This document provides detailed replies to Clark et al.'s responses to each of the 16 points raised by Munday et al. 2020 (doi.org/10.1038/s41586-020-2803-x) in their Matters Arising entitled "*Methods matter in repeating ocean acidification studies*". Additional detail and clarification that was not possible in the Matters Arising published by *Nature* is provided here. Original comments by Munday et al. in the Matters Arising are shown in italics. Clark et al.'s response is in standard text and our reply to their response is in bold italics.

Since our Matters Arising was prepared, *Nature* has published an important comment on how attempts to replicate previous research should be coordinated. To reduce conflict and advance knowledge in the field, Nosek and Errington (2020) argue that replicators and original researchers should discuss the methods and approach to be used, before the replication study is done, so that both sides can reach agreement on the study design from the start. We agree whole-heatedly, and argue that this should be the standard for ethical replication in scientific research. We would have welcomed communication with Clark et al. about the methods they proposed to use and how these differed from our previous experiments. We would have valued the opportunity to find common ground regarding the best approach to replicating our earlier studies. We would have willingly supplied fish from our breeding populations of clownfish (*Amphiprion percula*), so they could have used this key species in their study, as well as the damselfish *Acanthochromis polyacanthus*, to prevent the use of an inbred population from a public aquarium. Unfortunately, Clark and colleagues did not initiate communication with us about their attempt to replicate past studies, neither before they began the studies, nor when they got the initial results in the first year, nor during the subsequent two years when they repeated their experiments. In 2020, Clark et al. published their paper stating that the results of our past studies could not be repeated, that there were no effects of ocean acidification (OA) on reef fish behaviour, and that the results of six earlier studies were statistically improbable based on a comparison of variances obtained with their methods and fish populations. The accusatory tone of Clark et al. (2020) resulted in our rebuttal in a Matters Arising commentary in *Nature*, describing the many crucial ways that their experiments differed from the experiments preformed in previous studies. These differences can explain why their results differed from our earlier findings and, thus, why their direct statistical comparisons of data distributions are invalid. We ask readers to think critically about whether one single study with vast methodological differences, refutes more than a decade of research on this topic by many different researchers, not only from our group, but from other groups as well.

References

Clark et al. (2020). Ocean acidification does not impair the behaviour of coral reef fishes. Nature, https://doi.org/10.1038/s41586-019-1903-y.

Munday et al. (2020). Methods matter in repeating ocean acidification research. Nature, https://doi.org/10.1038/s41586-020-2803-x

Nosek, B.A. and Errigton, T.M. (2020). The best time to argue about what a replication means? Before you do it. *Nature* 583, 518-520 (2020) doi: 10.1038/d41586-020-02142-6

Response to risk-cue (predator or alarm cue)

1. "*Three of the key papers[4-6] with which Clark et al.[1] compared their results tested larval and naïve juvenile clownfish. Clark et al.[1] did not test clownfish and, therefore, cannot claim to have repeated these studies. At least three other studie[6-8] have confirmed the previously described results in clownfish. Furthermore, in of of the studies[8], feeding strikes were recorded and the data were extracted by researchers who were blind to treatment*."

Clark et al. response

Munday et al. suggest that the absence of clownfish in our study is a primary reason why we were unable to replicate their previous findings. They cite six papers co-authored by Munday as evidence to support their findings for clownfish. Clownfish are a subfamily (Amphiprioninae) of fishes in the damselfish (Pomacentridae) family; we included six species of the latter family in Clark et al. (2020). Notably, wild-caught damselfish (*Pomacentrus wardi*) were studied alongside clownfish (*Amphiprion percula*) in one of the papers cited above by Munday et al. (Munday et al., 2010). The results were essentially identical between these two Pomacentrid species (Fig. 1a-b of our main document, and Part B Paper 3 below). Munday and colleagues frequently argue the generality of their results across fish and even invertebrate taxa (e.g., Munday et al., 2010; Lönnstedt et al., 2013; Watson et al., 2014; Dixson et al., 2015). Based on this reasoning, they should indeed expect that the effects of OA would apply to confamilial species of damselfish.

