REVISED MANUSCRIPT (31 January) – Track change

Ideas and Perspectives: When ocean acidification experiments are not the same, repeatability is not tested

Deleted: roducibility

Phillip Williamson[1], Hans-Otto Pörtner[2], Steve Widdicombe[3], Jean-Pierre Gattuso[4,5]

[1]School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

[2]Alfred Wegener Institute for Polar and Marine Research, 27515 Bremerhaven, Germany

[3]Plymouth Marine Laboratory, Plymouth, PL1 3DH, UK

[4]Sorbonne Université, CNRS, 06230 Villefranche-sur-mer, France

[5]Institute for Sustainable Development and International Relations, 75006 Paris, France

Correspondence to: Phillip Williamson (p.williamson@uea.ac.uk)

Abstract

Can experimental studies on the behavioural impacts of ocean acidification be trusted? That question was raised in early 2020 when a high-profile paper failed to corroborate previously-observed responses of coral reef fish to high CO_2. New information on the methodologies used in the "replicated" studies now provides a plausible explanation: the experimental conditions were substantially different. High sensitivity to test conditions is characteristic of ocean acidification research; such response variability shows that effects are complex, interacting with many other factors. Open-minded assessment of all research results, both negative and positive, remains the best way to develop process-based understanding. As in other fields, replication studies in ocean acidification are most likely to contribute to scientific advancement when carried out in a spirit of collaboration rather than confrontation.

Deleted: impacts on the

1. Introduction

Ocean acidification involves a reduction in seawater pH (increased hydrogen ion concentration), currently caused by increased carbon dioxide (CO_2) in the atmosphere. Associated chemical changes include increased concentration of bicarbonate ions and dissolved inorganic carbon, and reduction in the concentration of carbonate ions in the ocean and, unless compensated for, the body fluids of marine organisms. Although the chemistry of the carbonate system has been well understood for decades, research on the biological and ecological implications of anthropogenic ocean acidification only began in earnest about 20 years ago (Gattuso and Hansson, 2011). A wide range of potential consequences have since been identified, with an early appreciation of the diverse vulnerability of plant and animal species (Kroeker et al., 2013; Wittmann and Pörtner, 2013). Effects on the production of shells and skeletons have been a major research focus, but reduced

Deleted: refers to

Deleted: over an extended period, typically decades or longer,

Deleted: primarily

Deleted: uptake of

Deleted: from

Deleted: variable

40 calcification is not the only impact; there is also strong evidence for low pH affecting many other physiological processes (Pörtner et al., 2014; Baumann, 2019; Hurd et al., 2020), including vertebrate and invertebrate behaviour (Clements and Hunt, 2015; Cattano et al., 2018; Zlatkin and Heuer, 2019). Laboratory experiments have investigated the biological impacts of ocean acidification through a reductionist approach; i.e. conditions are deliberately simplified. This approach has the advantage of enabling statistical testing of cause and effect for single factors, yet necessarily omits many of the complexities

45 of natural conditions, that may involve temporal as well as biotic and abiotic environmental factors (Kapsenberg and Cyronak, 2019).

2. The challenge of contradictory results

A two-step experiment has been used by many research groups to investigate the possible effects of ocean acidification on

50 fish behaviour. Initially, individual fish are given a binary choice of water conditions in a flume tank, with one choice including an odour (e.g. from predators or a conspecific alarm cue) known to elicit an avoidance response. Those observations of discriminatory ability then provide the "control" strength of preference, for comparison with treatment results using the same choice under raised CO_2 (lowered pH) conditions throughout the test tank. Several versions of such experimental conditions and treatments have been developed, with differences between protocols known to affect the

55 strength of the response change (Jutfelt et al., 2017).

