Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-394-AC6, 2021

Interactive comment on "Ideas and Perspectives: When ocean acidification experiments are not the same, reproducibility is not tested" *by* Phillip Williamson et al.

Phillip Williamson et al.

p.williamson@uea.ac.uk

Received and published: 15 January 2021

We thank RC2 for supportive and constructive comments. We respond here to the issues raised for attention.

COMMENT: . . . the content of the Williamson et al. essay could be richer, which would help it appeal to a larger portion of the scientific community. The essay currently omits important aspects of the scientific process that led to the situation described and lacks concrete suggestions for how to avoid similar situations. The authors might also more carefully examine the language they used to avoid participating in a "toxic" exchange.

RESPONSE: We welcome the overall suggestion (also made by others) that the content of the manuscript should be broadened, and we have done do so - both within the existing text and in an additional, concluding section "Wider implications". As given below, with Figure 1 appended. Yet we are wary of trying to do too much, within an Ideas and Perspectives article that has to be limited to a few pages. Instead, we hope that this manuscript will stimulate further attention to the important issues raised regarding replication and reproducibility in general, and the more specific issues relating to ocean acidification.

"5. Wider implications

"The concept of generalizability (Nosek and Errington, 2020a) would seem crucial to the broader debate on replication. Under what conditions should conclusions derived from one study be considered applicable (generalizable) to another, therefore enabling the underlying hypothesis to be tested, and potentially disproved, by the latter? The scientific benefits of that framing are greatest when the outcome of a replicability test is accepted by two research groups that initially favour different hypotheses - thereby requiring a more nuanced, non-confrontational framework for experimental planning, analysis and interpretation (Fanelli, 2018; Nosek and Errington, 2020a,b).

[Figure 1 here]

"Figure 1 provides a diagrammatic summary of these issues, with situation (a) showing close congruence between two experimental studies, carried out by two research groups. If that very close match is recognised by both groups when Study #2 is planned (following the arrangements proposed by Nosek and Errington, 2020b), the replication provides a valid test of any hypotheses arising from Study #1. In contrast, situation (b) shows a pair of studies that only partly overlap, i.e. they differ in many regards, and where prior agreement between research groups on their congruence may not have been achieved. If results from both studies in situation (b) are consistent, the generalizability of Study #1 is extended. However, if inconsistent, the generalizability of

Study #1 and Study #2 will each be constrained to its specific experimental conditions, with evidence from other studies providing the context for interpretation of the different outcomes. A range of intermediate situations between (a) and (b) can also occur.

"The above proposals for clearer "rules of engagement" for future replication studies could be greatly encouraged if research funders not only recognized that major insights can arise from closely similar or repeated work, but also required liaison between competing research teams as a condition of award in such circumstances. Our final recommendation is that high-profile publishers should give additional attention to the quality-control of potentially controversial papers, whilst also providing the opportunity for rapid, and preferably simultaneous, publication of responses by other researchers who may consider that their work has been unfairly criticized".

COMMENT: The authors might also more carefully examine the language they used to avoid participating in a "toxic" exchange.

RESPONSE: We have carefully reviewed our choice of words, and have made several edits to ensure that our text is more neutral. We respond to specific issues below.

COMMENT: 1) This essay could strengthen its arguments by better incorporating the ideas presented in Nosek and Errington (2020). . . this difference between failure and generalizability should be addressed in Williamson et al as it pertains to maintaining good norms of conduct in the field. Failure of a single study to support a hypothesis is a learning opportunity, not a reason to cast doubt on the rigor of prior work. A similar line of thinking is presented in the response by McCoy and in an Oceanography article by Busch, O'Donnell et al 2015 (http://dx.doi.org/10.5670/oceanog.2015.29).

RESPONSE: We appreciate these comments with associated recommendations, and have acted on them. Our submitted manuscript already cited Nosek and Errington (2020a,b), and more detailed discussion of their ideas is now given in the new Wider Implications section, as provided above. The reminder re the relevance of Busch et al. (2015) (now also cited) is also appreciated: those authors consider both reducible

and irreducible uncertainties, and the methods by which they might be addressed and communicated. Busch et al. also discuss the potential bias in meta-analyses arising from the non-publication of "no-effect" papers.

