

THE ROYAL SOCIETY
PUBLISHING

ROYAL SOCIETY
OPEN SCIENCE

Why are we not evaluating multiple competing hypotheses in ecology and evolution?

Gustavo S. Betini, Tal Avgar and John M. Fryxell

Article citation details

R. Soc. open sci. 4: 160756.

<http://dx.doi.org/10.1098/rsos.160756>

Review timeline

Original submission: 22 September 2016

1st revised submission: 28 September 2016

2nd revised submission: 2 December 2016

Final acceptance: 5 December 2016

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

Note: This manuscript was transferred from another Royal Society journal with peer review.

Review History

Decision letter (RSOS-160735)

23-Sep-2016

Dear Dr Betini:

I write you in regards to manuscript # RSOS-160735 entitled "Why are we not evaluating multiple competing hypotheses?" which you submitted to Royal Society Open Science. In light of the appropriateness of your manuscript for our journal, your manuscript has been denied publication in Royal Society Open Science.

As is our policy, the journal is only able to consider invited reviews for publication. Having said this, the journal does welcome unsolicited review proposals, but all proposals must first be agreed by the relevant Subject Editor. Proposals for reviews should be no longer than 1 side of A4, include a structural outline with sub-headings to briefly explain description of content, and must succinctly identify the core issue(s) to be addressed. If you would like to propose a review for submission, please contact the journal directly via the email address openscience@royalsociety.org.

For more information about the article types considered by the journal, please follow the link below:

<http://rsos.royalsocietypublishing.org/content/information-authors#question2>

© 2017 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited

Thank you for considering Royal Society Open Science for the publication of your research. I hope the outcome of this specific submission will not discourage you from the submission of future manuscripts.

Regards,
Alice Power
Editorial Coordinator, Royal Society Open Science
openscience@royalsociety.org

RSOS-160756.R0 (Resubmission)

Review form: Reviewer 1 (Alessandro Minelli)

Is the manuscript scientifically sound in its present form?

Yes

Are the interpretations and conclusions justified by the results?

Yes

Is the language acceptable?

Yes

Is it clear how to access all supporting data?

N/A

Do you have any ethical concerns with this paper?

No

Have you any concerns about statistical analyses in this paper?

No

Recommendation?

Accept with minor revision (please list in comments)

Comments to the Author(s)

My only critical remarks, both minor, are:

Title

In the title, no specific discipline is mentioned, but the article specifically addresses current practice in biology and, within biology, in ecology and evolution specifically. This should be reflected in the article's title

Multiple causes

(last lines of p. 8 to first lines of p. 9) "The profound utility of the null Hardy-Weinberg equilibrium is that, if trait frequencies do exhibit change over time, then it must be because one of the six conditions has been violated." - Comment: in any given case, more than just one condition may have been violated.

Alessandro Minelli - University of Padova

Review form: Reviewer 2

Is the manuscript scientifically sound in its present form?

No

Are the interpretations and conclusions justified by the results?

No

Is the language acceptable?

Yes

Is it clear how to access all supporting data?

No - see specific comment to authors.

Do you have any ethical concerns with this paper?

No

Have you any concerns about statistical analyses in this paper?

Yes

Recommendation?

Reject

Comments to the Author(s)

The manuscript addresses an important topic and a potential mismatch between what researchers are doing, and what they perhaps believe they ought to be doing. The aim and theme are very worthwhile. In general, I would like to see more published on the topic of how science is done and ought to be done, with contrasts between the two highlighted. However, I find deficiencies in the execution of the manuscript.

Overall, the tone of this manuscript makes me uncomfortable. Is it a polemic? Is it a review of literature? Is it a contribution to philosophy of science? I ask, not to reinforce arbitrary disciplinary boundaries, but because the text has an intermediate level of detail that I find frustrating as a reader. Concepts are introduced but not sufficiently linked to the literature or to examples - whether hypothetical thought-experiments or actual published experiments.

I supply a few suggestions, below. I am relaxed about whether the authors follow these specific suggestions, or choose to increase the depth and rigour of the manuscript in other ways. For example: re-focusing the text on the empirical study of literature could be another approach.

"Needless to say, exploratory analysis and data mining should be conducted with extra attention to type I errors." (p. 6)

Is this needless - in which case, does it have to be said at all? My instinct is, it is not needless. But then, I am left hungry for detail. What type of attention is required (for a "philosophy of science" essay)? And/or: are there a few references that could be cited, covering this (for a "review article")? Can something pithy be said about work that gets this wrong (for a "polemic")?