Based on that binary-choice approach and with the intention of replicating previous work, Clark et al. (2020a) reported their findings in an unambiguously-titled paper: "Ocean acidification does not impair the behaviour of coral reef fishes". To exclude the possibility of inadvertent observer bias, they deployed video recording and automatic tracking software in their

60 study, making that digital information openly available. They also used data simulations to conclude that previously reported results were "highly improbable", with an estimated likelihood of 0 out of 10,000 − assuming identical experimental conditions and that their own data were valid. 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 either flawed or fraudulent, reflecting earlier concerns expressed by Clark et al. (2016) and Clark (2017).

65

In the context of a reported "crisis" in research reproducibility for many disciplines (Baker, 2016; Nature, 2018), Clark et al. (2020a) attracted media coverage and scientific responses, including praise for its thoroughness by several independent commentators (Enserinck, 2020; Science Media Centre, 2020). However, those initial reactions also identified three potential weaknesses. First, Clark et al. (2020a) *did* find several significant ocean acidification effects, contrary to the paper's title,

70 although less dramatic than those previously reported. Second, their analysis gave scant attention to the extensive literature on factors causing variability in ocean acidification research. The third, more fundamental, concern related to how closely the original experiments had been repeated, and whether that issue had been thoroughly checked before the paper was published.

3. Experimental differences

Any deficiencies in the peer review of Clark et al. (2020a) were addressed nine months after its publication, with a detailed (online) critique by Munday et al. (2020a) that challenged the effectiveness of the claimed replication: "Clark et al. did not closely repeat previous studies, as they did not replicate key species, used different life stages and ecological histories, and changed methods in important ways that reduce the likelihood of detecting the effects of ocean acidification".

Experimental differences identified by Munday et al (2020a) between the original and repeated results included the following:

- Clark et al. (2020a) did not use clownfish, one of the original test species
- Adult and sub-adult fish were mostly used, rather than larvae and small juveniles (with older fish known to be less responsive to risk cues)
- For one species, the juveniles were from an inbred aquarium population (likely to be pre-adapted to high CO_2, and hence less sensitive)
- Many experiments were carried out during a marine heatwave (with high temperatures known to reduce or reverse responses in the studied species)
- Dissolved CO_2 levels were unstable, with average daily pCO_2 range of 581 µatm in 2016 treatments. Such variability can reduce behavioural impacts (Jarrold et al., 2017) and did not match the stable conditions of directly-compared studies.

There were also crucial changes to the design of testing apparatus; the dilution and nature of odour cues; and the duration of tests. Such changes weakened the control response, hence reducing the likelihood of significant CO_2 treatment effects. In total, 16 differences between the original studies and the re-runs were described by Munday et al. (2020a), any one of which could potentially invalidate the comparisons.

The counter-argument, made at the time of the original publication (Enserink, 2020) and subsequently re-iterated by Clark et al. (2020b), is that minor experimental differences are inevitable, and can be considered as reflecting natural environmental variability. They should not matter if the original findings are widely applicable and robust. The question of what does or does not constitute a valid replication is therefore critical, yet inherently problematic. Since it is widely accepted that a fully-exact repeat of a biological study is impossible, due to the dynamic nature of both animate and inanimate factors ("No man ever steps in the same river twice; it is not the same man, nor is it the same river"; Heraclitus), it is valid to distinguish "reproducibility" from "replicability". Whilst both terms relate to the repeatability of outcomes, the test for reproducibility is conventionally limited to conditions where very tight control is achievable, e.g. in data treatments, or when re-using the original experimental set-up. In contrast, greater flexibility is allowed for investigating replicability, reflected in a definition

of replication as "a study for which any outcome would be considered diagnostic evidence about a claim from prior research" (Nosek & Errington, 2020a). This broad definition has merit, although consistency is needed across disciplines (e.g. Stevens, 2017; Junk and Lyons, 2021), to avoid contributing to semantic confusion in a contested topic area.