In this context, other foundational papers using two-current choice flumes similarly reported severe disruption of predator chemical cue avoidance following 4+ days of high CO_2 exposure in a variety of coral reef fishes. These include species such as the coral grouper *Plectropomus leopardus* (Family Serranidae) (Munday et al., 2013) (Fig. 1c of our main document, and Part B Paper 5), two other species of damselfish (*Dascyllus aruanus*, *Pomacentrus moluccensis*, both tested in Clark et al. 2020) (Fig. 1d-e, g-h, and Part B Paper 6), and two species of cardinalfish (Family Apogonidae: *Apogon cyanosoma, Cheilodipterus quinquelineatus*) (Munday et al., 2014) (Part B Paper 6). Thus, there is substantial overlap between the species tested in Clark et al. (2020) and those tested previously, as well as broad support in the literature by Munday and colleagues that the behavioural effects of CO_2 on coral reef fishes are not specific to any one species or family (reviewed in Clements and Hunt, 2015).

Munday et al. reply

Contrary to Clark et al.'s conclusion that "they should indeed expect that the effects of OA would apply to confamilial species of damselfish", it is already well known that behavioural effects of OA vary greatly among confamilial species. This was clearly described by Ferrari et al. (2011), who found a gradient in sensitivity to behavioural effects of OA among four closely related species of damselfishes from the genus Pomacentrus. Two species exhibited much larger behavioural changes than the others, and one species was relatively unaffected by elevated CO_2. This confamilial variation in behavioural sensitivity to OA among damselfishes was further confirmed by McCormick et al. (2013). Furthermore, both these studies were done with the observers blinded to the treatments. Therefore, confamilial variation in sensitivity to OA among damselfishes has been known for nearly a decade.

The clownfish (Amphiprion percula) has repeatedly and consistently been shown to be sensitive to behavioural effects of OA, including in recent studies with blinded observers and video recorded trials (Munday et al. 2016, McMahon et al. 2018). Three of the six earlier papers criticized by Clark et al. used clownfish. Inexplicably, Clark et al. 2020 did not test clownfish in any of their experiments, despite this species being readily available from commercial collectors and breeders. They cannot claim to have repeated past experiments if they did not use the same species as used in the previous studies. Equally, they cannot claim to have data contrary to the repeated and consistent evidence for significant behavioural effects of elevated CO_2 on clownfish from studies over a 11 year period (Munday et al. 2009, 2010, 2016, Dixson et al. 2012, Simpson et al. 2011, Nilsson et al. 2012, Jarrold et al. 2017, McMahon et al. 2018).

2. "*Clark et al.[1] did not use the life stages and ecological histories of fish species used in previous studies. They tested adults, sub-adults and some reef-resident juveniles. All previous studies considered by Clark et al.[1], with the exception of one at a CO_2 seep[9], used larvae and small juveniles, which were naïve to reef-based cues and which were either collected in light traps or reared in the laboratory. The response of naïve larvae and juveniles to risk cues is different to adults and to juveniles that have previously been exposed to risk cues. Indeed, it is already known that previous exposure to risk cues mitigates the magnitude of behavioural impairment in ocean acidification conditions[10].*"

See response to point 3.

3. "*The only species Clark et al.[1] collected in light traps had not previously been tested for CO_2 sensitivity, and variation between different species in behavioural sensitivity to ocean acidification is already well known[3,11].*"

(Combined response to points 2 and 3).