Syllogisms (p. 6). I think the authors' point is, the first syllogism follows from the statements but one of the statements is false. However, this could be made more explicit.

Conflict between logic and prior knowledge (pp. 6-7). The authors should be more explicit about whether this is necessarily a problem. They may or may not agree with Stegenga (2013, "Evidence in biology and the conditions of success", *Biol Philos* 28:981-1004) - but some reference to Stegenga's paper, or similar arguments, would clarify and enrich the text.

"... a statistical null model is usually generated in opposition to a single hypothesis" (p. 9). This appears to contradict the author's outline of HWE (p. 8), in which the null model is presented as being in opposition to six hypotheses. This might be covered by the authors' use of "usually". But this requires explanation.

"... information-theoretic basis" (p. 9). In context it appears the authors are talking about AIC (or BIC) - in which case, they should make it clear. Otherwise it is not immediately obvious whether they mean AIC, or mutual information, entropic priors, or some other approach linked to information theory.

"The candidate set of models tested ... considered in the context of the full set of alternatives" (p. 9). This is intriguing. However, I did not fully understand. The authors should consider expanding this section, either with a thought-experiment or with reference to an actual study in biology.

"Null models don't come easily. ..." (pp. 9-10). An explanatory example - again, whether a thought-experiment or an actual study - is required to clarify the text here.

"The only remedy is to ..." (p. 10). If the authors wish to take a polemical tone, it seems to me this requires further support from their own text. For a philosophy-of-science angle, it requires deeper analysis. Or for a review, it requires references to the literature and discussion of those references. Alternatively, "only remedy" could be rephrased in a less inflammatory way. (Of course: combinations of these approaches are possible.)

"Good scientists always follow the literature ... Forcing oneself to follow other 'schools of thought' ... is vitally important". This paragraph is too light. For example: consider the broader "sociology of science" angle, in which it may be argued rival scientists with entrenched points of view contribute to progress of the field as a whole (Hull 1988, *Science as a Process*)? The authors may not agree - which is fine - but some attention to this point of view would strengthen the manuscript.

"We propose that concerted collaboration between opposite sides of the same debate is the only way one could reach such far-reaching conclusions." (pp. 10-11) As-is, this statement comes across as false. Major advances in science have been made by researchers not noted for their collaboration with the opposite side (e.g. RA Fisher, Darwin). The authors' statement is also in opposition to the thesis of Hull (1988, cited above). The authors might explicitly consider counter-examples, and/or explicitly delimit the extent to which their proposal applies.

Crowdsourcing (p. 11). This is intriguing and - like the authors - I have not heard of it applied in ecology and evolution. It would be helpful to outline a thought experiment, showing in principle how it could work.

"The fallacy of factorial design." (pp. 11-12) There are many questions in ecology and evolution which cannot be addressed by factorial design, necessarily being based on observations (e.g. phylogeny, and in general perhaps the evolution of wild populations). Some attention could be given to these.

"Although not a law, this tension between new and old ideas essentially reflects a conflict between new and old generations." This is another rather inflammatory statement which requires further detail, discussion and references to the literature (beyond the single reference cited), and/or reference to individual cases. Otherwise, counter-examples come to mind (e.g. Lamarck, who made his most major contributions as an older man).

Decision trees (p. 13). Like crowdsourcing (above), this is intriguing but, if included, requires coverage with further depth. How might a decision tree work for a particular example in ecology or evolution? How would it compare to a more typical approach?

"... before spending large amounts of money and years in the field." Again, some (hypothetical or real) example seems necessary here, illustrating the tension between laboratory and field work.

Box 1. The authors should clarify "alternative possible outcomes". For example, for continuous data (as might be used to estimate a mean), there are - in principle - infinite possible outcomes.

"lacked any clear motivation or hypothesis" (p. 21; also Figure 1 legend). Could this be rephrased or expanded? It comes across as a negative verdict on the quality of three papers, but without justification. It occurs to me (without knowing the papers), they might have a strong implicit motivation that was simply not covered in the paper itself (e.g. a genome sequencing project, to generate a widely usable resource; other possibilities come to mind).

Citations and/or DOI for the papers analysed in Figure 1 should be included as a spreadsheet or table as supplementary data. Optionally, I would also like this table/spreadsheet to include the authors' categorisation of the papers.