120

Three further generic issues are also relevant here. First, it is important that the design of a replication study adequately addresses all key components of existing hypotheses; for example, the strong life-stage dependence of the response to high CO_2 conditions. Second, the limitations of statistical analyses need to be recognized: statistically non-significant results do not necessarily mean there is no effect (Amrhein et al.,2019). Third, any single study does not disprove the consensus, since

125 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. Munday et al. (2020a; Supplementary Table 1) identified 110 research papers published between 2009-2019 that investigated how ocean acidification might, or might not, affect the behaviour and sensory physiology of fish. Out of 44 that involved coral reef fish, 41 of those studies (carried out by 68 researchers at 35 institutions in 15 countries), reporting significant effects, including

130 several that used video recording, blind-testing, and raw-data publication. The remaining 66 papers (for other tropical, temperate and polar fish; marine and freshwater) provided additional support: 44 of those reported significant behavioural effects of ocean acidification. We are aware of five more recent publications on this topic, in addition to Clark et al. 2020a (Steckbauer et al., 2020; Rong et al., 2020; Jarrold et al., 2020; McIntosh, 2020; Roche et al., 2020); four of those reported significant effects.

135

A closely similar result was found in a meta-analysis of 95 marine and freshwater studies by Clements et al. (2020), with T.D. Clark included in the authorship team: they found that 64 of those papers reported either strong or weak behavioural effects. Whilst the proportion showing a strong effect declined over the period 2009 -2019, that decrease is unsurprising, since the early strong-effect studies were all on the most sensitive (marine) species. Additional independent evidence is

140 provided by molecular studies, showing direct effects of high CO_2 on neurotransmission in fish (e.g. Schunter et al., 2019) and other taxa (e.g. Moya et al., 2016; Zlatkin and Heuer, 2019); further biochemical and pharmacological examples are given by Munday et al. (2020a). 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, species and the life stage tested.

145

4. Taking account of response variability

A recent IPCC assessment (Bindoff et al., 2019) confirmed the pervasive and complex effects of high CO_2 and warming, not only on marine organisms and ecosystems but also on ecosystem services and society. Improved knowledge of all these response levels is crucial for effective mitigation and adaptation. This increasing appreciation of the interactions between

150 ocean acidification and other biochemical, physiological, behavioural, ecological and physical processes is both scientifically

Deleted: all

Deleted: terminological

Deleted: wo

Deleted: and the

155 exciting and sobering, showing the difficulty in developing comprehensive understanding of this important component of ocean climate change. The complexity of these interactions should, however, not be surprising, since marine species have experienced natural variability in pH and CO_2 levels throughout their evolution and also in their diverse habitats (Kapsenberg and Cyronak, 2019). Species will inherently have differently vulnerabilities and different ways of responding, and response differences can therefore be expected to occur in experimental studies.

160

Recognition of widespread response variability in ocean acidification experiments is not novel. It was noted for studies on survival, calcification, growth and reproduction in early meta-analyses (Kroeker et al., 2013), and subsequently provided the focus for much national and international research. It is therefore now well-established that closely-related marine species can respond very differently to experimental pH treatments, and that the magnitude of single species responses can be
165 affected by many factors. These influences include: length of exposure; population-level genetic differences due to local adaptation; food availability; interactions with other stressors; seasonality; energy partitioning; life stage and the sex of the organisms used in experiments (e.g. Thomsen et al., 2012; Suckling et al, 2014; Sunday et al., 2014; Breitburg et al, 2015; Vargas et al., 2017; Ellis et al., 2017; Dahlke et al., 2018); as well as physico-chemical conditions (Riebesell et al., 2011).

