Dear Mr Edmiston:  
  
I am writing to inform you that your manuscript RSPB-2017-1610 entitled "The emergence of words from vocal imitations" has, in its current form, been rejected for publication in Proceedings B.  
  
This action has been taken on the advice of the Associate Editor and the referees. Whilst they all see much merit in the paper, the referees have indicated that substantial revisions are necessary. Further, on a really important point, we can see the potential relevance of the topic for a biological journal, but it needs to be presented much more clearly. For example, your cover letter notes that the paper is "of broad appeal, bridging between linguistics, cognitive psychology, and semiotics", without mentioning biology - so I wonder if you are aiming for a different audience? I am therefore not clear as to whether it fits with Proc B's remit of addressing fundamental biological principles.  Thus we would be willing to consider a resubmission, provided the comments of the referees are fully addressed and provided you can make the relevance to a biological journal clear.  However please note that this is not a provisional acceptance and you should consider whether the paper would actually be better suited to a different journal.  
  
The resubmission will be treated as a new manuscript.  However, we will approach the same reviewers if they are available and it is deemed appropriate to do so by the Editor. Please note that resubmissions must be submitted within six months of the date of this email. In exceptional circumstances, extensions may be possible if agreed with the Editorial Office. Manuscripts submitted after this date will be automatically rejected.  
  
Please find below the comments made by the referees, not including confidential reports to the Editor, which I hope you will find useful. If you do choose to resubmit your manuscript, please upload a ‘response to referees’ document including details of how you have responded to the comments, and the adjustments you have made. We require a copy of the manuscript with revisions made since the previous version marked as ‘tracked changes’ to be included in the ‘response to referees’ document.  
  
To upload a resubmitted manuscript, log into [http://mc.manuscriptcentral.com/prsb](http://mc.manuscriptcentral.com/prsb" \t "_blank) and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Resubmission." Please be sure to indicate in your cover letter that it is a resubmission, and supply the previous reference number.  
  
Sincerely,  
  
Professor Loeske Kruuk  
Editor, Proceedings B  
mailto: [proceedingsb@royalsociety.org](mailto:proceedingsb@royalsociety.org" \t "_blank)  
  
  
Associate Editor, Dr Katie Slocombe  
  
Comments to Author:  
My main concern is whether this article addresses a fundamental biological principle, and is therefore a good fit for the journal. R2 shares some concerns about the suitability of the paper for the readership of the journal, so in your revisions you need to consider how you can make the relevance of your findings to understanding fundamental biological principles clear.  
  
  
Reviewer(s)' Comments to Author:  
  
Referee: 1  
  
Comments to the Author(s)  
Edminston et al, present an interesting and very innovative work aimed at studying the role of vocal imitation in the evolution of language. In my view, the methodology and results are very sound. In this regard, I would like to raise some brief questions whose answers should be included in the text  
  
1. In the introduction the authors write (80-81): does the imitation of a particular water-splashing sound become, over generations of repeated  imitation, a better label for the more general category of water-splashing sounds? What does this mean in linguistic terms? Do the authors refer in any way to the iconicity of language?  
2. In “Collecting vocal imitations”, 111-112, “Participants were allowed to listen  
to each target sound multiple times” How many times were allowed to listen to each target? Was there a maximum number of repetions allowed to each participant?  
3.  Figures in the text, please, the graphs must include the headings on the x and y axes  
  
  
Referee: 2  
  
Comments to the Author(s)  
Review of “The emergence of words from vocal imitation” by Peirce et al.  
  
The paper presents 3 experiments asking whether vocal imitation for environmental sounds, through a process of cultural transmission, can give rise to word-like labels which crucially still maintain iconicity. Experiment 1 elicited vocalizations in response to specific environmental noises that then were passed through generations of speakers. Next, these vocalizations were rated/scored for their acoustic similarity and it was found that within each chain, vocalizations produced later in the chains were rated as more similar to one another than those produced earlier on (although different across chains). Vocalizations were then transcribed in written form and there was more agreement among naïve trascribers in their transcriptions of later labels. Experiment 2 showed that produced vocalization across generations were iconic of the original sound. Finally, Experiment 3 argues that later vocalizations are more iconic of the category of sounds rather than the specific label.  
  
I enjoyed reading this well-written paper. The experiments are clever and the methods interesting. I have however a number of general issues with the paper, some of which I am sure the authors can address quite easily. But for me the main issues relate to whether the findings are clear cut and general enough to warranty publication in the journal.  
  
1.My first general issue concerns what these findings tell us. The paper shows that modern English-speaking humans show biases, that are present in their language, in the specific task. Namely, they can imitate sounds (and we have plenty examples of onomatopoaies in English as well across languages), and their imitations become more word-like (and there are plenty indications from previous studies using iterated learning - see point 2 below) that through generations language-like properties emerge in this paradigm. Thus, the findings are not surprising.  
  
Still, unsurprising findings can be informative about language evolution. Here, however, I also have a problem. In addition to the usual difficulty in making any sort of claim concerning language origin and evolution starting from modern humans (who already know a language), the present findings do not allow us to claim that imitation in vocalizations is a sine qua non, nor that they have been the first step. They do not falsify the possibility that gesture came first, nor that a combination of gesture and vocalization was key. They simply show that imitation in vocalization might have played a part. Thus, I find the argument proposed in the paper not to be fully justified on the basis of the findings.  
  
2. There is by now an important tradition within language studies in using the method presented here in Experiment 1 (and multiple variations on this methods). This method is referred as iterated learning has been introduced by Simon Kirby and colleagues (Kirby et al., 2008). It is rather puzzling that the present paper does not make any reference to the previous papers, to the name of the paradigm and, importantly, to the learning mechanisms that have abundantly discussed in the literature as underscoring the processes simulated with the use of this paradigm. This, in my mind, is especially important. On line 419, they write “Our results show that through simple repetition...”, which seems to me to trivialise what is going on in the study. In fact, they don’t really address what is going on – what the learning mechanisms may be that give rise to rendering imitations more word-like through generations. It seems strange to me to use a very particular and well-discussed experimental paradigm and then not discuss at all why you used it or what the effects of it are.  
  
3. The authors call ‘imitative’ what, in the literature is more commonly referred as ‘iconic”. They do not define what they mean by ‘imitative’ nor whether ‘imitative’ is still a correct label to use for the word-like productions that they observe in later generations (where the vocalizations are indeed less imitative). This is an important point if we want to understand the mechanisms that underscore the effects found.  
  
4. I am unsure about the reasoning behind experiment 3. They argue that if vocalizations are more word-like, then they should more likely refer to categories rather than specific exemplars. I am always uneasy when talking about categorization levels as we do not know, really, whether the specific sounds within each category might already be categories standing for various tokens of the same type. That having been said, I know that some people have argued that iconicity is at the level of categories, not exemplars (e.g., Monagham, 2012). However, we also know that speakers may have more or fewer specific labels within a category (e.g., Eskimos have 10 different words for snow, English speakers only have one) depending upon whether it is important to differentiate or not. Hence one can argue that there is a bias toward category-level labels just because making finer distinctions is not so relevant to the task.  
  
5. The different results for spoken and written labels that they got in Experiment 3 may be task related: in the spoken version, there were three questions, only 2 in the written version. This alternative account needs to be addressed.  
  
6. Related to 5, the issue of differences in difficulty between questions in Experiment 3, I think, is not sufficiently addressed.