Figure 1. Since these bars represent an estimate (being based on a random sample of literature), some indication of the certainty of results or precision of estimates would be useful - whether in the figure, legend or text.

Decision letter (RSOS-160756)

28-Oct-2016

Dear Dr Betini,

The editors assigned to your paper ("Why are we not evaluating multiple competing hypotheses?") have now received comments from reviewers. We would like you to revise your paper in accordance with the referee and Subject Editor suggestions which can be found below (not including confidential reports to the Editor). Please note this decision does not guarantee eventual acceptance.

Please submit a copy of your revised paper within three weeks (i.e. by the 20-Nov-2016). If we do not hear from you within this time then it will be assumed that the paper has been withdrawn. In exceptional circumstances, extensions may be possible if agreed with the Editorial Office in advance. We do not allow multiple rounds of revision so we urge you to make every effort to fully address all of the comments at this stage. If deemed necessary by the Editors, your manuscript will be sent back to one or more of the original reviewers for assessment. If the original reviewers are not available we may invite new reviewers.

To revise your manuscript, log into <http://mc.manuscriptcentral.com/rsos> and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision. Revise your manuscript and upload a new version through your Author Centre.

When submitting your revised manuscript, you must respond to the comments made by the referees and upload a file "Response to Referees" in "Section 6 - File Upload". Please use this to document how you have responded to the comments, and the adjustments you have made. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response.

In addition to addressing all of the reviewers' and editor's comments please also ensure that your revised manuscript contains the following sections as appropriate before the reference list:

- Ethics statement (if applicable)

If your study uses humans or animals please include details of the ethical approval received, including the name of the committee that granted approval. For human studies please also detail whether informed consent was obtained. For field studies on animals please include details of all permissions, licences and/or approvals granted to carry out the fieldwork.

- Data accessibility

It is a condition of publication that all supporting data are made available either as supplementary information or preferably in a suitable permanent repository. The data accessibility section should state where the article's supporting data can be accessed. This section should also include details, where possible of where to access other relevant research materials such as statistical tools, protocols, software etc can be accessed. If the data have been deposited in an external repository this section should list the database, accession number and link to the DOI for all data from the article that have been made publicly available. Data sets that have been deposited in an external repository and have a DOI should also be appropriately cited in the manuscript and included in the reference list.

If you wish to submit your supporting data or code to Dryad (<http://datadryad.org/>), or modify your current submission to dryad, please use the following link:
<http://datadryad.org/submit?journalID=RSOS&manu=RSOS-160756>

- Competing interests

Please declare any financial or non-financial competing interests, or state that you have no competing interests.

- Authors' contributions

All submissions, other than those with a single author, must include an Authors' Contributions section which individually lists the specific contribution of each author. The list of Authors should meet all of the following criteria; 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published.

All contributors who do not meet all of these criteria should be included in the acknowledgements.

We suggest the following format:

AB carried out the molecular lab work, participated in data analysis, carried out sequence alignments, participated in the design of the study and drafted the manuscript; CD carried out the statistical analyses; EF collected field data; GH conceived of the study, designed the study, coordinated the study and helped draft the manuscript. All authors gave final approval for publication.

- Acknowledgements

Please acknowledge anyone who contributed to the study but did not meet the authorship criteria.

- Funding statement

Please list the source of funding for each author.

Once again, thank you for submitting your manuscript to Royal Society Open Science and I look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Yours sincerely,
Alice Power
Editorial Coordinator
Royal Society Open Science

on behalf of Kevin Padian
Subject Editor, Royal Society Open Science
openscience@royalsociety.org

Associate Editor's comments:

Associate Editor: 1

Comments to the Author:

The referees have made a number of useful suggestions as to how the manuscript can be revised in order to strengthen the way the message is put across to readers. One referee, in particular, has a list of suggestions for the authors to consider, largely relating to the style or tone of the manuscript and its overall aim. These suggestions should be helpful to the authors and give the authors a choice in how they want to accommodate the recommendations. It would clearly be beneficial to accommodate as many of these comments as possible in order to improve the manuscript.