COMMENT: 2) The essay omits a major player that precipitated the situation described in the paper: the publisher. In considering this essay and reading the exchange between Clark et al and Munday et al, I found it shocking that the editors and Nature and the reviewers involved in peer-review efforts let the situation unfold as it did. For example, in a short essay on its website titled "Challenges in irreproducible research," (https://www.nature.com/collections/wjsrmrdnsm) Nature states "No research paper can ever be considered to be the final word" (a parallel of the sentence in Williamson et al's line 104), though this is what Nature seems to have declared when publishing an article by Clark et al titled "Ocean acidification does not impair the behaviour of coral reef fishes". Why did editors and peer-reviewers not balk at this? Exploring this question would give the Williamson et al essay greater relevance within the literature.

RESPONSE: We share these concerns, and (briefly) address the role of the publisher as a concluding remark in the Wider Implications section of.our revised manuscript, as given above. Whilst recognising that there are unavoidable imperfections and risks in the peer review process, there do seem to have been avoidable editorial failings with regard to Clark et al. (2020) and associated issues, including the long delay between the January submission of the Munday et al (2020) response and its October publication. We also point out here (but not in our manuscript) that Nature's policy regarding Correspondence currently excludes timely publication of comments relating to published material. In particular, a letter to the editor submitted in January 2020 by three of us, on the issues arising from Clark et al. (2020), was rejected as out of scope.

COMMENT: The review by Dupont makes a good point when it suggests expanding the Williamson et al essay to include another controversy that played out in Nature. Along a related line of thought, the essay also omits consideration of why scientists like Clark et al and journals like Nature might have chosen to frame their work as they did and the implications their actions have on public trust of scientific information. Exciting and controversial articles in high-profile journals reward the authors and journals with media attention and higher "scores" in algorithms that aim to characterize prominence. Neither of these metrics denotes quality or trustworthiness, two characterizes that are vital for the public to trust scientific information and advice related to carbon dioxide emissions. The Williamson et al essay is important in setting boundaries for acceptable behavior in research science, but, as it is currently written, it does not touch on the larger consequences of the unfortunate actions it focuses on.

RESPONSE: The additional discussion of other controversies is tempting, yet could be diversionary. A separate response is given to Dupont to cover that specific suggestion. We recognise the importance of the wider issue of "impact", and its implications for quality and trustworthiness, yet have decided not to develop the current manuscript in that regard.

COMMENT: 3) I would like to see the authors outline what would have been a better path for the Clark et al team at each step of the process of their work (initiation during grant writing, designing, interpreting) to avoid the situation that unfolded. Text on this idea could also include discussion of the value of institutional efforts in this process, for example the community building work done for US scientists by the Ocean Carbon Biogeochemistry program of the Woods Hole Oceanographic Institution.

RESPONSE: The new, concluding section of our revised manuscript, as provided above in response to an earlier comment, gives (brief) attention to the practices that would enable more constructive replication studies. The ideas of Nosek and Errington (2020b) are highly relevant in that regard.

COMMENT: 4) Williamson et al bring up that the use of language in Clark et al deviated from typical academic literature in its boldness. I support Williamson et al's decision to convey the tone of the Clark et al arguments. Yet, they did not explicitly call out Clark et al for conveying information in an unemotional way, and I ask Williamson et

al to consider if it would be valuable to do so. I acknowledge that emotions are not supposed to be included in papers on marine biology, but I see this Williamson et al paper as focused on the human actors doing science, which makes emotive language fair game for discussion. I also ask Williamson et al to consider the value of their use of emotional language in their essay. I see objective language as more powerful than language that relies on emotive words to emphasize a point. For example, language like "unambiguously-titled" (line 39) made me chuckle, but may be a bit too cheeky to include. Reviewer 1 also remarked on this line "For the purpose of the appearance of objectivity, I recommend removing the phrase 'an unambiguously titled' and replacing it with the phrase 'the paper titled'. It allows the reader to draw their own conclusions about the Clark et al. 2020a paper's title from the argument that you present below." In another example, language on lines 43-45 ("Since Clark et al. went to 'great lengths' to replicate earlier work yet failed to get the same results,") feels sarcastic to me as a reader, which I don't appreciate in this type of professional setting.