Clark et al. response

We went to great lengths to include as many individuals, species and life stages as possible given our time and equipment constraints. This includes >900 individuals of six species over three years from adults, sub-adults, and reef-resident juveniles, and one species of naïve (pre-settlement) larvae caught in light traps. Four out of the six species tested in Clark et al. (2020) have previously been reported by Munday and colleagues to show behavioural impairments from CO_2 exposure (*D. aruanus*, *P. moluccensis*, *Pomacentrus amboinensis* and *Acanthochromis polyacanthus* (used in: Ferrari et al., 2012a; Ferrari et al., 2012b; Munday et al., 2014; Welch et al., 2014)). In fact, strong effects of CO_2 on behaviour have been reported by Munday and colleagues in almost all of their papers covering a multitude of species (e.g., damselfishes, cardinalfishes, groupers, sharks, and marine snails) from both wild and captive-reared populations (Munday et al., 2010; Munday et al., 2013; Munday et al., 2014; Watson et al., 2014; Dixson et al., 2015). Of the 38 studies on coral reef fish behaviour authored by Munday and colleagues in their Supplementary Table 1, 37 of them reported an effect of elevated CO_2. Thus, it is unrealistic to argue that, by chance, we selected species and individuals within a species that were behaviourally tolerant of elevated CO_2 when the behavioural impairments are reported to be widespread.

Munday et al. state that the response of adults and juveniles pre-exposed to risk cues is different to that of naïve fish. However, the strong effects of CO_2 on fish behaviour reported by Munday and colleagues have included adult fishes as well as reef-resident juveniles that were pre-exposed to risk cues (e.g., Devine et al., 2012; Munday et al., 2014; Dixson et al., 2015; Heuer et al., 2016). For example, Munday et al. (2014) reported from choice flume experiments that the behavioural impairments of reef-resident juvenile damselfish and cardinalfish from OA exposure at CO_2 seeps are just as extreme as those reported for naïve larval fish (e.g., cf. Fig. 1a-b vs. 1d-e of our main document). We also note that Munday and colleagues have reported effects of CO_2 on the behaviour of adult fish of several species when using techniques other than fluming, including on antipredator responses measured by quantifying changes in activity levels in response to predator cues (e.g., see Cripps et al., 2011; Ferrari et al., 2011b; Devine et al., 2012; Dixson et al., 2015). These studies indicate that there is no evidence in the previous publications by Munday and colleagues that older fish have greater behavioural resilience to OA.

To illustrate that neither species nor life stage explain why our findings contradict those of Munday and colleagues, we have provided a side-by-side comparison of data from Munday and colleagues (Fig. 1d-e of our main document) versus those presented in Clark et al. (2020) (Fig. 1g-h), when standardising for species and life stage.

Munday et al. reply

Clark et al.'s own response illustrates that their results are anomalous – as they point out, more than 40 previous studies, by more than 20 different lead authors (with > 45 different co-authors), have demonstrated significant effects of elevated CO_2 on the behaviour of coral reef fishes (Supplementary Table 1). Yet, Clark et al. claim there are no effects of OA on the behaviour of coral reef fishes. Are they trying to argue that all of these other studies by many different authors over an 11 year period are incorrect? The most parsimonious explanation is that Clark et al. (2020) used vastly different methods to other studies, which explains why they did not detect effects that many others have observed.

Clark et al. state that significant effects reported in previous studies on adult fishes "indicate that there is no evidence that older fish have greater behavioural resilience to OA". However, most studies with adult fishes have used very different techniques to those with larval and juvenile fishes and therefore a direct comparison of between age groups is often not possible. Nevertheless, the few past studies with adult coral reef fishes that use similar two-choice flume methods to the previous studies with larvae and small juveniles do report a smaller magnitude of effect (Cripps et al. 2011, Heuer et al. 2016), suggesting that adults may indeed be more tolerant to OA conditions.