Comments to Author:

Reviewers' Comments to Author:

Reviewer: 1

Comments to the Author(s)

My only critical remarks, both minor, are:

Title

In the title, no specific discipline is mentioned, but the article specifically addresses current practice in biology and, within biology, in ecology and evolution specifically. This should be reflected in the article's title

Multiple causes

(last lines of p. 8 to first lines of p. 9) "The profound utility of the null Hardy-Weinberg equilibrium is that, if trait frequencies do exhibit change over time, then it must be because one of the six conditions has been violated." - Comment: in any given case, more than just one condition may have been violated.

Alessandro Minelli - University of Padova

Reviewer: 2

Comments to the Author(s)

The manuscript addresses an important topic and a potential mismatch between what researchers are doing, and what they perhaps believe they ought to be doing. The aim and theme are very worthwhile. In general, I would like to see more published on the topic of how science is done and ought to be done, with contrasts between the two highlighted. However, I find deficiencies in the execution of the manuscript.

Overall, the tone of this manuscript makes me uncomfortable. Is it a polemic? Is it a review of literature? Is it a contribution to philosophy of science? I ask, not to reinforce arbitrary disciplinary boundaries, but because the text has an intermediate level of detail that I find frustrating as a reader. Concepts are introduced but not sufficiently linked to the literature or to examples - whether hypothetical thought-experiments or actual published experiments.

I supply a few suggestions, below. I am relaxed about whether the authors follow these specific suggestions, or choose to increase the depth and rigour of the manuscript in other ways. For example: re-focusing the text on the empirical study of literature could be another approach.

"Needless to say, exploratory analysis and data mining should be conducted with extra attention to type I errors." (p. 6)

Is this needless - in which case, does it have to be said at all? My instinct is, it is not needless. But then, I am left hungry for detail. What type of attention is required (for a "philosophy of science" essay)? And/or: are there a few references that could be cited, covering this (for a "review article")? Can something pithy be said about work that gets this wrong (for a "polemic")?

Syllogisms (p. 6). I think the authors' point is, the first syllogism follows from the statements but one of the statements is false. However, this could be made more explicit.

Conflict between logic and prior knowledge (pp. 6-7). The authors should be more explicit about whether this is necessarily a problem. They may or may not agree with Stegenga (2013, "Evidence in biology and the conditions of success", *Biol Philos* 28:981-1004) - but some reference to Stegenga's paper, or similar arguments, would clarify and enrich the text.

"... a statistical null model is usually generated in opposition to a single hypothesis" (p. 9). This appears to contradict the author's outline of HWE (p. 8), in which the null model is presented as being in opposition to six hypotheses. This might be covered by the authors' use of "usually". But this requires explanation.

"... information-theoretic basis" (p. 9). In context it appears the authors are talking about AIC (or BIC) - in which case, they should make it clear. Otherwise it is not immediately obvious whether they mean AIC, or mutual information, entropic priors, or some other approach linked to information theory.

"The candidate set of models tested ... considered in the context of the full set of alternatives" (p. 9). This is intriguing. However, I did not fully understand. The authors should consider expanding this section, either with a thought-experiment or with reference to an actual study in biology.

"Null models don't come easily. ..." (pp. 9-10). An explanatory example - again, whether a thought-experiment or an actual study - is required to clarify the text here.

"The only remedy is to ..." (p. 10). If the authors wish to take a polemical tone, it seems to me this requires further support from their own text. For a philosophy-of-science angle, it requires deeper analysis. Or for a review, it requires references to the literature and discussion of those references. Alternatively, "only remedy" could be rephrased in a less inflammatory way. (Of course: combinations of these approaches are possible.)

"Good scientists always follow the literature ... Forcing oneself to follow other 'schools of thought' ... is vitally important". This paragraph is too light. For example: consider the broader "sociology of science" angle, in which it may be argued rival scientists with entrenched points of view contribute to progress of the field as a whole (Hull 1988, *Science as a Process*)? The authors may not agree - which is fine - but some attention to this point of view would strengthen the manuscript.

"We propose that concerted collaboration between opposite sides of the same debate is the only way one could reach such far-reaching conclusions." (pp. 10-11) As-is, this statement comes across as false. Major advances in science have been made by researchers not noted for their collaboration with the opposite side (e.g. RA Fisher, Darwin). The authors' statement is also in opposition to the thesis of Hull (1988, cited above). The authors might explicitly consider counter-examples, and/or explicitly delimit the extent to which their proposal applies.

Crowdsourcing (p. 11). This is intriguing and - like the authors - I have not heard of it applied in ecology and evolution. It would be helpful to outline a thought experiment, showing in principle how it could work.

