong & Wilcox 2008; Corballis 2002; Hockett 1978; Tomasello 2008). For example, (Pinker and Jackendoff 2005) argued that, “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” (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 (Goldin-Meadow 2016; Kendon 2014).

But although most words of contemporary spoken languages are not clearly imitative in origin, there has been a growing recognition of the preponderance of imitative words in spoken language (Dingemanse, Blasi, Christiansen, Lupyan & Monaghan 2015; Perniss, Thompson, & Vigliocco 2010)and the frequent use of vocal imitation and depiction in spoken discourse (Clark 2016; Lewis 2009), and some researchers have argued 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 showing that people can, in fact, be highly effective at using vocal imitations to refer to different kinds of sounds – in some cases, even more effective than with the use of words (Lemaitre and Rocchesso 2014). The effectiveness of these imitations arises not because people are able to mimic environmental sounds with high-fidelity, but because they are able to represent 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 sound it represents: 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). Recent work has also shows that people are able to create novel imitative vocalizations for more abstract meanings (e.g. ‘slow’, ‘rough’, ‘good’, ‘many’), in ways that are understandable to naïve listeners (Perlman, Dale, and Lupyan 2015).

Thus, research shows that people can use vocal imitation as an effective means to communicate about the various sounds of their environment and even more abstract concepts. 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 was a recording of an environmental sound. The first generation participant imitated the sound, the next generation imitated the previous imitation, and so on for up to 8 generations.

In a series of subsequent studies, we systematically answered the following questions. 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. 16 seed sounds were selected, four in each category of environmental sound. Participants imitated each seed sound, and then the next generation of participants imitated the imitations and so on for 8 generations.

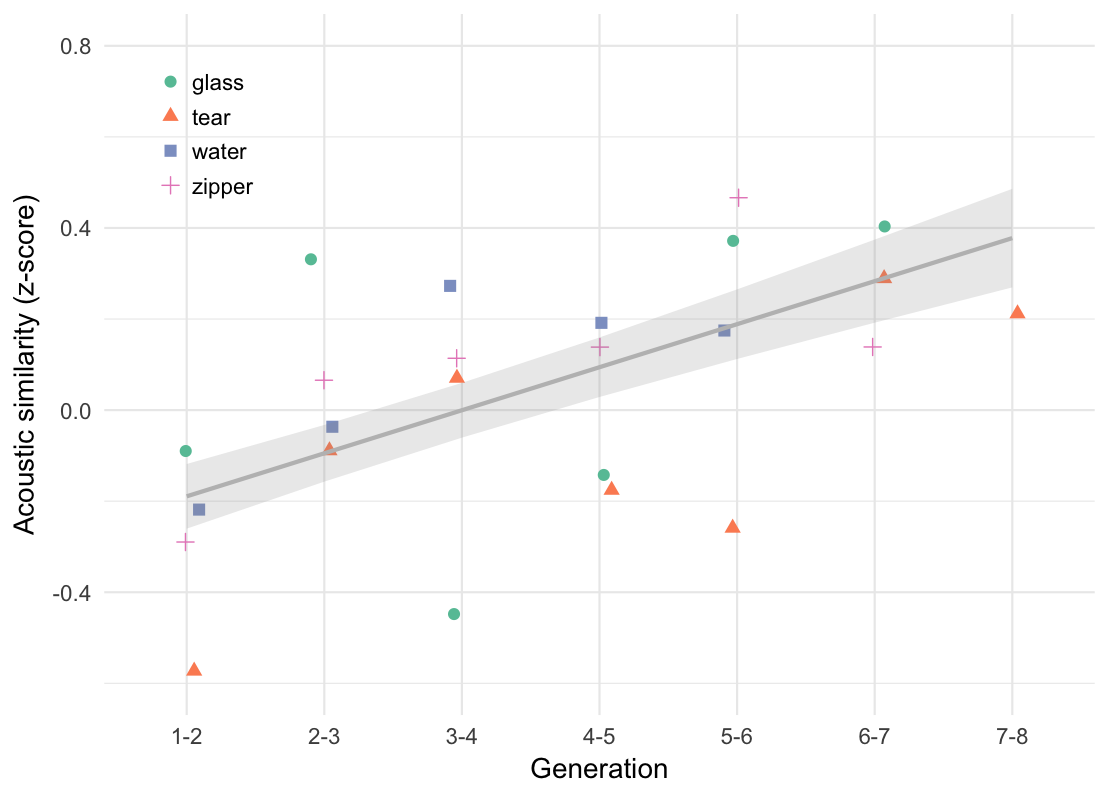
# Results

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

## Imitations stabilized over generations

We began by collecting 480 imitations from 94 participants using Amazon Mechanical Turk. The final set included 365 imitations along 105 contiguous transmission chains (see Methods).

