

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

### **Anonymous Referee #2**

Received and published: 18 December 2020

Review of “Ideas and Perspectives: When ocean acidification experiments are not the same, reproducibility is not tested” by Williamson et al.

This article summarizes the debate in the literature between Clark et al and Munday et al on the sensitivity of Australian coral-reef fish behavior to ocean acidification. It does not present new data; rather it weighs in on the process of doing science. I applaud the authors for taking up the debate between Clark et al and Munday et al. It is important for senior scientists that are not conflicted to engage given the high level of attention the Clark/Munday debate has gotten inside and outside of the academic community. Disagreement is healthy to science when done constructively; Williamson et al have

C1

justly called out non-constructive behavior. I appreciate how Williamson et al framed their essay and that they confronted the controversy head on. That said, 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.

1) This essay could strengthen its arguments by better incorporating the ideas presented in Nosek and Errington (2020). The most troubling step taken by Clark et al was that they elevated their findings to a level in which their work could singularly suggest a failure of the hypothesis that fish behavior is sensitive to carbon dioxide conditions instead of it being part of a broader constellation of research used to refine the “generalizability space” (term per Nosek and Errington) of the hypothesis. I argue that 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>). As brought up in section 3 of Williamson et al, Clark et al's failure to properly account for the mechanistic explanations for observed effects of OA on fish behavior when interpreting their work is likely part of what led Clark et al to their conclusions. In doing so, they not only refute observation of the behavioral expression but the physiological work under-pinning it.

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,”

C2

(<https://www.nature.com/collections/wjsrmdnsm>) 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. 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.

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.

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 con-

C3

sider 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.

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.

---

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

C4