"The fallacy of factorial design." (pp. 11-12) There are many questions in ecology and evolution which cannot be addressed by factorial design, necessarily being based on observations (e.g. phylogeny, and in general perhaps the evolution of wild populations). Some attention could be given to these.

"Although not a law, this tension between new and old ideas essentially reflects a conflict between new and old generations." This is another rather inflammatory statement which requires further detail, discussion and references to the literature (beyond the single reference cited), and/or reference to individual cases. Otherwise, counter-examples come to mind (e.g. Lamarck, who made his most major contributions as an older man).

Decision trees (p. 13). Like crowdsourcing (above), this is intriguing but, if included, requires coverage with further depth. How might a decision tree work for a particular example in ecology or evolution? How would it compare to a more typical approach?

"... before spending large amounts of money and years in the field." Again, some (hypothetical or real) example seems necessary here, illustrating the tension between laboratory and field work.

Box 1. The authors should clarify "alternative possible outcomes". For example, for continuous data (as might be used to estimate a mean), there are - in principle - infinite possible outcomes.

"lacked any clear motivation or hypothesis" (p. 21; also Figure 1 legend). Could this be rephrased or expanded? It comes across as a negative verdict on the quality of three papers, but without justification. It occurs to me (without knowing the papers), they might have a strong implicit motivation that was simply not covered in the paper itself (e.g. a genome sequencing project, to generate a widely usable resource; other possibilities come to mind).

Citations and/or DOI for the papers analysed in Figure 1 should be included as a spreadsheet or table as supplementary data. Optionally, I would also like this table/spreadsheet to include the authors' categorisation of the papers.

Figure 1. Since these bars represent an estimate (being based on a random sample of literature), some indication of the certainty of results or precision of estimates would be useful - whether in the figure, legend or text.

Author's Response to Decision Letter for (RSOS-160756)

See Appendix A.

Decision letter (RSOS-160756.R1)

05-Dec-2016

Dear Dr Betini,

I am pleased to inform you that your manuscript entitled "Why are we not evaluating multiple competing hypotheses in ecology and evolution?" is now accepted for publication in Royal Society Open Science.

You can expect to receive a proof of your article in the near future. Please contact the editorial office (openscience_proofs@royalsociety.org and openscience@royalsociety.org) to let us know if you are likely to be away from e-mail contact. Due to rapid publication and an extremely tight schedule, if comments are not received, your paper may experience a delay in publication.

Royal Society Open Science operates under a continuous publication model (<http://bit.ly/cpFAQ>). Your article will be published straight into the next open issue and this will be the final version of the paper. As such, it can be cited immediately by other researchers. As the issue version of your paper will be the only version to be published I would advise you to check your proofs thoroughly as changes cannot be made once the paper is published.

In order to raise the profile of your paper once it is published, we can send through a PDF of your paper to selected colleagues. If you wish to take advantage of this, please reply to this email with the name and email addresses of up to 10 people who you feel would wish to read your article.

On behalf of the Editors of Royal Society Open Science, we look forward to your continued contributions to the Journal.

Kind regards,
Alice Power
Royal Society Open Science
openscience@royalsociety.org
<http://rsos.royalsocietypublishing.org/>

Associate Editor Comments to Author:

Thank you for taking such care to accommodate all of the reviewers' comments and for providing a very clear response letter and annotated revised manuscript. The manuscript constitutes an interesting and thought-provoking paper.



COLLEGE OF BIOLOGICAL SCIENCE
Department of Integrative Biology

November 29th, 2016

Prof. Kevin Padian
Subject Editor, Royal Society Open Science

Dear Professor Padian,

We thank you for the chance to resubmit our paper entitled “*Why are we not evaluating multiple competing hypotheses?*” to Royal Society Open Science and thank the two reviewers for comments on the original version of this manuscript. We have now addressed all of these comments and think that the manuscript has been significantly improved as a result. Below, we have copied the referees comments and included our replies in bold. We have highlighted in red all changes we have made in the manuscript.

We look forward to hearing back from you and please do not hesitate to contact us if you require further clarification or additional information.

Sincerely,

Gustavo S. Betini

Tal Avgar

John M. Fryxell

Department of Integrative Biology,

University of Guelph,

Guelph, Ontario N1G 2W1, Canada.