Trained research assistants coded these imitations for acoustic similarity using a blinded, pairwise comparison procedure (see Methods). Inter-rater reliability was high, ICC = 0.39, 95% CI [0.32, 0.47], F(170, 680) = 4.18, *p* < 0.001. Similarity ratings were fit with a hierarchical linear model predicting similarity from generation with random effects for rater and for category. Imitations from later generations were rated as sounding more similar to one another than imitations from earlier generations, *b* = 0.09 (0.02), *t*(4.5) = 4.42, *p* = 0.009 (Fig. 2). This result suggests that imitations may be stabilizing on particular acoustic forms through repetition.

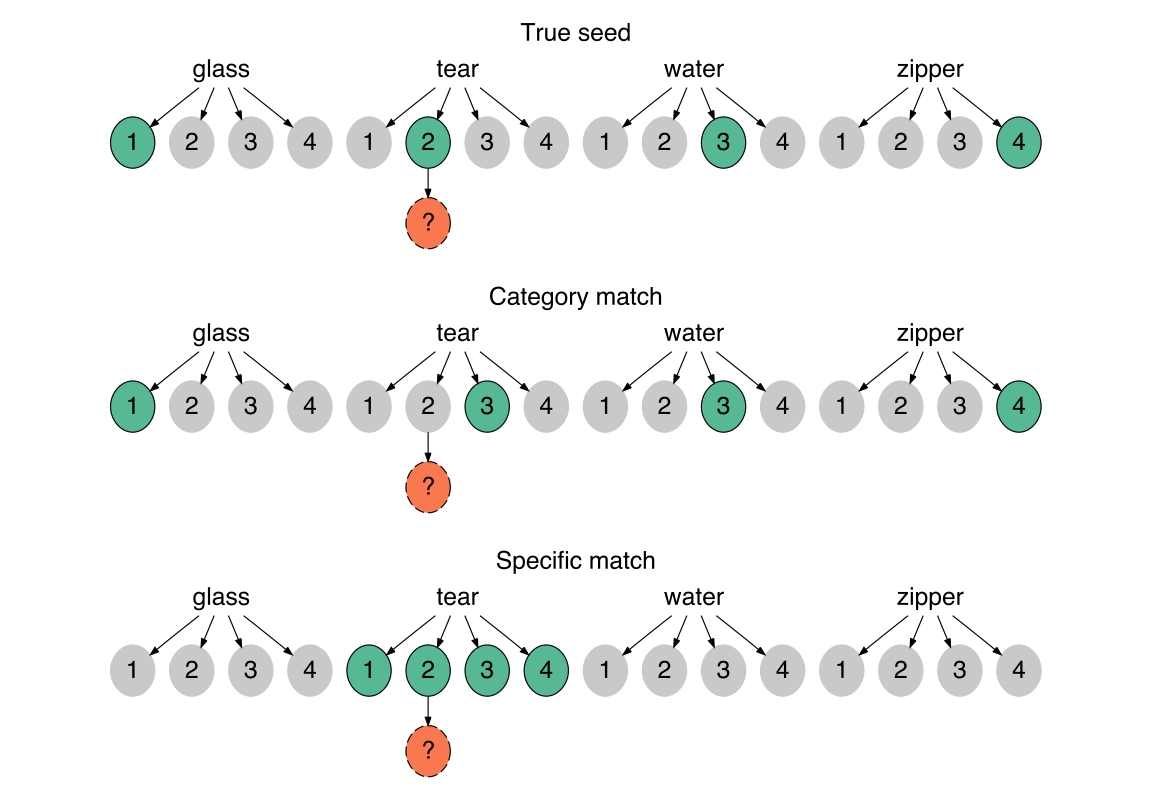


Change in acoustic similarity over generations of repetition. Points show mean acoustic similarity ratings for imitations in each category of environmental sounds. The line shows the linear predictions of a hierarchical model with random effects for rater and category, with error bands designating +/-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

