

## ***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 RC1 for constructive general comments and for careful identification of ~30 specific issues where improvements to the manuscript might be made. Our responses follow.

COMMENT: . . . I think that adding in a table or figure on this topic [criteria that should be used in replication versus reproducibility studies, and the implications for each] to help authors in the future to ensure that they meet these criteria in their studies would be helpful.

C1

RESPONSE: We welcome this suggestion, and have added Figure 1 (appended here) to summarise the issues relating to the match between original and replication studies, and their implications. This is for the purpose of illustrating general principles rather than identifying specific criteria to separate the concepts of replication and reproducibility studies.

COMMENT: I think that the authors should consider editing the article in its entirety to prevent using subjective language throughout. I think that this topic is important and the audience as a whole will be more receptive if the authors' suggested ideas are laid out objectively and without subjective criticism of the published work on both sides.

RESPONSE: We welcome the explicit raising of issues relating to subjectivity and objectivity, recognizing the core importance of the latter in science. In response, we have made several edits to our manuscript, to make it more “neutral” in its language. Yet we are also aware of the following:

- i) There is a timeless conundrum relating to absolute and relative truth: what is perceived as objective by one individual may be considered subjective by another. Cultural issues, semiotics (value-linked associations) and semantics (differences in word usage and intended meaning) can all affect the philological separation of objectivity and subjectivity.
- ii) The Ideas and Perspectives format provides the opportunity to explore wider considerations relating to interpretation, opinion and framing, not being limited to the (objective) presentation and discussion of factual information.
- iii) The main protagonist in this debate has had several high-profile opportunities for (subjective) communication of his opinions (e.g. Clark et al., 2016; Clark, 2017, 2020c), without the scrutiny of peer review.

We defend our own objectivity on a three-way basis: the lack of any collaborative engagement or other conflict of interests with the research groups that are in dispute;

C2

our wider involvement in national and international research assessments, evaluations and reviews; and the different disciplinary routes by which we individually acquired ocean acidification expertise.

COMMENT: Lines 17-19: Is this the authors' argument? That the replication study was "confrontational"? It seems rather subjective and less constructive than the remainder of the abstract. Rather, it would be useful to provide an assessment of how to interpret and take further actions to understand the discrepancy when they arise from different experimental conditions here to close out the abstract. Below, you provide some really interesting suggestions for criteria for categorizing replication vs reproducibility studies, and I think a sentence along those lines would be more useful here.

RESPONSE: RC1 identifies the term "confrontational" as an example of unhelpful subjectivity. We disagree: the word was carefully chosen as a factual description of the approach taken by Clark (2017, 2020) and Clark et al. (2020a,b). Being confrontational is not necessarily undesirable; in certain circumstances, such action may be fully appropriate. But in this case we consider that it has been an obstacle, not an aid, to scientific progress.

COMMENT: Line 22: Add in an "i.e." in front of "increased"

RESPONSE: Accepted.

COMMENT: Line 26: Change "very well understood" to "highly studied"

RESPONSE: There is a substantive difference in meaning between our text ("the chemistry of the carbonate system has been very well understood for decades") and the proposed edit ("the chemistry of the carbonate system has been highly studied for decades"). Presumably the reason for the proposed change is that RC1 considers that the chemistry is not "very well understood", and therefore chemical aspects of ocean acidification remain uncertain. We recognize that pH-related chemical interactions can be complex (e.g. relating to use of buffers, calibrations, constants and pH scales), and

C3

have therefore deleted "very". Nevertheless, we consider that it is important to point out that our knowledge of carbonate chemistry is much more complete than for the biological impacts of ocean acidification. Whilst this difference partly arises from the time period that the two systems have been investigated, more fundamental causes of complexity are also involved, affecting the predictability of system behaviours.

COMMENT: Line 35: Past studies have frequently used conspecific chemical alarm cues as the stimulus in these flume studies, and thus this stimulus should also be noted in the parentheses (such as in Welch et al. 2014 and Heuer et al. 2016). [Also Line 35]: I would recommend adding some examples of references here that have used this methodology so that interested readers can further read studies that have employed this methodology.

RESPONSE: We have changed "(e.g. from predators)" to "(from predators or conspecific alarm cues)", and have provided the following additional background:

"Several versions of such experimental conditions and treatments have been developed, with differences between protocols known to affect the strength of the response change (Jutfelt et al., 2017)".

The cited reference provides a review of this methodology that may be of interest to some readers.

COMMENT: Line 37: Again, provide some references here for examples when this methodology has been employed.