Reviewer: 1

Comments to the Author(s)

My only critical remarks, both minor, are:

Title

In the title, no specific discipline is mentioned, but the article specifically addresses current practice in biology and, within biology, in ecology and evolution specifically. This should be reflected in the article's title

Thanks for the suggestions. We have now modified the title: “Why are we not evaluating multiple competing hypotheses in ecology and evolution?”

Multiple causes

(last lines of p. 8 to first lines of p. 9) “The profound utility of the null Hardy-Weinberg equilibrium is that, if trait frequencies do exhibit change over time, then it must be because one of the six conditions has been violated.” – Comment: in any given case, more than just one condition may have been violated.

Alessandro Minelli - University of Padova

We have modified the text in the manuscript following reviewer 1 suggestion (l. 214). Now it reads: “... then it must be because at least one of the six conditions...”

Reviewer: 2

Comments to the Author(s)

The manuscript addresses an important topic and a potential mismatch between what researchers are doing, and what they perhaps believe they ought to be doing. The aim and theme are very worthwhile. In general, I would like to see more published on the topic of how science is done and ought to be done, with contrasts between the two highlighted. However, I find deficiencies in the execution of the manuscript.

Overall, the tone of this manuscript makes me uncomfortable. Is it a polemic? Is it a review of literature? Is it a contribution to philosophy of science? I ask, not to reinforce arbitrary disciplinary boundaries, but because the text has an intermediate level of detail that I find frustrating as a reader. Concepts are introduced but not sufficiently linked to the literature or to examples - whether hypothetical thought-experiments or actual published experiments.

I supply a few suggestions, below. I am relaxed about whether the authors follow these specific suggestions, or choose to increase the depth and rigour of the manuscript in other ways. For example: re-focusing the text on the empirical study of literature could be another approach.

We thank reviewer 2 for the comments. During the preparation of the manuscript, we gave full consideration to this issue. We believe our manuscript is a perspective piece that gives our personal account on the subject but it is also supported by empirical examples. With this in mind, we followed reviewer 2 suggestions and provided more examples from the ecological and evolutionary literature that could make the concepts and ideas clearer. We have also clarified this in the text (l. 69)

"Needless to say, exploratory analysis and data mining should be conducted with extra attention to type I errors." (p. 6)

Is this needless - in which case, does it have to be said at all? My instinct is, it is not needless. But then, I am left hungry for detail. What type of attention is required (for a "philosophy of science" essay)? And/or: are there a few references that could be cited, covering this (for a "review article")? Can something pithy be said about work that gets this wrong (for a "polemic")?

We agree with reviewer 2 here. We have expanded our discussion, including examples from the literature to explain why pattern seeking is a problem in data mining (l. 132-143). Briefly, even randomly generated variables can have low p-values and high R^2 depending on how the analysis is conducted. Without a set of predictions in mind, one might unconsciously bias the search for results that support their favourite hypothesis.

Syllogisms (p. 6). I think the authors' point is, the first syllogism follows from the statements but one of the statements is false. However, this could be made more explicit.

The text now reads: "The conclusion follows the premises, but one of the premises is false (i.e. whales do not walk)." (l. 155)

Conflict between logic and prior knowledge (pp. 6-7). The authors should be more explicit about whether this is necessarily a problem. They may or may not agree with Stegenga (2013, "Evidence in biology and the conditions of success", *Biol Philos* 28:981-1004) - but some reference to Stegenga's paper, or similar arguments, would clarify and enrich the text.

We believe this is a problem because researchers might judge evidence based on their favourite hypothesis (prior knowledge), not based on logic. This could be even more important because, as stated in Stegenga (now cited in the main text), it is not clear how researchers define whether some evidence is reliable or not. We re-arrange the paragraph and introduced empirical evidence to support our claim (l. 165-169).

"... a statistical null model is usually generated in opposition to a single hypothesis" (p. 9). This appears to contradict the author's outline of HWE (p. 8), in which the null model is presented as being in opposition to six hypotheses. This might be covered by the authors' use of "usually". But this requires explanation.

By statistical null model we mean the H_0 in the classical statistical test, not a research hypothesis, as in the case of the HWE. We have clarified this in the text (l. 224-225).

"... information-theoretic basis" (p. 9). In context it appears the authors are talking about AIC (or BIC) - in which case, they should make it clear. Otherwise it is not immediately obvious whether they mean AIC, or mutual information, entropic priors, or some other approach linked to information theory.