170 Given this known variability, the results from any single ocean acidification study cannot provide the final word, over-riding the consensus of other findings. Whilst many important uncertainties remain (Busch et al., 2016; Baumann, 2019; Hurd et al., 2020), we consider that scientific progress can be hindered by the sequence of polarising criticisms (Clark 2017, Clark et al 2020a), rebuttal (Munday et al 2020a), reply (Clarke et al.2020b) and a further point-by-point response (Munday et al. 2020b). A more constructive approach would involve experimental co-design in a collaborative, comparative framework
175 (Boyd et al., 2018), with appropriate rigour (Cornwall and Hurd, 2016) – that can still be consistent with scientific scepticism, replication tests and the reporting of negative results (Browman, 2016). Future ocean acidification experiments would also benefit from an updating of Riebesell et al. (2011), to provide improved guidance on the key parameters that can affect laboratory results. Since a very wide range of factors are potentially important, pragmatism will be needed with regard to associated issues of resource deployment and measurement accuracy, recognizing that chemists and biologists may
180 have different priorities on such matters.

5. Wider implications

The concept of generalizability (Nosek and Errington, 2020a) seems crucial to the broader debate on replication. Under what conditions should conclusions derived from one study be considered applicable (generalizable) to another, therefore
185 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).

195

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 both groups recognize that there is a very close match 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.

200 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.

205
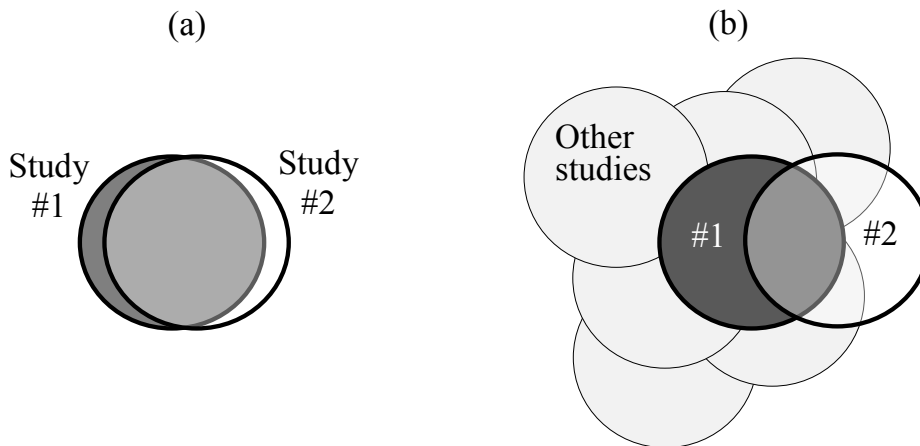
Figure 1: Visual summary of contrasting situations relating to (a) very close matching and (b) part-matching of pairs of studies where Study #2 is intended to provide a test of repeatability (and generalizability) of Study #1. Whilst 'other studies' are also relevant to situation (a), their importance is increased when interpreting results from situation (b). See text for more detailed explanation and discussion, including the importance of experimental co-design between research groups with contradictory hypotheses.

210

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.

220 **Author contributions:** PW prepared the original draft with subsequent input and editing by all co-authors.

Competing interests: The authors declare that they have no conflicts of interest.

Acknowledgements: We are grateful for the constructive comments by Tyler Cyronak, Christian Duarte, Sam Dupont,
225 Sophie McCoy, Martin Solan and two anonymous referees that have significantly improved this paper.

References

Amrhein, V., Greenland, S. and McShane, B.: Scientists rise up against statistical significance. Nature, 567, 305-307, https://doi.org/10.1038/d41586-019-00857-9, 2019.

230 Baumann, H.: Experimental assessments of marine species sensitivities to ocean acidification and co-stressors: how far have we come? Can. J. Zool., 97, 399-408, https://doi.org/10.1139/cjz-2018-0198, 2019.

Bindoff, N. L., Cheung, W. W. L., Kairo, J. G., Arístegui, J., Guinder, V. A., Hallberg, R., Hilmi, N., Jiao, N., Karim, M. S., Levin, L., O'Donoghue, S., Purca Cuicapusa, S. R., Rinkevich, B., Suga, T. Tagliabue A., and Williamson, P.: Chapter 5. Changing ocean, marine ecosystems, and dependent communities, in: IPCC Special Report on the Ocean and Cryosphere
235 in a Changing Climate, edited by Pörtner, H.-O., Roberts, D. C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K.,Alegría, A., Nicolai, M., Okem, A., Petzold, J.,Rama, B., and Weyer, N. M., Intergovernmental Panel on Climate Change, https://www.ipcc.ch/srocc/chapter/chapter-5, 2019.