We also calculated automated analyses of imitation fidelity using Mel Frequency Cepstral Coefficients (MFCCs) as a measure of acoustic distance. However, for our stimuli the correlation between automated analyses of acoustic similarity and rater judgments was low, r = 0.20, 95% CI [0.16, 0.25], suggesting that the automated analyses do not capture the acoustic features driving the perception of acoustic similarity. This is possibly due to the non-verbal nature of the imitations as well as variation in recording quality between participants in the online study. We report the results of these automated analyses in the Supporting Information.

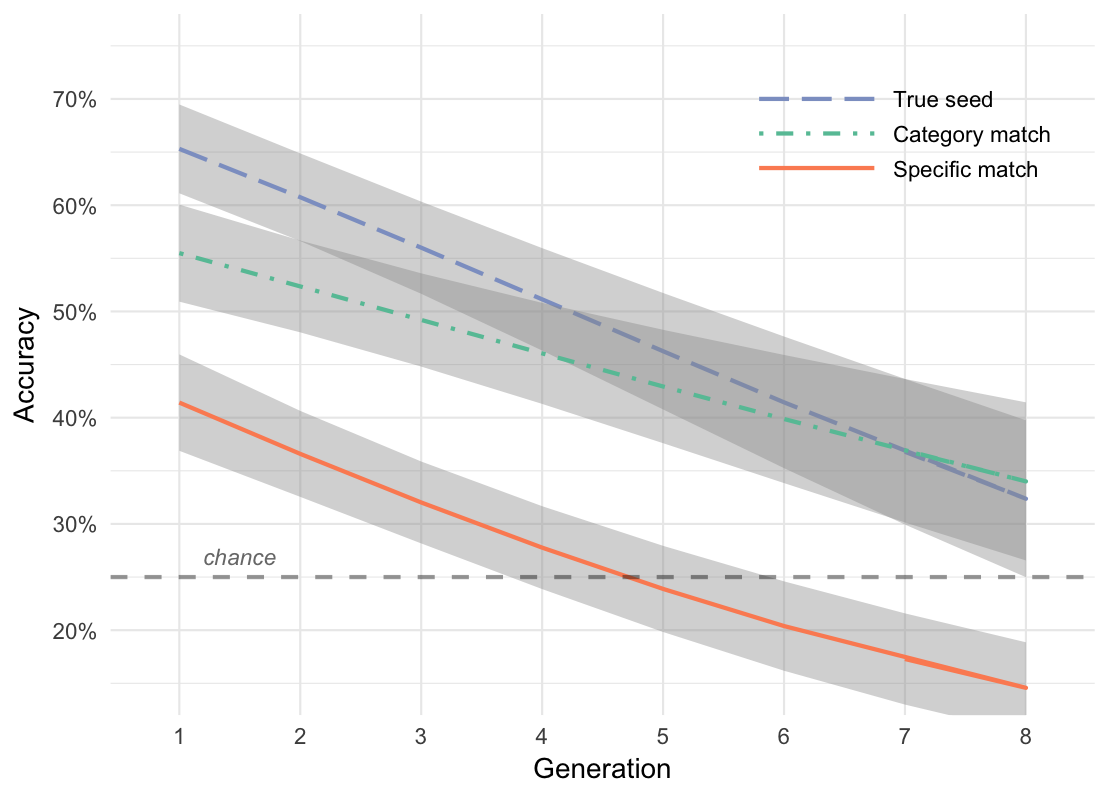
## Imitations retained seed category information

Were the imitations stabilizing on arbitrary acoustic forms or were they maintaining some aspect of the original environmental sound? To test this, we measured the ability of participants naive to the design of the Telephone game to match each imitation back to its original source relative to other seed sounds from either the same category or from different categories (Fig. 3). All 365 imitations were tested in the three conditions depicted in Fig. 3. These conditions differed in the relationship between the imitation and the four seed sounds serving as the choices in the 4 alternative forced choice (4AFC) task. Responses were fit by hierarchical generalized linear models predicting match accuracy as different from chance (25%) based on the type of question being answered (True seed, Category match, Specific match) and the generation of the imitation.