RESPONSE: Covered by above.

COMMENT: Line 39: 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.

RESPONSE: We consider the descriptor "unambiguously" to be objective, not subjective.

C4

tive (nor pejorative). It draws attention to an important issue: the title of Clark et al. (2020a) is part of the problem. If it had been reworded "Ocean acidification does not always impair..." or even "Behaviour of coral reef fishes unaffected by ocean acidification" it would have been less strident in its assertion that this study invalidated previous work; i.e. less confrontational. Such a change of tone would have facilitated dialogue between the research groups directly involved - and may have obviated the need for this Ideas and Perspectives manuscript. The directness of the title of Clark et al (2020a) would seem to be deliberately provocative; it was reinforced by the accompanying press release by a co-author's institution (SLU, 2020) that stated "a new study, published in Nature, shows that ocean acidification does not affect the behaviour of coral reef fish at all".

COMMENT: Line 41: Replace "they" with "the authors"

RESPONSE: "The authors" could be used instead of "They". However, we do not consider that such a change significantly either improves clarity, removes potential misunderstanding or is relevant to the subjectivity/objectivity issue.

COMMENT: Line 43: Add in the year following Clark et al. (Clark et al. 2020a)

RESPONSE: For completeness, the date has been added.

COMMENT: Line 43.45: I recommend re-writing this statement in order to summarize the assertions in Clark et al. 2020a objectively. I recommend the following edit: "Since Clark et al. 2020 claim to have attempted to replicate the results obtained from earlier work without success, they imply that the earlier work was either unreliable, flawed, or fraudulent (Clark et al. 2017)."

RESPONSE: We have made minor edits to this sentence, as follows:

"Since Clark et al. (2020a) went to "great lengths" (in their own words) to replicate earlier work yet failed to observe the same effects, there was the implication that other researchers' work was flawed or even fraudulent. The latter interpretation is consistent

C5

with earlier concerns expressed by Clark et al. (2016) and Clark (2017)".

We have retained "went to 'great lengths'" whilst now making clear that those words were used by Clark et al. (2020a) to assert that their experiments used, as far as possible, the same protocols and methods of previous studies. We therefore do not regard this as a subjectivity/objectivity issue.

COMMENT: Line 48: I recommend adding "substantial" here ("...attracted substantial media...").

RESPONSE: We do not consider "substantial" is justified. Whilst there was some media coverage, this was relatively specialised and mostly in the scientific press (e.g. Enserik 2020), without coverage by any major popular outlets (national press, TV or radio) as far as we are aware.

COMMENT: Line 48: I again recommend adding the year here ("... the apparent thoroughness of the approach described in Clark et al. (2020a)...").

RESPONSE: The date has been added.

COMMENT: Line 50: Please be clear about WHO identified these potential weaknesses. Also, as detailed in the reply by Munday et al. 2020a Nature, there were actually many more criticisms than three, so you should be clear that there were SEVERAL criticisms, and that you are highlighting three in particular.

RESPONSE: The three potential weaknesses were identified by combining quoted comments by Munday, given in a report by Enserick (2020), and independently by Widdicombe and Williamson in the collation of expert comments by the Science Media Centre (2020). All those comments were, however, relatively informal and of a rapid-response nature; they were made well before the much more comprehensive response by Munday et al. (2020), giving many more criticisms. Such information is readily available from the cited references. We do not consider it appropriate to say that we are highlighting three criticisms out of many, since we focus in this paragraph on the early

C6

response - when most of the comments were relatively favourable.

COMMENT: Line 63: I recommend editing to include the phrase "but not limited to" (as in, "...included (but not limited to) the following...") .

RESPONSE: The proposed edit is redundant: "including" has the meaning that what follows is not comprehensive, i.e. is not limited to the information given.

COMMENT: Line 89: I recommend editing to say "...avoid creating confusion..." instead of "...avoid contributing to the confusion...". The recent Clark et al. 2020a paper created some confusion about the generality of CO<sub>2</sub>-induced effects on behaviour rather than contributing to existing confusion.

RESPONSE: The confusion referred to here is not specific to the effects of ocean acidification on fish behaviour. Instead, it was intended to refer to the different usages and definitions that exist in different disciplines relating to replication/replicability, repeatability and reproducibility. To clarify, the following re-wording is now used:

"This broad definition [of replication] has merit, although consistency is needed across all disciplines . . . to avoid contributing to terminological confusion in a contested topic area".

