Decision for round #1 : Revision needed

Recommender: François Munoz

Thank you for submitting your interesting work to PCI Ecology.

You will find enclose two reviews that provide useful suggestions for more in-depth analyses and discussion of the findings.

Please consider their suggestions in a revised version of the manuscript.

reply#1: Thank you, the anonymous reviewer and François Massol for appreciating our work, and for providing very useful and thoughtful suggestions that helped us sharpening the analyses and the message we aim to deliver. We have addressed and incorporated almost all comments, and we have uploaded the revised manuscript in bioRxiv (version 3, which should be soon updated online). Please, refer to our replies below to comments made by the Reviewers (numbered and in blue text, as are in blue text the changes implemented in the manuscript).

Review by anonymous reviewer 1

I was asked to review the manuscript entitled "On the quest for novelty in ecology". I enjoyed reading the straightfoward and clear study, that is concise and precise. I think the framework in which the authors thought about the study is clear. The methods are sound and well explained, and the analyses easy to reproduce as all of the materials are openly shared by the authors. The paper is a good quantitative contribution to the debate of how papers are increasingly framed as "novel". I, however, have the impression that the paper falls short on its initial goal of testing to what extent papers in ecology use novelty terms. While I fully understand the appeal of a clear manuscript centered around a single question with a single analysis to answer it, I have the impression that the manuscript could be made stronger with minimal additional analyses. Also, I think the discussion could be broaden to link to more current pressing issues as well as taking stronger stances regarding the results of the paper.

reply#2: Thanks for your positive assessment of our short paper, and for offering insights that stimulated us to broaden the contextualization and generalization of our findings, as well as refining the analyses. Please, refer to our replies to your specific comments below.

The authors mainly focus on one analysis in their manuscript regarding the temporal trend of the proportion of novelty versus confirmatory terms in the abstracts of the corpus of articles. The results are clear, but I think much more could be added to papers, with minimal additional analyses and could deepen the message of the manuscript. For example, the authors evaluated the prevalence of novelty terms in abstracts, but they don't discuss its prevalence in titles. I would expect that titles are even greater vehicle to convey novelty as they are the first thing that would attract a potential reader. If not in the main text, I, and I'm sure any potential reader, would ask how these trends compare across journals.

reply#3: Thanks for this suggestion. We thought about having titles included also in previous versions of our paper, but unfortunately we could not run any meaningful analysis because title-derived data are highly zero-inflated (see plot below).



Could we see, as supplementary material, the raw trends across the 17 journals? Are there differences across journals? Also society versus commercial journals? In particular, given that the authors mention 65% of journals having novelty as a criteria for publication, are these specific journals showing a stronger trend in novelty terms than the others? Similarly, the used linear mixed-models don't have they variance reported (see Nagakawa's R²). I would be interested to see to how large the variance explained by the journal random effect is relative to one of fixed effects.

reply#4: Thanks for this comment, which we found especially useful. For within-journal trends, we produced individual plots showing the temporal patterns of novel vs confirmatory terms and included this in the revised version of the manuscript (see new Fig. 3). These trends are generally aligned with the overall pattern reported in Fig. 2, with only one exception (Austral Ecology). In the new Fig. 3, we also identified (with different

symbols) those journals which mention novelty as a criterium in their descriptions (authors' guidelines, aims and scopes), as for Table S1.

Concerning the possible comparison between society vs commercial journals, while potentially highly interesting, our sample prevented us to test this – most of the 17 journals belong to/are associated with societies (such as British Ecological Society, Ecological Society of America, OIKOS) and published by Wiley.

As for reporting the variance explained (R^2) in the models, we have now included this information – see new Fig. 4.

I also have the impression that the discussion and conclusion of the manuscript don't take any stance regarding novelty and confirmatory terms. While a neutral stance gives the impression of rising above it all, I do think that given the framework of the manuscript, a more engaged stance should be adopted by the authors. What do they think of their results? Are there concerning? What does "novelty"" really means? Should novelty matters in ecology or not? I would have liked to have a slightly longer discussion and recommendation on these points. In particular, I find the conclusion half-hearted. I understand the authors message to consider each paper in itself with its specificities without focusing on novel versus confirmatory results, but it goes against the introduction of the paper. In the introduction it is pointed out that science is cumulative with rare breakthrough, meaning confirmatory results should be more important for its advancement and novel results rarer. I do think that this should emphasize the importance of dropping the novelty criterion for journals and increase the importance of confirmatory results. Given the current debate on the science reproducibility crisis, I would also have expected some mention of the importance of confirmatory results in this regard.

reply#5: We are very glad that the Reviewer asked us for extending upon this crucial point. We took a more neutral stance in the previous version as, in prior submissions of this paper to other journals, we have been criticized for being too directional (or even too speculative). We are certainly in favor of elaborating more around possible implications as well as ways forward, better linking back to the reasoning presented in the introductory paragraph. We have therefore edited/added text in the discussion and concluding paragraphs (L181-195; L238-248).