Types of 4AFC matching questions depicted in relation to the original set of 16 seed sounds. For each question, participants listened to an imitation (orange dashed circle) and had to guess which of 4 sound choices (green solid circles) they thought the person was trying to imitate. (Top) True seed questions contained the actual sound that generated the imitation in the choices, and the three distractor sounds were sampled from different categories. (Middle) Category match questions also used distractor sounds from different categories but the "correct" sound was not the actual seed, but a different sound within the same category. (Bottom) Specific match questions pitted the actual seed against the other seeds within the same category.

Matching accuracy for all question types started above chance for the first generation of imitations, *b* = 1.65 (0.14) log-odds, odds = 0.50, *z* = 11.58, *p* < 0.001, and decreased steadily over generations, *b* = -0.16 (0.04) log-odds, *z* = -3.72, *p* < 0.001. We tested whether this increase in matching difficulty was constant across the three types of questions or if some question types became more difficult at later generations than others. The results are shown in Fig. 4. Performance decreased over generations more rapidly for questions requiring a within-category distinction than for between-category questions, *b* = -0.08 (0.03) log-odds, *z* = -2.69, *p* = 0.007, suggesting that between-category information was more resistant to loss through transmission. One explanation for this result is that the within-category match questions are simply more difficult because the sounds are more acoustically similar to one another than the between-category questions and therefore performance might be expected to drop off more rapidly with repeated imitations. However, performance also decreased for the easiest type of question where the correct answer was the actual seed generating the imitation (True seed questions; see Fig. 3); the advantage of having the true seed among between-category distractors decreased over generations, *b* = -0.07 (0.02) log-odds, *z* = -2.77, *p* = 0.006.



Changes in matching accuracy over generations. Matching accuracy is the ability to guess the sound most likely to have generated the imitation relative to other seed sounds used in the experiment. Performance is separated by question type, which describes the relationship between the imitation and the choices in the 4AFC task (see Fig. 3). Lines show predictions from a generalized linear mixed effects model along with +/- 1 standard error of the model predictions. The "category advantage" (category match v. specific match) increased over generations, while the "true seed advantage" (true seed v. category match) decreased. These results suggest that imitations lose within-category information more rapidly than between-category information.

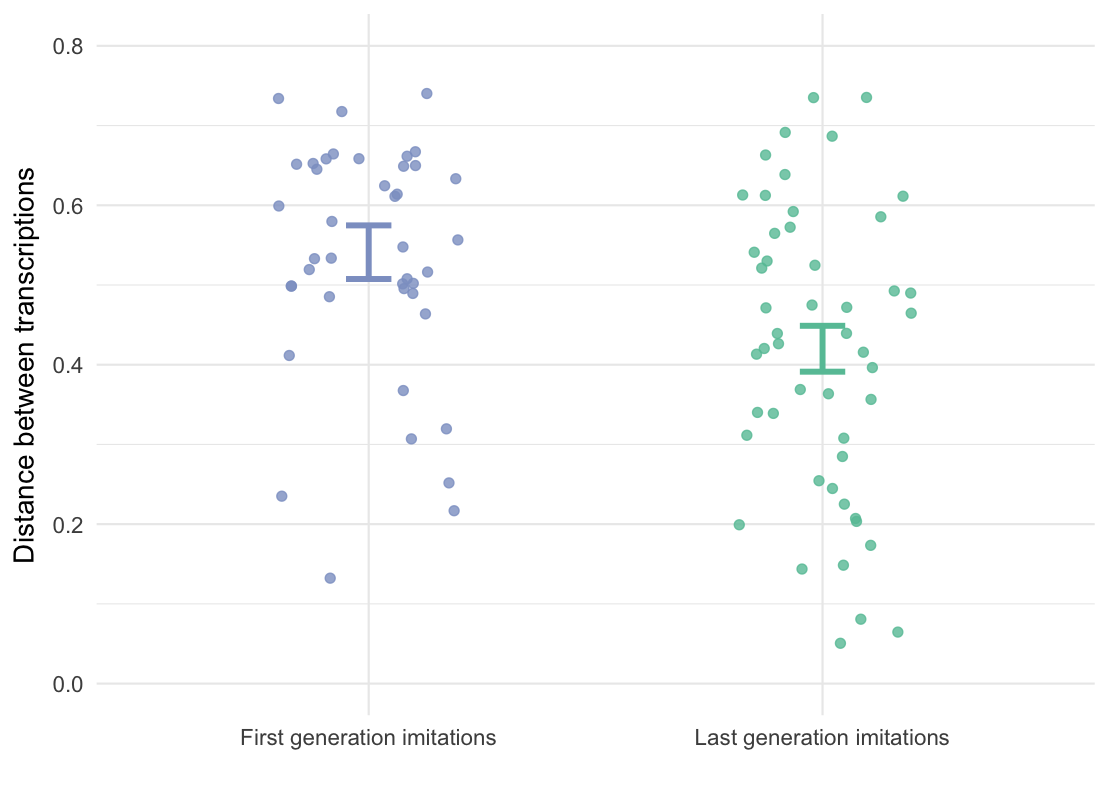
These results indicate that as imitations are repeated they lose exemplar information more rapidly than category information. Later generation imitations were just as likely to be recognized as identifiers of an entire category of environmental sounds as they were of particular sounds within the category.