Yes, the reviewer is correct here. We have now explicitly mentioned both AIC and BIC in the text (l. 229-230).

"The candidate set of models tested ... considered in the context of the full set of alternatives" (p. 9). This is intriguing. However, I did not fully understand. The authors should consider expanding this section, either with a thought-experiment or with reference to an actual study in biology.

The goal of this paragraph is to emphasize that model competition is not a stepwise regression procedure. We have now included this information in the main text and elaborated more on this point (l. 236-237 and l. 240-242).

"Null models don't come easily. ..." (pp. 9-10). An explanatory example - again, whether a thought-experiment or an actual study - is required to clarify the text here.

Good point. We have now used the HWM as an example to explain why researchers rarely think about null models. We wrote: "Evolutionary biologists want to understand how evolution happens, so they think about the mechanisms driving changes in allele frequencies, not about the mechanisms that might keep allele frequencies constant or situations under which these mechanisms are never at work. (l. 244-250).

"The only remedy is to ..." (p. 10). If the authors wish to take a polemical tone, it seems to me this requires further support from their own text. For a philosophy-of-science angle, it requires deeper analysis. Or for a review, it requires references to the literature and discussion of those references. Alternatively, "only remedy" could be rephrased in a less inflammatory way. (Of course: combinations of these approaches are possible.)

Our goal was not to be polemical, but provide a way to avoid cognitive bias. We, however, do not know of any study showing that the use of creative thinking minimize cognitive bias. We have now elaborated this point in the manuscript (l. 257-263).

"Good scientists always follow the literature ... Forcing oneself to follow other 'schools of thought' ... is vitally important". This paragraph is too light. For example: consider the broader "sociology of science" angle, in which it may be argued rival scientists with entrenched points of view contribute to progress of the field as a whole (Hull 1988, *Science as a Process*)? The authors may not agree - which is fine - but some attention to this point of view would strengthen the manuscript.

We agree with reviewer 2 that the section is too light. After reading the manuscript with this comments in mind, we realized that this section is not necessary. Our goal is to emphasize the potential benefits of working with researchers with different perspective, which we covered in the next section '*Work with the enemy*'.

"We propose that concerted collaboration between opposite sides of the same debate is the only way one could reach such far-reaching conclusions." (pp. 10-11) As-is, this statement comes across as false. Major advances in science have been made by researchers not noted for their collaboration with the opposite side (e.g. RA Fisher, Darwin). The authors' statement is also in opposition to the thesis of Hull (1988, cited above). The authors might explicitly consider counter-examples, and/or explicitly delimit the extent to which their proposal applies.

We agree that the debate among different scientific traditions is crucial for the development of science, but it is often the case that these traditions have much more in common than it is first assumed. Collaboration among opposite sides of a debate is a good way to evaluate the differences and similarities among opposite traditions, which could help in the development of the whole field. We have now clarified this in the text (l. 268-272 and 279-281).

Crowdsourcing (p. 11). This is intriguing and - like the authors - I have not heard of it applied in ecology and evolution. It would be helpful to outline a thought experiment, showing in principle how it could work.

We have provided two ways by which this could be accomplished (l. 287-290): "In the simplest case, one might ask a colleague to change labels of treatments prior to analysis (e.g. A and B instead of high and low). But blind analysis could also involve different teams of researchers applying their own statistical approaches to answer the question at hand (also called crowdsourcing [36])." We have also added references for examples in physics (l. 292). In addition, crowdsourcing sometimes is used to refer to data gather, not only data analysis. We have now clarified this in the main text and gave examples from the ecological literature (l. 292-294)

"The fallacy of factorial design." (pp. 11-12) There are many questions in ecology and evolution which cannot be addressed by factorial design, necessarily being based on observations (e.g. phylogeny, and in general perhaps the evolution of wild populations). Some attention could be given to these.

This is true. However, at the beginning of the manuscript, we stated that we do not see studies based on observation a barrier to use MH. In l. 84 we stated: “Here we exclude the common argument that ecology and evolution are less prone to strong inference than molecular biology, because of the need to rely on observational and comparative studies, which are not suitable for strong inference [5–7]. Although this seemingly attractive argument has generated much debate in our field [8,9], Chamberlain was a geologist and developed the method of multiple hypotheses with observational data in mind. Therefore, it is reasonable to suppose that Chamberlain’s method should also be appropriate for other areas of scientific inquiry in which controlled experiments are not readily available. In fact, we argue that it is the often observational nature of ecological and evolutionary research that makes the use of multiple hypotheses so important, as means of considering all plausible processes that might have produced the observed pattern.”