Also, we have replaced all the titling of the paragraphs (i.e. Introduction, Methods, Results, Discussion, Conclusions) with a less "orthodox" heading, yet more tightly (and ideally more punchily) connected to the main point we would like the reader to get away with in each section. We are convinced that this rewording better reflects the inherently thought-provoking nature of our concise piece.

Review by François Massol

In this paper, Ottaviani and colleagues show that the use of novelty-esque words has increased in the abstracts of ecology papers since 1997, while the use of words associated with replication and reproducibility has stayed stable. Based on this analysis, they offer some interpretations of these findings, one of which being that journals are partially responsible for the increase in novelty-related adjectives.

Overall, I really appreciate the work performed in this study. Although short, this paper will be useful to the community in these times of redefinition of research assessments, journal rankings, and viability of research system models.

I have a few comments, some of them methodological, some linked to the existing literature on the topic. I hope these will help the authors improve their manuscript.

reply#6: Thanks for your very thoughtful comments which helped us refining both the analytical framework and narrative.

1. Regarding the myth of the "lone genius" (page 3), to me the most enlightening book on the topic has been Clifford Conner's "People history of science". I very strongly suggest this reading to the authors. It is full of counter-examples to the common myth and it also suggests that "lone scientists" (even when they are brilliant) make mistakes while groups are more likely to self-correct.

Conner, C. (2005) A People's History of Science: Miners, Midwives, and Low Mechanicks, Nation books.

reply#7: Great – we did not know this book, now cited in L48.

2. Because your work was performed on abstracts of ecology papers, I was wondering whether you could contrast your findings with those of Weinberger et al. (2015) -- they looked at whether common rules for writing abstracts were really associated to more citations (or less).

Weinberger, C. J., Evans, J. A. & Allesina, S. (2015) Ten Simple (Empirical) Rules for Writing Science. PloS Computational Biology, 11, e1004205.

reply#8: Again, thanks for spotting this out, which nicely aligns with our work and with that of Mammola 2020 (Trend Ecol Evol) – we therefore cite this paper in L62, L211.

3. The biggest correction I would suggest to this paper is about statistics: I do appreciate the use of linear models (with mixed effects), but I am not sure that using a Gaussian model (here, an 'lm' or 'lmer' in R) is appropriate; rather I think using a generalized linear model with a binomial distribution for the error would be better (your data is actually a collection of articles with a 0/1 response to the questions "is novelty in the abstract?" and "is reproducibility in the abstract?"). Or if you really prefer to work with proportions rather than counts, at least make the distribution "beta". However, in my opinion, there is really a large bonus to working with counts here insofar as it makes use of the information contained in the number of papers published by each journal each year (with proportions, you don't see this piece of information, and so the result might be biased by journal that publish fewer papers).

reply#9: Good point. We have now revised the temporal trend analysis as suggested, generating two Bernoulli models (i.e. 0-1 discrete models) to test the relationship between the use (1) or non-use (0) of confirmatory or novelty terms and publication year. The overall results remain unchanged, showing that the probability of abstracts containing novelty terms (but not confirmatory terms) has increased significantly in recent years. The major advantage of this approach is that the analysis is now much more robust, as it is based on the full dataset (>50,000 individual articles) rather than 20 data points (the yearly proportions).

To visualize these trends, however, we found the percentage plot more intuitive, so we retained it (Fig. 2 and the new Fig. 3), including a linear fit added solely for visual presentation.

4. The last parts of the analysis (IF against novelty, citations against novelty, and IF against confirmatory) work on residuals of a previous analysis (a GAM on citations per paper corrected for age of the paper). In all statistical textbooks (and when you discuss with statistician colleagues), it is always more advisable to make a single model with all factors rather than make a second model on the residuals of the first. Could you rather do that here, maybe through a GAM that forces your linear predictor to be linear but still keeps the spline effect of paper age (if you really want to keep it that way)?

reply#10: Thanks for this very helpful comment. We followed your suggestion and fitted two new models – one for number of citations, one for impact factor as response variables. There, we set novel terms, confirmatory terms, the year of publication, and number of words in the abstract as fixed effects. Then, for the number-of-citation model, we set journal as random effect, whereas for the impact-factor model we ran a simple linear model as it did not make sense to set 'journal' as random effect (being highly collinear with IF). We created a new Fig. 4 to synthetically illustrate these new results.

5. More generally, the M&M is too short -- I had to read through the R script to really understand what was being done, this is not very friendly to colleagues allergic to stats...

reply#11: We have amended that part, better specifying what we did (L105-131).

6. In the reporting of the results, we don't have a clue as to what are the "beta's" and which tests were performed to get these p-values.

reply#12: In the revised version, we indicate the new test statistics to get the p-values.

7. Last methodological point: can you control for the size of the abstract of each article (number of words per abstract)? Because longer abstracts could potentially contain more room for both confirmatory and novelty terms, so that might be a factor explaining part of the effect (or at least you could take that into account).

reply#13: This is another excellent hint, thanks – we controlled for this (as an additional fixed effect) in the new models (see **#10**).