As noted in Point 2, above, all previous studies considered by Clark et al., with the exception of Munday et al. (2014) at a natural CO_2 seep, used larvae and small juveniles that were naïve to reef-based cues, either collected in light traps or reared in the laboratory. Ferrari et al. (2017) showed that prior exposure to risk cues can greatly reduce the effects of elevated CO_2 on the behavioural response to risk cues. Naïve juveniles have impaired response to risk cues in elevated CO_2, whereas juveniles with repeated past experience of risk cues do not. With the exception of one species that has not previously been tested, Clark et al.'s large sample size consisted exclusively of fish that had prior exposure to reef-based risk cues. Therefore, Clark et al. tested animals that are less likely to exhibit consistent effects of elevated CO_2 on the response to predator odour compared with earlier studies.

Clark et al. note that Munday et al. (2014) reported impaired behaviours in small reef-resident juveniles collected at a CO_2 seep. In that same study, Munday et al. also reported that predator density was lower at the CO_2 seep compared with nearby control reefs. Therefore, the small juveniles from the CO_2 seep study came from a relatively low risk environment, which might explain why they still exhibited impaired response to predator cue. Nagelkerken et al. (2016) observed a marked reduction in startle distance (a measure of risk aversion) in small fishes at two other CO_2 seeps, which is broadly consistent with the altered predator avoidance behaviours reported by Munday et al. (2014) at the coral reef CO_2 seep.

4. "*The ocean acidification chemical conditions in experiments at Lizard Island described by Clark et al.[1] did not meet the necessary standards of stability. The average (±s.d.) within-day pCO_2 range of 581 µatm ± 508 in their CO_2 treatment in 2016 is probably sufficient to diminish the behavioural effects[7] of elevated CO_2, especially in combination with the high temperatures that occurred in their experiment (Supplementary Information and Supplementary Table 2).*"

Clark et al. response

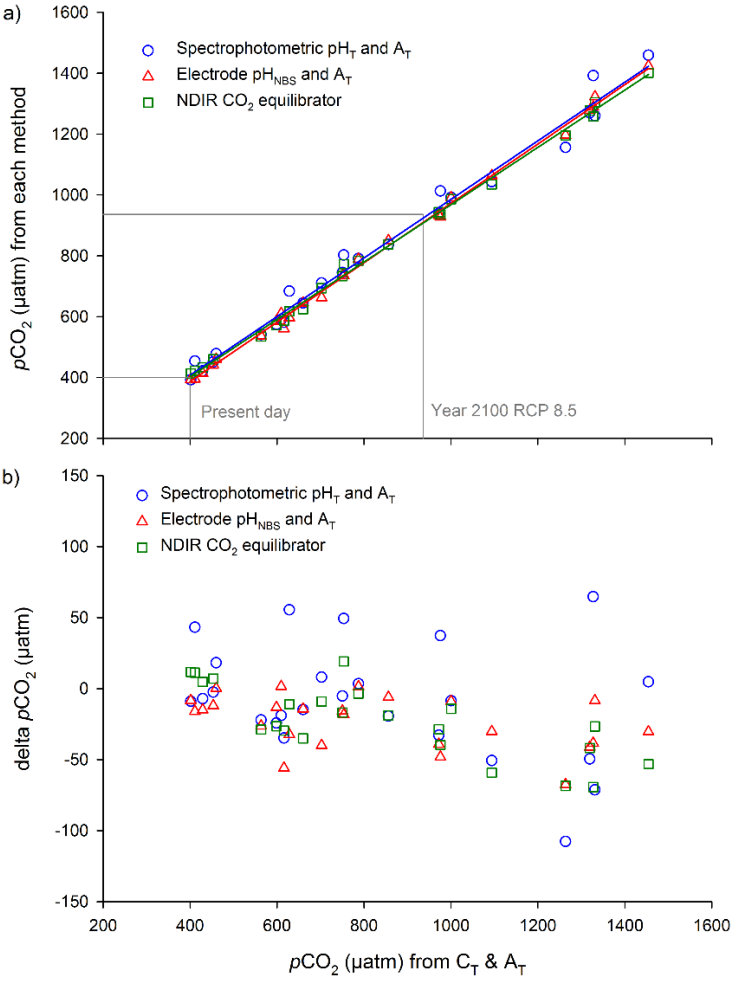
We took care to regularly measure pCO_2 (i.e., direct measurements of pCO_2) in our holding tanks and experimental arenas, in contrast to measuring pH with NBS-calibrated probes to calculate pCO_2 (i.e., indirect measurements) as done by Munday and colleagues in many of their studies (e.g., Munday et al., 2009; Dixson et al., 2010; Ferrari et al., 2012b; Nilsson et al., 2012; Lönnstedt et al., 2013; Chivers et al., 2014a; Chung et al., 2014; Domenici et al., 2014; Dixson et al., 2015; McMahon et al., 2018). Several prominent papers describe in detail why using NBS-calibrated pH probes to calculate pCO_2 in seawater is problematic (Riebesell et al., 2010; Moran, 2014; Bockmon and Dickson, 2015). Note that pH measurements made in seawater using conventional probes can be as far as 0.1 pH units off (equivalent to an error in pCO_2 of ~200 µatm in coral reef environments) when measured by specialists at state-of-the-art laboratories (Bockmon and Dickson, 2015), and this error is likely to be much higher in ecological experiments with non-specialists using less sophisticated pH measuring equipment. Such measurement error introduces a far greater uncertainty in the pH-calculated estimate of pCO_2 compared with measuring pCO_2 directly.