## Transcription agreement increased over generations

We next tested whether the imitations became more clearly distinguishable as particular words, as opposed to non-linguistic, i.e., non-English sounds. We had English-speaking participants transcribe the imitations into English orthography, and then we measured whether transcription agreement increased over generations. We selected the first and final three imitations in each transmission chain to be transcribed. As a control, we also obtained "transcriptions" of the seed sounds themselves. 216 participants generated a total of 2163 or approximately 20 transcriptions per sound (imitation and seed sounds). Transcriptions containing actual English words and those from participants who failed a catch question were excluded from analysis (n\_transcriptions\_dropped).

To measure transcription agreement we took the average orthographic distance (longest contiguous matching subsequence) between the most frequent transcription and all other transcriptions of a given imitation. Hierarchical linear models were fit predicting orthographic distance from the type of imitation being transcribed (First generation imitations, Last 3 generation imitations) with random effects for transmission chains nested within categories.

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



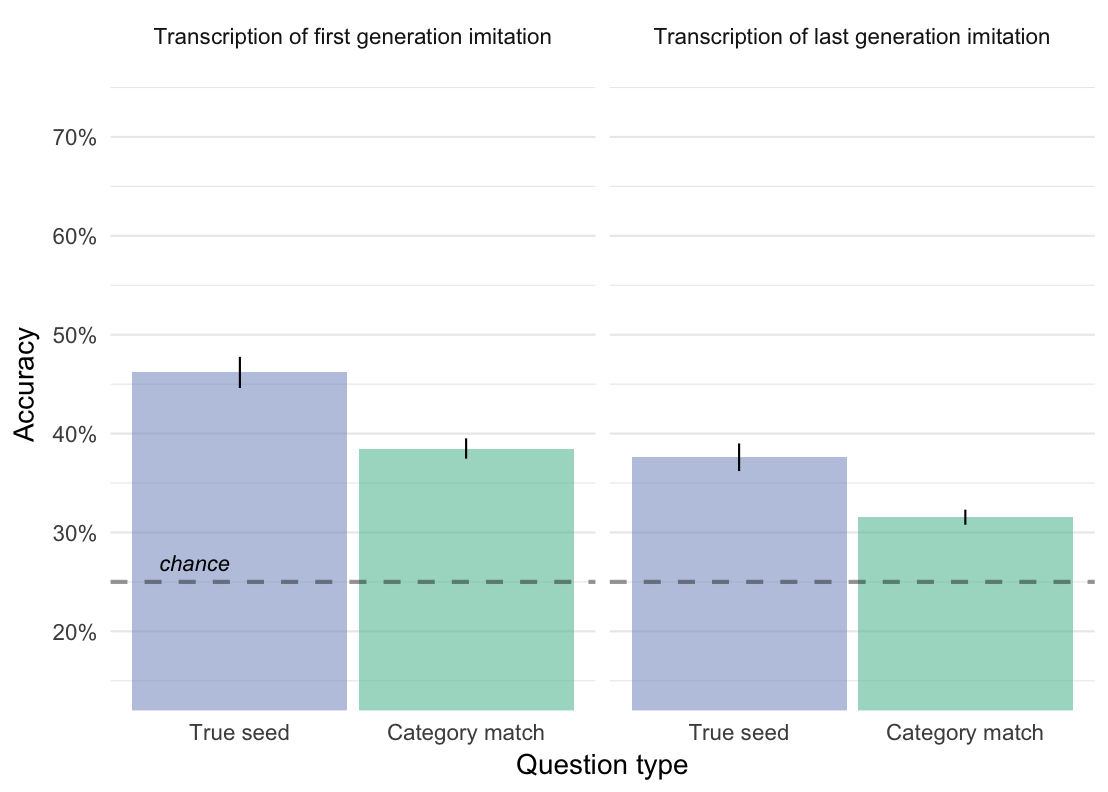
Average orthographic distance among transcriptions of imitations taken from first and last generations. Each point shows the average orthographic distance between the most frequent transcription and all other transcriptions of a single imitation. Error bars are +/- 1 standard error of the linear mixed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