RESPONSE: We consider that the use of non-technical, "direct English" generally improves communication. Thus there is no need to avoid boldness, in either a scientific or non-scientific context - provided that the words used are honest and appropriate; i.e. deliver the intended meaning, and are well-justified. What should be avoided is exaggeration or inaccuracy, not supported by evidence, or personal criticisms. Such failings may, however, become more apparent when the underlying message is more directly expressed. Whilst we recognise that the same words can be read in different ways, common usage should be the default. Thus we do not consider that "unambiguously", as used here, is an emotive word (cause for amusement or other strong reaction) - nor a subjective one, as regarded by the other Anonymous Reviewer. Similarly, the sentence "Clark et al. went to 'great lengths' to replicate earlier work yet failed to get the same results" was not intended to be sarcastic. Instead it stated, in those authors' own words (as we now make clear in the revised text), that they considered they had met all criteria for hypothesis-testing replication.

COMMENT: 5) A minor point: on lines 28-29 the authors should consider using more current references to characterize the ocean acidification literature as Kroeker et al. 2013 and Wittmann and Pörtner 2013 are too old to include the vast majority of literature on species sensitivity to ocean acidification.

RESPONSE: We agree that there is much more recent ocean acidification literature, and additional reviews are now cited in the Introduction. Nevertheless, those meta-analyses are still considered valid - and there have not been more recent attempts to provide systematic syntheses of a comprehensive nature, perhaps because of the magnitude of that task. Part of the point we wished to make in the Introduction (as scene-setting for later discussion) was that knowledge of response variability is well-established for ocean acidification experiments, and therefore provides the basic context for interpreting novel or unexpected results. References identifying important factors that are known to cause such response variability were identified later (lines 127-128) in our original manuscript.

REFERENCES CITED IN RESPONSES

Busch, D. S., O'Donnell, M. J., Hauri, C., Mach, K. J., Poach, M., Doney, S.c. and Signorini, S.R.: Understanding, characterising, and communicating responses to ocean acidification: challenges and uncertainties, Oceanogr., 28(2), 30-39, http://dx.doi.org/10.5670/oceanog.2016.29, 2016.

Clark, T. D., Raby, G. D., Roche, D. G., Binning, S. A., Speers-Roesch, B., Jutfelt, F. and Sundin, J.: Ocean acidification does not impair the behaviour of coral reef fishes. Nature, 577, 370-375, https://doi.org/10.1038/s41586-019-1903-y, 2020.

Munday, P. L., Dixson, D. L., Welch, M. J., Chivers, D. P., Domenici, P., Grosell, M., Heuer, R. M., Jones, G. P., McCormick, M. I., Mark Meekan, M., Nilsson, G. E., Ravasi, T. and Watson, S.-A.: Methods matter in repeating ocean acidification studies. Nature (online), doi: 10.1038/s41586-020-2803-x, 2020.

Nosek, B. A. and Errington, T. M.: What is replication? PLoS Biol. 18, e3000691, https://doi.org/10.1371/journal.pbio.3000691, 2020a.

Nosek, B. A. and Errington, T. M.: Argue about what a replication means before you do it. Nature, 583, 518-520, https://doi.org/10.1038/d41586-020-02142-6, 2020b.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2020-394, 2020.

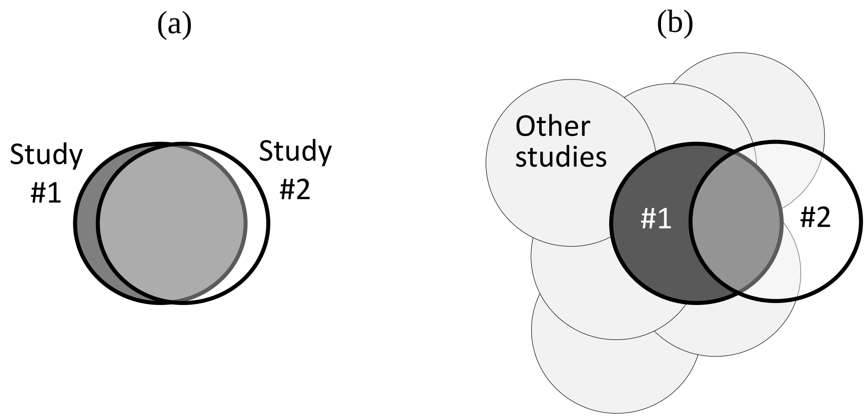

Fig. 1. Visual summary of contrasting situations relating to (a) very close matching and (b) part-matching of pairs of studies where Study #2 is intended to provide a test of repeatability of Study #1