Munday et al. reply

We deal with Clark et al.'s responses paragraph-by-paragraph here because each paragraph contains factual errors and/or misleading statements that need to be corrected.

In their response, above, Clark et al. imply that their measurements of pCO_2 are superior to previous studies by Munday and colleagues, because they measured pCO_2 directly using non-dispersive infrared (NDIR) technology instead of the standard approach in OA research of estimating pCO_2 from measurements of pH and total alkalinity (A_T). There are a number of problems with their argument. First, measuring pCO_2 directly with a relatively low resolution instrument (Valisala GMT222) like Clark et al. have done is most definitely not more accurate than careful measurements of pH and alkalinity to estimate pCO_2. In fact, pCO_2 estimated by careful use of glass electrodes to measure pH_{NBS} in conjunction with A_T has an almost identical measurement uncertainty as direct measurement of pCO_2 using a high performance Vaisala GMP343. Each method has an error of 3.5-3.6% compared with pCO_2 estimated from total carbon (CT) and A_T (Watson et al. 2017). Moreover, Clark et al. used a low resolution Vaisala GMT 222 (±75 ppm + 2% of reading for 0-5000 ppm range) instead of the more accurate Vaisala GMP343 (±5 ppm + 2% of reading for 0-5000 ppm range), which would substantially increase the error in their measurements.

In their response, above, Clark et al. also neglect to report that Munday and colleagues used NDIR to cross-validate estimated pCO_2 from pH_{NBS} and A_T. In other words, Munday et al. cross checked their estimates of pCO_2 derived from standard carbonate chemistry methods with the NDIR method advocated by Clark et al. This cross validation is reported in previous studies, but mysteriously ignored by Clark et al. in their responses. As Clark et al. are aware from previous correspondence in Journal of Experimental Biology, our estimates of pCO_2 from carbonate chemistry are within a few µatm of measurements with high performance NDIR using a Vaisala GMP343 (see Munday, P.L., Watson, S.A., Chung, W.S., Marshall, N.J., Nilsson, G.E. (2014). Response to "The importance of accurate CO_2 dosing and measurement in ocean acidification studies'. Journal of Experimental Biology, 217: 1828-1829).



Seawater pCO_2 calculated from CT and A_T, compared with three other methods: 1) spectrophotometric pH_T and A_T (n = 25), 2) electrode pH_{NBS} and A_T (n = 25), and 3) direct measurement of seawater CO_2 with high resolution NDIR using a Vaisala GMP343 (n = 23); a) for pCO_2 data and b) for the difference in pCO_2 compared to pCO_2 derived from CT and A_T. The figure shows the three methods yield values of pCO_2 values close to those from the gold-standard of CT and A_T.