## Transcriptions retained seed category information

We previously demonstrated that people were able to accurately guess the source of an imitation after 8 repetitions, but what about the source of a transcription of an imitation? Do these invented words still resemble the category of sounds that was originally imitated? We tested the top 4 most frequent transcriptions for each imitation in a modified version of the "Guess the Seed" game (see Fig. 3). Participants were given a novel word and had to guess which sound they thought the person who invented the word was talking about. The distractors for all questions were between-category, i.e. Specific match questions were not tested with transcriptions.

Participants were able to guess the correct meaning of the transcribed word above chance even after 8 generations of repetition, *b* = 0.83 (0.13) log-odds, odds = -0.18, *z* = 6.46, *p* < 0.001 (Fig. 6). This was true both for "True seed" questions containing the actual seed generating the transcribed imitation, *b* = 0.75 (0.15) log-odds, odds = -0.28, *z* = 4.87, *p* < 0.001, and for "Category match" questions where participants had to associate transcriptions with a particular category of environmental sounds, *b* = 1.02 (0.16) log-odds, odds = 0.02, *z* = 6.39, *p* < 0.001.

Interestingly, the effect of generation did not vary across these question types, *b* = 0.05 (0.10) log-odds, *z* = 0.47, *p* = 0.637. This indicates that transcriptions of imitations may capture idiosyncratic elements of specific category members more than the imitations themselves. Possible reasons for this asymmetry between imitations and transcriptions are explored in the Discussion.