Boyd, P. W., Collins, S., Dupont, S., Fabricius, K., Gattuso, J.-P., Havenhand, J., Hutchins, D. A., Riebesell, U., Rintoul, M. S., Vichi, M., Biswas, H., Ciotti, A., Gao, K., Gehlen, M., Hurd, C. L., Kurihara, H., McGraw, C. M., Navarro, J. M.,
240 Nilsson, G. E., Passow, U., and Pörtner, H.-O.: Experimental strategies to assess the biological ramifications of multiple drivers of global ocean change – A review, Glob. Change Biol., 24, 2239-2261, https://doi.org/10.1111/gcb.14102, 2018.

Breitburg, D. L., Salisbury, J., Bernhard, J. M., Cai, W. J., Dupont, S., Doney, S. C., Kroeker, K. J., Levin, L. A., Long, W. C., Milke, L. M., and Miller, S. H.: And on top of all that… coping with ocean acidification in the midst of many stressors. Oceanogr., 28, 48-61, https://doi.org/10.5670/ oceanog2015.31, 2015.

245 Browman, H. I.: Applying organized scepticism to ocean acidification research, ICES J. Mar. Sci., **73**, 529-536, https:/doi.org/10.1093/icesjms/fsw010, 2016.

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, 30-39, http://dx.doi.org/10.5670/oceanog.2016.29, 2016.

250 Cattano, C. J., Claudet, J., Domenici, P., and Milazzo, M.: Living in a high CO_2 world: A global meta-analysis shows
    multiple trait-mediated fish responses to ocean acidification, Ecol. Monogr., 88, 320-335,
    https://doi.org/10.1002/ecm.1297, 2018.

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

Clark, T.D., Binning, S.A., Raby, G.D., Speers-Roesch, B., Sundin, J., Jutfelt, F., and Roche, D.G.: Scientific misconduct:
255   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.

260 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.

Clements, J. C. and Hunt, H. L.: Marine animal behaviour in a high CO_2 ocean, Mar. Ecol. Progr. Ser., 536, 259-279,
    https://doi.org/10.3354/meps11426, 2015.

Clements, J. C., Sundin, J., Clark, T. D., and Jutfelt, F. : An extreme decline effect in ocean acidification fish ecology,
265   EcoEvoRxiv [preprint], https://doi.org/10.32942/osf.io/k9dby, 17 September 2020.

Dahlke, F. T., Butzin, M., Nahrgang, J., Puvanendran, V., Mortensen, A., Pörtner, H.- O., and Storch, D.: Northern cod
    species face spawning habitat losses if global warming exceeds 1.5°C, Sci. Adv., 4, eaas8821,
    https://doi.org/10.1126/sciadv.aas8821, 2018.

Ellis, R. P., Davison, W., Queirós, A. M., Kroeker, K. J., Calosi, P., Dupont, S., Spicer, J. I., Wilson, R. W., Widdicombe, S.,
270   and Urbina, M. A.: Does sex really matter? Explaining intraspecies variation in ocean acidification responses, Biol.
    Lett., 13, 20160761, https://doi.org/10.1098/rsbl.2016.0761, 2017.

Enserink, M.: Study disputes carbon dioxide-fish behavior link, Science, 367. 128-129,
    https://doi.org/10.1126/science.367.6474.128, 2020.

Fanelli, D.: Is science really facing a reproducibility crisis, and do we need it to? Proc. Natl. Acad. Sci. USA, 115, 2628-
275   2631, https://doi.org/10.1073/pnas.1708272114, 2018.