From: Watson et al. (2017). Quantifying pCO_2 in biological ocean acidification experiments: a comparison of four methods. PLoS One, 12: e0185469. https://doi.org/10.1371/journal.pone.018

Clark et al. response

Seawater at the Lizard Island Research Station (where our 2014 and 2016 experiments were conducted) enters the aquarium system directly from the reef, and thus natural, temporal fluctuations in pCO_2 are expected. Additionally, the CO_2 dosing systems (Aqua Medic AT-Control System) used in our study and in studies by Munday and colleagues rely on feedback from pH probes to trigger solenoids and diffuse CO_2 gas into the experimental water. These systems are useful for applications in ocean acidification biology, but they are not infallible and can occasionally overshoot pCO_2 targets for short periods (minutes). By measuring pCO_2 directly and frequently, we achieved higher temporal resolution and captured natural and dosing-related daily fluctuations in pCO_2, which we report transparently. It would be unrealistic to argue that such pCO_2 fluctuations were not present in earlier studies by Munday and colleagues, many of which were also carried out at the Lizard Island Research Station in the same holding aquaria and using the same CO_2 dosing methods. It is more probable that Munday and colleagues were not aware of these pCO_2 fluctuations because they did not measure pCO_2 directly and frequently, but instead used less reliable pCO_2 calculations from infrequent pH measurements. Despite this, extreme effects of CO_2 were regularly reported by Munday and colleagues.

Munday et al. reply

Contrary to Clark et al.'s assertions above, Munday et al.'s previous experiments maintained stable pH and pCO_2 due to a high level of expertise perfected over many years of OA research. The reason our experiments achieve such tight control of pH and pCO_2 is that we use a precision needle valve to control delivery of CO_2 gas to the mixing tank. Fine-tuning of the needle valve is crucial as it enables a slow and steady stream of CO_2 into the mixing tank. The CO_2 gas was then fed into the impeller of a small submersible pump in the mixing tank, where it is immediately dissolved and rapidly mixed throughout the tank. This ensures a very even and stable pH and pCO_2 in the treatment water. The use of a large header tank (60 l) further buffers any small CO_2 variation. Clark et al. do not report if they used a precision needle valve or not, and they do not appear to have used a pump to dissolve and mix CO_2 gas, instead relying on the slow dissolution of large gas bubbles from aquarium air stones. The absence of a needle valve, or use of one with coarse control, makes it difficult to control CO_2, causing overdosing whenever the solenoid is triggered, and quickly leads to massive overdosing if a solenoid fails. Inability to maintain a narrow range of pH in the mixing tanks (and thus adequately control pCO_2) is exacerbated by relying on airstones to deliver CO_2 gas into the mixing tank instead of rapidly dissolving the gas throughout the mixing tank with a submersible pump, as Munday et al. have done.

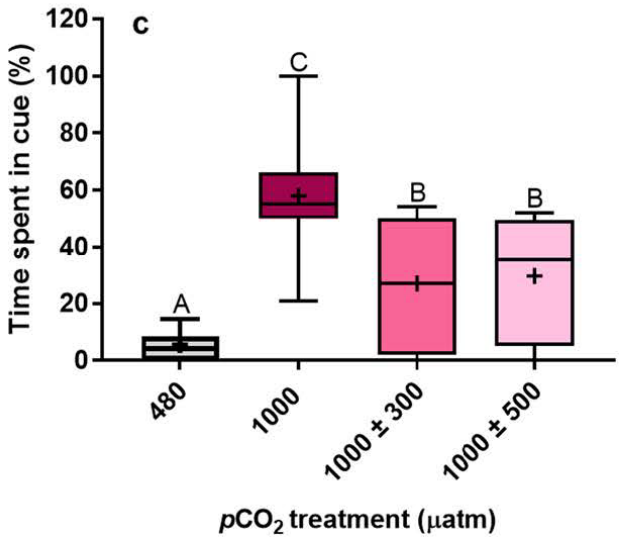
That Clark et al. were unable to maintain sufficient control of their CO_2 treatments is exemplified in their response above, where they report that they captured natural and dosing-related daily fluctuations in pCO_2 in their elevated CO_2 treatment. The elevated treatment level of ~1000 µatm CO_2 is nearly double the maximum pCO_2 of water being delivered to experiments at Lizard Island Research Station for the surrounding lagoon (see Hannan et al. 2020). Therefore, well constrained CO_2 dosing to a set pH should result in a highly stable elevated CO_2 treatment that removes all natural CO_2 variation. That Clark et al.'s elevated CO_2 treatment had a within-day pCO_2 range of 581 µatm ± 508 S.D. is clear evidence of inadequate control of CO_2 dosing in their elevated CO_2 treatment.