COMMENT: Line 90-91: I suggest editing this sentence to say: "First, it is important that replication studies examine key components of the original hypotheses, such as the life-stage dependence of the response to altered CO<sub>2</sub> conditions."

RESPONSE: We have made minor edits to improve the clarity of this sentence. It is now closely similar to the suggested re-wording, as follows:

"First, it is important that the design of replication studies adequately addresses all key components of existing hypotheses; for example, the strong life-stage dependence of the response to high CO<sub>2</sub> conditions".

COMMENT: Line 91-93: Your point in this sentence is not entirely clear. I believe

C7

you are trying to say that all available evidence must be considered when evaluating potential controversies. Is that right? If so, please re-word so that aim is clear.

RESPONSE: We have re-wording this sentence to improve its clarity, as follows:

" . . . any single study does not disprove the consensus, since broadening the concept of replication has the clear corollary that novel outcomes need to be interpreted using all available lines of evidence, with awareness of both similarities and differences in relation to previous work".

COMMENT: Line 94: For non-experts on this topic, provide some context for how studies of sensory physiology link to the studies on behaviour, which are the focus here.

RESPONSE: We now say: "sensory physiology (and hence behaviour)". Whilst additional details of the linkage could be provided, we do not consider this to be necessary.

COMMENT: Line 94: I believe that you mean Table 1 in Munday et al. 2020a, not supplementary table 1. Is that correct? If so, please correct this in the text.

RESPONSE: The reference to Supplementary Table 1 is correct: it can be accessed via the link given under Supplementary Information in the online version of Munday et al (2020). That information was compiled in January 2020, and did not include any papers later than 2019. Based on searches for the key words "fish, behaviour, acidification", we have now included data on five additional 2020 publications.

COMMENT: Line 95-96: For the sake of readability, I recommend using parentheses instead of commas around the phrase summarizing the researchers, institutions, and countries involved.

RESPONSE: Accepted.

COMMENT: Line 107: Add in a period after et al.

RESPONSE: Accepted.

C8

COMMENT: Line 107-109: I recommend rewriting this sentence to the following: "Given the plethora of independent evidence, ocean acidification likely does have adverse impacts on fish behaviour. However, the resilience of fishes to altered CO<sub>2</sub> is likely to vary depending on the species and circumstances under investigation."

RESPONSE: We have re-worded as follows:

"An objective summary of the global evidence is that ocean acidification can adversely affect fish behaviour under experimental conditions, whilst also recognising that the occurrence and scale of such impacts vary with circumstances and the species tested".

COMMENT: Line 112: Add in a comma between ecosystems and which.

RESPONSE: This sentence has been re-worded; comment no longer applicable.

COMMENT: Line 113: Change good to strong.

RESPONSE: This edit is not considered necessary.

COMMENT: Line 114: Rewrite this sentence as follows: "Our increasing appreciation for the complex relationship between ocean acidification and the ocean's biochemical, physiological, behavioural and ecological interactions are both scientifically exciting and sobering..."

RESPONSE: Sentence now re-worded, as follows:

"This increasing appreciation of the interactions between ocean acidification and other biochemical, physiological, behavioural, ecological and physical processes is both scientifically exciting and sobering..."

COMMENT: Line 116-119: These two sentences are somewhat difficult to understand. Can you re-write to improve their clarity and brevity? I believe that your point here is that the complexity of the relationship between ocean acidification and natural processes is to be expected given the variability that these species would have experienced throughout their evolutionary history. If so, just say that.

C9

RESPONSE: Sentence now re-worded, as follows:

"The complexity of these interactions should, however, not be surprising, since marine species will have experienced natural variability in pH and CO<sub>2</sub> levels throughout their evolution. Species will have found different ways of responding, and response differences can therefore be expected to occur in experimental studies".

COMMENT: Line 130: The terms over-simplistic and unscientific are perhaps not the most constructive adjectives to use here. Consider changing this sentence to something like the following: "Given this known variability, the results from any one ocean acidification study will therefore be unable to overshadow all of the other previous findings from that area of research."

RESPONSE: Sentence now re-worded, as follows:

"Given this known variability, the results from any single ocean acidification study cannot provide the final word, over-riding other findings".

COMMENT: Line 131-132: I am not entirely clear what you are trying to assert with this sentence. My initial reading of this paragraph is that this sentence can be deleted in its entirety, to instead focus on how practices in the field can be improved generally, through more widespread publication of replication studies and negative results, which are notoriously difficult to publish and often end up in low impact journals, causing researchers not to prioritize dissemination of their important results.