"Although not a law, this tension between new and old ideas essentially reflects a conflict between new and old generations." This is another rather inflammatory statement which requires further detail, discussion and references to the literature (beyond the single reference cited), and/or reference to individual cases. Otherwise, counter-examples come to mind (e.g. Lamarck, who made his most major contributions as an older man).

We agree with reviewer 2, and that is why we do not state this as a rule. We have now made this point clearer in the manuscript. We have also added a counter-example recently published in Nature showing that important contributions can be made at any point of a researcher’s career (l. 335-336 and 338-339).

Decision trees (p. 13). Like crowdsourcing (above), this is intriguing but, if included, requires coverage with further depth. How might a decision tree work for a particular example in ecology or evolution? How would it compare to a more typical approach?

We have now provided an hypothetical example of how decision trees work (l. 350-358).

"... before spending large amounts of money and years in the field." Again, some (hypothetical or real) example seems necessary here, illustrating the tension between laboratory and field work.

We have now added an example to make this point clearer (l. 366-372).

Box 1. The authors should clarify "alternative possible outcomes". For example, for continuous data (as might be used to estimate a mean), there are - in principle - infinite possible outcomes.

Good point. We have now explained that Platt gave special attention to the design of “alternatives and crucial experiments” that would exclude scientific hypothesis, not simply the design of any or all experiment designs and outcomes (l. 411-413)

"lacked any clear motivation or hypothesis" (p. 21; also Figure 1 legend). Could this be rephrased or expanded? It comes across as a negative verdict on the quality of three papers, but without justification. It occurs to me (without knowing the papers), they might have a strong implicit motivation that was simply not covered in the paper itself (e.g. a genome sequencing project, to generate a widely usable resource; other possibilities come to mind).

These three papers focused on describing a pattern and thus lacked a hypothesis. Note that we do not see this as a criticism to more descriptive studies. We have now clarified this point in the manuscript (l. 459-459 and 63-65).

Citations and/or DOI for the papers analysed in Figure 1 should be included as a spreadsheet or table as supplementary data. Optionally, I would also like this table/spreadsheet to include the authors' categorisation of the papers.

We have now provided all the papers and how we classified them in the electronic supplementary material.

Figure 1. Since these bars represent an estimate (being based on a random sample of literature), some indication of the certainty of results or precision of estimates would be useful - whether in the figure, legend or text.

Good point. We have now provided the bootstrapped 95% confidence interval in Figure 1 and Table 1.

Journal Name: Royal Society Open Science

Journal Code: RSOS

Online ISSN: 2054-5703

Journal Admin Email: openscience@royalsociety.org

Journal Editor: Emilie Aime

Journal Editor Email: emilie.aime@royalsociety.org

MS Reference Number: RSOS-160756

Article Status: SUBMITTED

MS Dryad ID: RSOS-160756

MS Title: Why are we not evaluating multiple competing hypotheses?

MS Authors: Betini, Gustavo; Avgar, Tal; Fryxell, J

Contact Author: Gustavo Betini

Contact Author Email: gsbetini@gmail.com

Contact Author Address 1: 50 Stone Road East

Contact Author Address 2: Guelph, Ontario, N1G 2W1

Contact Author Address 3:

Contact Author City: Guelph

Contact Author State: Ontario

Contact Author Country: Canada

Contact Author ZIP/Postal Code: N1G 2W1

Keywords: Platt, Chamberlain, strong inference, model selection, scientific method

Abstract: The use of multiple working hypotheses to gain strong inference is widely promoted as a means to enhance the effectiveness of scientific investigation. Only 21 of a 100 randomly selected studies from the ecological and evolutionary literature tested more than one hypothesis and only 8 tested more than two hypotheses. The surprising rarity of application of multiple working hypotheses suggests that this gap between theory and practice might reflect some fundamental issues. Here we identify several intellectual biases and practical barriers that discourage us from using multiple hypotheses in our scientific investigation. While scientists have developed a number of ways to avoid biases, such as the use of double-blind controls, we suspect that few scientists are fully aware of the potential influence of cognitive bias on their decisions and they have not yet adopted many techniques available to overcome intellectual and practical barriers in order to improve scientific investigation.

EndDryadContent