Gattuso, J.-P., and Hansson, L. (Eds.): Ocean Acidification. Oxford University Press, Oxford, 326 pp, 2011.

Hurd, C.L., Beardall, J., Comeau, S., Cornwall, C.E., Havenhand, J.N., Munday, P.L., Parker, L.M., Raven, J.A., and
    McGraw, C.M.: Ocean acidification as a multiple driver: how interactions bewteeen changing seawater carbonate
    parameters affect marine life, Mar. Freshwater Res., 71, 263-274, https://doi.org/10.1071/MF19267, 2020.

280 Jarrold, M. D., Humphrey, C., McCormick, M. I., and Munday, P. L..: Diel CO_2 cycles reduce severity of behavioural
    abnormalities in coral reef fish under ocean acidification, Sci. Rep., 7, 10153, https://doi.org/10.1038/s41598-017-10378-
    y, 2017.

Jarrold, M. D., Welch, M. J., McMahon, S. J., McArley, T., Allan, B. J. M., Watson, S.-A., Parsons, D. M., Pether, S.M.J., Pope, S., Nicol, S., Smith, N., Herberet, N., and Munday, P. L.: Elevated CO_2 affects anxiety but not a range of other behaviours in juvenile yellowtail kingfish, Mar. Env. Res., 157, 104863, https://doi.org/10.1016/j.marenvres.2019.104863, 2020.

Junk, J. R., and Lyons, L.: Reproducibility and replication of experimental particle physics results, arXiv [preprint], arXiv:2009.06864(physics.data-an), 8 January 2021.

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.

Kapsenberg, L. and Cyronak, T.: Ocean acidification refugia in variable environments. Glob. Change Biol., 25, 3201-3214, https://doi.org/10.1111/gcb.14730, 2019.

Kroeker, K. J., Kordas, R. L., Crim, R., Hendriks, I. E., Ramajo, L., Singh, G. S., Duarte, C. M., and Gattuso, J.-P.: Impacts of ocean acidification on marine organisms: quantifying sensitivities and interaction with warming, Glob. Change Biol., 19, 1884-1896, https://doi.org/10.1111/gcb.12179, 2013.

McIntosh, E. The Effect of Environmental Stressors on the Development and Behaviour of Larval *Oryzias latipes*, Honours Thesis, University of Winnipeg, Winnipeg, Canada, http://hdl.handle.net/10680/1804, 2020.

Moya, A., Howes, E.L., Lacoue-Labarthe, T., Forêt, S., Hanna, B., Medina, M., Munday, P. L., Ong, J. S., Teyssié, J. L., Torda, G., Watson, S.-A., Miller D. J., Bijma J. & Gattuso J.-P.: Near-future pH conditions severely impact calcification, metabolism and the nervous system in the pteropod *Heliconoides inflatus*, Glob. Change Biol., 22, 3888-3900, https://doi.org/10.1111/gcb.13350, 2016.

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, 2020a.

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.: Additional response (online), https://doi.org/10.25903/7rz7-4640, 2020b.

Nature: Challenges in Irreproducible Research, Special Collection (online), www.nature.com/collections/prbfkwmwvz, 2018.

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.

Pörtner, H.-O., Karl, D. M., Boyd, P. W., Cheung, W. W. L., Lluch-Cota, S. E., Nojiri, Y., Schmidt, D. N., and Zavialov, P. O.: Ocean systems, in: Climate Change 2014: Impacts, Adaptation, and Vulnerability, Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate

Change, edited by Field, C. B., Barros, V. R., Dokken, D. J. Mach, K. J., Mastrandrea, M. D., Bilir, T. E., Chatterjee, M., Ebi, K. L., Estrada, Y. O., Genova, R. C.,Girma, B., Kissel, E. S. Levy, A. N., MacCracken, S., Mastrandrea, P. R., and White, L. L., Cambridge University Press, Cambridge, UK, 411-484, 2014.