RESPONSE: We have re-worked this paragraph, and have also included a 'Wider Implications' section to give more emphasis on how practices can be improved - whilst still noting the importance of publishing negative results.

COMMENT: Line 135-139: I am not entirely clear what you are suggesting here as an alternative approach. You suggest that the published back and forth between Clark et al. and Munday et al. is ineffective, but do not clearly layout an alternative framework for these sorts of controversies. What do you mean specifically by a "more nuanced,

C10

nonconfrontational framework"? These specifics would be useful for taking constructive steps in the future.

RESPONSE: We have now expanded and explained these ideas in a new Wider Implications section, provided below (this text provides the context for Fig 1, as appended and already mentioned in response to an earlier comment):

#### "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.

C11

"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."

#### REFERENCES CITED IN RESPONSES

Clark, T. D.: Science, lies and video-taped experiments. *Nature*, 542, 139, <https://doi.org/10.1038/542139a>, 2017.

Clark, T.D.: Is ocean acidification driving fish bonkers? *Nature Res. Ecol. Evol. Comm.* (online), <https://natureecoevocommunity.nature.com/posts/6519-is-ocean-acidification-driving-fish-bonkers>, 2020.

Clark, T.D., Binning, S.A., Raby, G.D., Speers-Roesch, B., Sundin, J., Jutfelt, F. and Roche, D.G.: Scientific misconduct: the elephant in the lab. A response to Parker et al. *Trends. Ecol. Evol.* 31, 899-900, <https://doi.org/10.1016/j.tree.2016.09.006>, 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>, 2020a.

Clark, T.D., Raby, G.D., Roche, D. G., Binning, S. A., Speers-Roesch, B., Jutfelt, F. and Sundin, J.: Reply to: Methods matter in repeating ocean acidification studies, *Nature* (online), <https://doi.org/10.1038/s41586-020-2803-x>, 2020b.

Jutfelt, F., Sundin, J., Raby, G. D., Krång, A.- S. and Clark, T.D.: Two-current choice flumes for testing avoidance and preference in aquatic animals. *Methods Ecol. Evol.*, 8, 379-390, <https://doi.org/10.1111/2041-210X.12668>, 2017.

C12

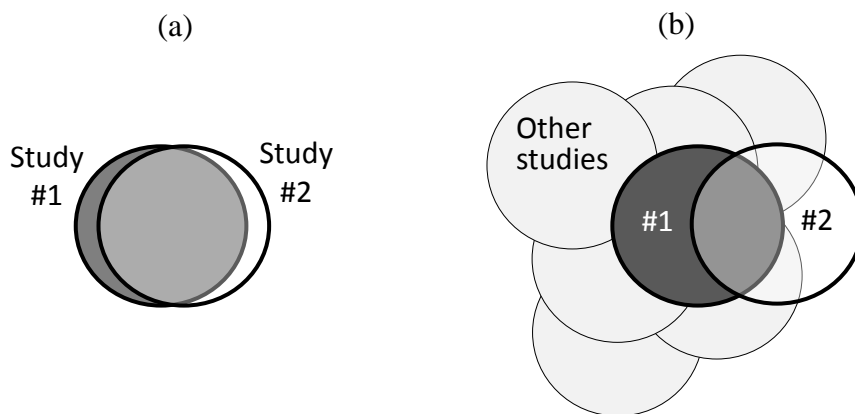
Munday, P. L., Dixon, D. L., Welch, M. J., Chivers, D. P., Domenici, P., Grosell, M., Heuer, R. M., Jones, G. P., McCormick, M. I., Mark Meekan, M., Nilsson, G. E., Ravasi, T., and Watson, S.-A.: Methods matter in repeating ocean acidification studies, Nature (online), doi: 10.1038/s41586-020-2803-x,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

Science Media Centre: Expert reaction to study looking at ocean acidification and coral reef fish behaviour. [www.sciencemediacentre.org/expert-reaction-to-study-looking-at-ocean-acidification-and-fish-behaviour](http://www.sciencemediacentre.org/expert-reaction-to-study-looking-at-ocean-acidification-and-fish-behaviour), 2020,

SLU: Ocean acidification is a big problem – but not for coral reef fish behaviour. SLU News/Swedish University of Agricultural Sciences, <https://www.slu.se/en/ew-news/2020/1/ocean-acidification-a-big-problem-but-not-for-coral-reef-fish-behaviour>, 2020.

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

C13



**Fig. 1.** Figure 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 (and gene

C14