Clark et al. response

Munday et al.'s argument also contradicts the majority of results by Jarrold and Munday (2018), who reported that diel cycles of CO_2 did little to reduce the behavioural impairments associated with constant elevated CO_2 exposure in *A. polyacanthus*. Additionally, water on coral reefs and especially around CO_2 seeps in Papua New Guinea is extremely variable in pCO_2 (Fabricius et al., 2011; Shaw et al., 2012), yet damselfishes and cardinalfishes from seeps were reported to show severe behavioural impairments, including strong attraction to predator chemical cues (Munday et al., 2014) (Part B Paper 6). Given these considerations, Munday et al.'s suggestion that behavioural impairments occur only under constant elevated CO_2 – but not under cycling CO_2 conditions that better reflect daily variations on the reef – is puzzling.

Munday et al. reply

Clark are categorically wrong in their statement that Jarrold and Munday (2018) reported that "diel cycles of CO_2 did little to reduce the behavioural impairments associated with constant elevated CO_2 exposure in A. polyacanthus". Both Jarrold et al. (2017) and Jarrold and Munday (2018) clearly show that a diel fluctuation of ± 500 uatm ameliorates the negative effect of a stable 1000 uatm CO_2 treatment on lateralization in A. polyacanthus. Fish exposed to a stable 1000 uatm CO_2 treatment were less lateralized than controls, but the inclusion of ± 500 µatm daily variation restored lateralization to control levels (i.e. there was no effect of elevated CO_2 when ± 500 µatm daily pCO_2 variation was present). The variation in pCO_2 in Clark et al.'s experiments could therefore explain why they did not find significant effects of elevated CO_2 on lateralization in A. polyacanthus and other species. Furthermore, Jarrold et al. (2017) showed that the impaired response to predator cue in clownfish was substantially ameliorated when a daily variation of ± 300 µatm or ± 500 µatm was present, compared with clownfish exposed to a stable 1000 uatm CO_2 treatment. These studies clearly show that daily CO_2 variation can diminish or ameliorate the effects of elevated CO_2 in laboratory experiments.



Percent time that juvenile clownfish (*Amphiprion percula*) exposed to 4 different CO_2 treatments spent in the predator-cue stream in a two-channel flume (n = 16 per treatment). Boxplots are the 25th and 75th quartiles, the line identifies the median, the cross is the mean and the whiskers span the minimum and maximum values.