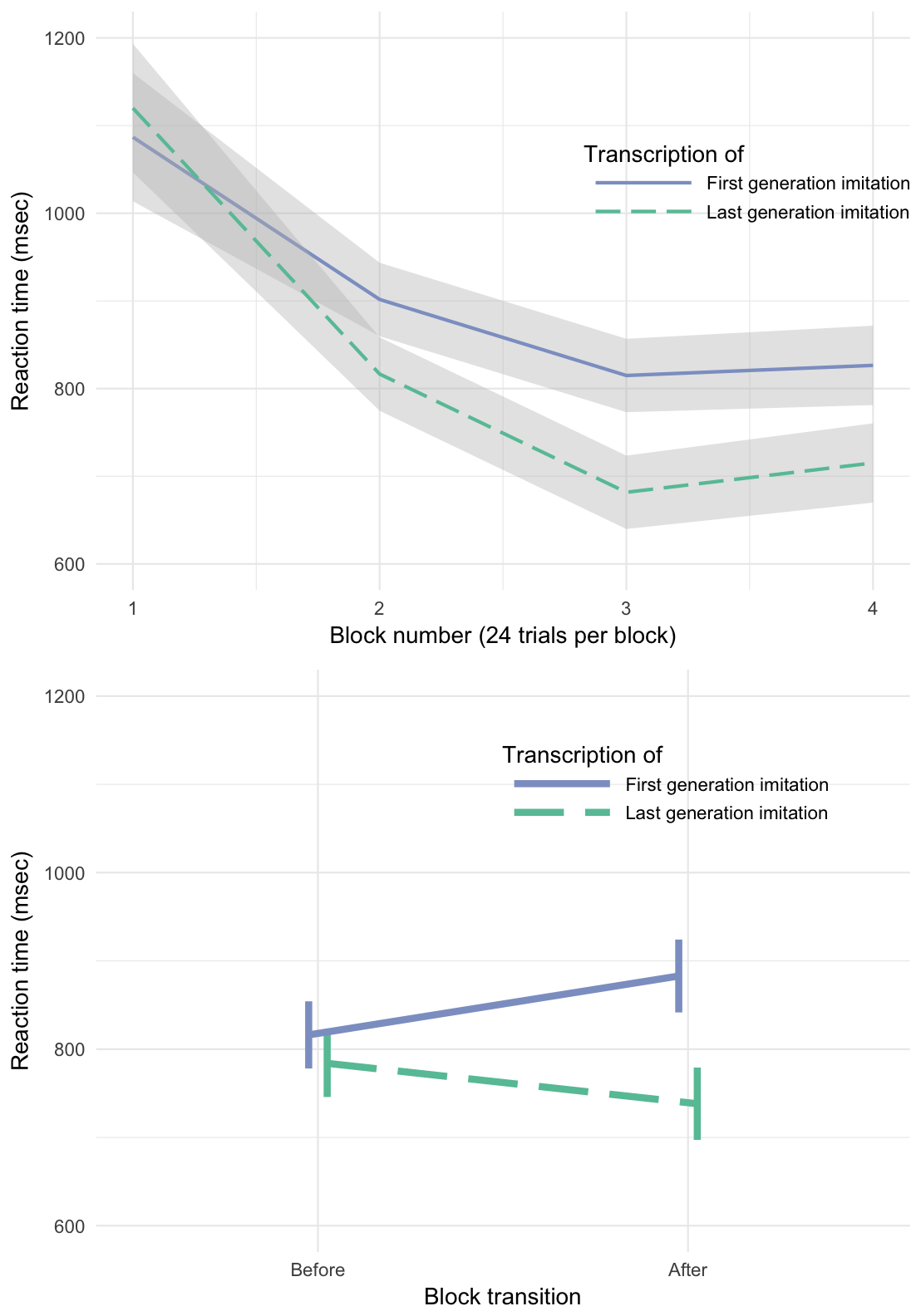
Matching accuracy for transcriptions of imitations taken from first and last generations. Bars represent the predicted means taken from the generalized linear mixed effects model with +/- 1 standard error of the model predictions. True seed questions contained transcriptions of the actual seed generating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

## Repeated imitations were easier to learn as category labels

Our hypothesis was that repetition of imitations would result in increasingly word-like forms, but what are the consequences of this transition for the language user? To examine this question, we tested whether the words created through repetition were easier to learn as category labels.

When participants learned some of the transcriptions as novel category labels for categories of environmental sounds, they were faster when the label came from transcriptions of later generation imitations than from transcriptions of first generation imitations, *b* = -114.13 (52.06), *t*(39.9) = -2.19, *p* = 0.034 (Fig. 7A). In addition to becoming more stable both in terms of acoustic and orthographic properties, imitations that have been more repeated were also easier to learn as category labels.

The effect can be further localized within each block. Comparing RTs on the trials leading up to a block transition and the trials immediately after the block transition revealed a reliable interaction between block transition and the generation of the transcribed label, *b* = -112.50 (48.96), *t*(1732.0) = -2.30, *p* = 0.022 (Fig. 7B). This result suggests that learning transcriptions from later generation imitations were easier to generalize to new category members.



(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

People can be effective at using vocal imitation to represent and communicate about the sounds in their environment, as well as more abstract concepts. Moreover, imitative (or “iconic”) words are found across the spoken languages of the world (Perniss et al. 2010; Dingemanse et al. 2015). However, little is known about the process by which vocal imitations can develop into standardized words. Must new words be deliberately invented as such, or can words form simply by iterated imitations of a sound – even when there is no intention to communicate. To examine this question, we conducted a large-scale, online, iterated vocal imitation experiment.

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 imitations became more word-like in their increased effectiveness as category labels. In a 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 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. Thus, 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, idiosyncratic features of the sound could become rendered as categorical phonological features. This process could exaggerate the features and 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 (Vigliocco, Perniss, & Vinson 2014), is documented to have hundreds of words for human and animal vocalizations and various kinds of environmental sounds (Rhodes 1994; Sobkowiak 1996). 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 1996; Voeltz & 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 et al. 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 can be a simple process: 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 show how words might be created from the repetition of one-shot vocal imitations of an original sound.