320 Riebesell, U., Fabry, V. J., Hansson, L., and Gattuso, J.-P.: Guide to Best Practices for Ocean Acidification Research and Data Reporting, European Commission, Brussels, http://dx.doi.org/10.2777/66906, 2011.

Roche, D. G., Amcoff, M., Morgan, R., Sundin, J., Andreassen, A. H., Finnøen, M. H., Lawrence, M. J., Henderson, E., Norin, T., Speers-Roesch, B., Brown, C., Clark, T.D., Bshary, R., Leung, B., Jutfelt, F., and Binning S.A.: Behavioural lateralization in a detour test is not repeatable in fishes, Anim. Behav. 167, 55-64,

325 https://doi.org/10.1016/j.anbehav.2020.06.025, 2020

Rong, J., Tang, Y., Zha, S., Han, Y., Shi, W., and Liu, G.: Ocean acidification impedes gustation-mediated feeding behaviour by disrupting gustatory signal transduction in the black sea bream, *Acanthopagrus schlegelii,* Mar. Env. Res. 162, 195182, https://doi.org/10.1016/j.marenvres.2020.105182, 2020.

Schunter, C., Ravasi, T., Munday, P. L., and Nilsson, G. E.: Neural effects of elevated CO_2 in fish may be amplified by a

330 vicious cycle, Cons. Physiol., 7, coz100, https://doi.org/10.1093/conphys/coz100, 2019.

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, last access 20 October 2020.

Steckbauer, A., Diaz-Gil, C., Alós, J., Catalán, I. A. and Duarte, C. M.: Predator avoidance in the European seabass after

335 recovery from short-term hypoxia and different CO_2 conditions, Front. Mar. Sci. 5, 350, https://doi.org/10.3389/fmars.2018.00350, 2018.

Stevens, J. R.: Replicability and reproducibility in comparative psychology. Front. Psychol., 8, 862, https://doi.org/10.3389/fpsyg.2017.00862, 2017.

Suckling, C. C., Clark, M. S., Richard, J., Morley, S. A., Thorne, M. A., Harper, E. M., and Peck, L. S.: Adult acclimation to

340 ambient temperature and pH stressors significantly enhances reproductive outcomes compared to short-term exposure, J. Anim. Ecol., 84, 773-784, https://doi.org/10.1111/1365-2656.12316, 2014.

Sunday, J. M., Calosi, P., Dupont, S., Munday, P. L., Stillman, J. H., and Reusch, T. B.: Evolution in an acidifying ocean. Trends Ecol. Evol., 29, 117-125, https://doi.org/10.1016/j.tree.2013.11.001, 2014.

Thomsen, J., Casties, I., Pansch, C., Körtzinger, A., and Melzner, F.: Food availability outweighs ocean acidification in

345 juvenile *Mytilus edulis*: laboratory and field experiments. Glob. Change Biol., 19, 1017-1027, https://doi.org/10.1111/gcb.12109, 2012.

Vargas, C. A., Lagos, N. A., Lardies, M. A., Duarte, C., Manríquez, P. H., Aguilera, V. M., Broitman, B., Widdicombe, S., and Dupont, S.: Species-specific responses to ocean acidification should account for local adaptation and adaptive plasticity, Nature Ecol. Evol., 1, 1-7, https://doi.org/10.1038/s41559-017-0084, 2017.

350   Wittmann, A. C., and Pörtner, H. O.: Sensitivities of extant animal taxa to ocean acidification, Nature Clim. Change, 3, 995-
       1001, https://doi.org/10.1038/nclimate1982, 2013.

       Zlatkin, R. L., and Heuer R. M.: Ocean acidification affects acid–base physiology and behaviour in a model invertebrate, the
       California sea hare (*Aplysia californica*), Roy. Soc. Open Sci., 6, 191041. http://dx.doi.org/10.1098/rsos.191041, 2019.