# Methods

## Selecting seed sounds

We selected inanimate categories of sounds because they were less likely to have lexicalized onomatopoeic forms already in English, and they were assumed to be less familiar and more difficult to imitate. Using an odd-one-out norming procedure (*N* = 105 participants), an initial set of36 sounds in 6 categories was reduced to a final set of 16 "seed" sounds: 4 sounds in each of 4 categories. The four final categories included: water, glass, tear, zipper. The results of the norming procedure are presented in the Supporting Information.

## Collecting imitations

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

## Measuring acoustic similarity

Acoustic similarity was measured by having research assistants listen to pairs of sounds and rate their subjective similarity. On each trial, raters heard two sounds played in succession. Then they rated the similarity between the sounds on a 7-point scale. They were instructed that a 7 on this scale meant the sounds were nearly identical, whereas a 1 meant the sounds were entirely different and would never be confused. Raters were encouraged to use as much of the scale as they could while maximizing the likelihood that, if they did this procedure again, they would reach the same judgments. Full instructions are provided in the Supporting Information. Ratings were normalized prior to analysis (z-scores).

## Matching imitations to seeds

Participants (*N* = 751) were paid to complete an online survey containing 4AFC questions. For each question in the survey, participants listened to an imitation and guessed which of four possible sounds they thought the person was trying to imitate. No feedback was provided.

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

## Collecting transcriptions of imitations

Participants (*N* = 216) were paid to transcribe sounds into words in an online survey. They listened to imitations and were instructed to write down what they heard as a single word so that the written word would sound as much like the message as possible. Instructions are provided in the Supporting Information.

Imitations were drawn at random from the first and last three generations of all imitations collected in the Telephone game. As a control, we also had participants "transcribe" words directly from listening to the environmental seed sounds. Transcriptions from participants who failed a catch trial were excluded (*N* = 2), leaving 2163 transcriptions for analysis. Of these, 179 transcriptions were removed because they contained English words, which was a violation of the instructions of the experiment.

## Matching transcriptions to seeds

Participants (*N* = 468) completed a modified version of the "Guess the seed" game. Instead of listening to imitations, participants now read a word (a transcription of an imitation), which they were told was an invented word. They were instructed that the word was invented to describe one of the four presented sounds, and they had to guess which one. Of all the unique transcriptions that were collected for each sound (imitations and seed sounds), only the top four most frequent transcriptions were used in the matching experiment. 6 participants failed a catch trial and were excluded, leaving 461 participants in the final sample.

## Learning transcriptions as category labels

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

Participants (*N* = 67) were randomly assigned four novel names for four categories of environmental sounds. Participants were assigned between-subject to learn labels from first or last generation imitations, as well as labels from transcriptions of seed sounds as a control. They learned the referents for these names in a trial-and-error category learning experiment. On each trial, participants heard one of the 16 seed sounds and then saw a word--one of the transcriptions of the imitations. They responded yes or no using a gamepad as to whether the sound and the label went together. Initially they were forced to guess, but because they received feedback on their performance, over trials they learned the names of the categories. 63 outlier participants were excluded from the final sample due to high error rates and slow reaction times.

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

# References

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 Richard J Gerrig. 1990. “Quotations as Demonstrations.” *Language, Cognition, and Neuroscience* 66 (4). Linguistic Society of America: 764–805. doi:[10.2307/414729?ref=search-gateway:df58c19715a2e512d551c6fd62e27164](https://doi.org/10.2307/414729?ref=search-gateway:df58c19715a2e512d551c6fd62e27164).

Crystal, D. 1987. *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.

Dingemanse, M. (2014). Making new ideophones in Siwu: Creative depiction in conversation. *Pragmatics and Society*, 5(3), 384-405. doi:10.1075/ps.5.3.04din.

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, 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, 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).

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*, *68*(4), 756–796. https://doi.org/10.2307/416852

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 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.

Taylor, Kevin J. 2006. KA-BOOM! A Dictionary of Comic Book Words, Symbols & Onomatopoeia. Mora Publications, November 1.

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