From Jarrold et al. (2017). Diel CO_2 cycles reduce severity of behavioural abnormalities in coral reef fish under ocean acidification. Scientific Reports, 7: 10153. DOI:10.1038/s41598-017-10378-y

Clark et al. are correct that damselfishes and cardinalfishes at a natural CO_2 seep exhibited impaired behavioural responses (Munday et al. 2014), despite variable pCO_2 at this site and other natural CO_2 seeps. However, the methodology used in that study is different to laboratory-based studies, because small juvenile fish permanently exposed to elevated CO_2 were collected from the CO_2 seep and tested in a two channel y-maze onboard a vessel within 12 hours of capture. By contrast, previous studies by Munday and colleagues, and those of Clark et al. (2020), all involved collecting fish from current-day ambient conditions and exposing them to elevated CO_2 for many days in the laboratory. It is possible the elevated stress from capture and rapid testing exacerbated the behavioural effects of elevated CO_2 at the CO_2 seep. Nevertheless, Nagelkerken et al. (2016) observed a marked reduction in startle distance measured directly underwater in small fishes at CO_2 seeps, which shows their behaviour is affected by elevated CO_2 in the absence of capture stress. More research is needed to understand the behavioural responses of fish at natural CO_2 seeps.

Clark et al. response

Munday et al. argue that our 2016 experiments at the Lizard Island Research Station may have been confounded by high temperatures. We respond to this criticism in point 7, where the topic was again raised in more detail by Munday et al.

Munday et al. reply

See point 7.

5. "*Welch et al.[12] used an alarm cue from a conspecific fish, rather than a predator cue, as it is a more reliable indicator of an immediate predation threat[13]. Clark et al.[1] did not use alarm cues in any of the experiments and their study is therefore not directly comparable to the study by Welch et al.[12], or other studies using alarm cue.*"

Clark et al. response

We are unclear as to why Munday et al. suggest that conspecific alarm cues are a "more reliable indicator of predation threat". Many papers by Munday and colleagues that have reported extreme effects of CO_2 on fish chemical cue detection have used predator cues like in our study (Fig. 1a-f of our main document) (Dixson et al., 2010; Munday et al., 2010; Nilsson et al., 2012; Munday et al., 2014; Munday et al., 2016). Thus, predator cues would appear to be very reliable for testing the effects of OA on the capacity of fish to chemically detect predation threat.

In any event, Munday and colleagues report large CO_2 treatment effects when examining a striking diversity of chemical cues (predators, conspecifics, food, anemones, leaves, grass, etc.) (e.g., Munday et al., 2009; Dixson et al., 2010; Munday et al., 2010; Nilsson et al., 2012; Munday et al., 2013; Munday et al., 2014; Welch et al., 2014; Dixson et al., 2015), so there is little reason to expect that responses to predator cues should be vastly different from responses to conspecific alarm cues.

Additionally, Chivers et al. (2013) reported that conspecific chemical alarm cue degrades very rapidly once released into buckets of seawater stored outside (its efficacy drops to ~64% after 10 min, ~33% after 20 min, and ~5% after 30 min; Fig. S1), and Chivers et al. (2014b) reported that alarm cue degradation occurs even faster (within 15 min) in seawater with elevated CO_2 (~905 µatm) under lab conditions (Fig. S1). Yet, inexplicably, choice flume studies by Munday and colleagues at different CO_2 levels have extracted alarm cue and then used it in batches of trials lasting ~100 min without reporting any changes in fish preference/avoidance behaviour through time. For example, Welch et al. (2014) dosed a 10 L header tank with alarm cue, which allowed ~9 x 11-min trials given that the header tank was draining at 100 ml min^{-1}. Thus, the alarm cue, at least in elevated CO_2 trials, should have been completely degraded for 7 out of the 9 trials, and consequently the fish should have spent equal time on each side of the choice flume. There is no indication of this pattern in the results of the paper and, instead, major behavioural effects of CO_2 are reported.

Munday et al. reply

Conspecific alarm cue is a reliable indicator of immediate predation threat because it is an honest signal that a conspecific has been injured and, thus, that a predator of that particular species is hunting (Smith 1992, Ferrari et al. 2010). Reef fishes may encounter hundreds of different species of potential predators and they need to be able to classify them as dangerous or safe (Mitchell et al. 2015). We know that prey fish have sophisticated recognition capabilities; they can detect predator identity, predator diet, predator size, as well as the proximity and density of predators through odours alone (Mathis and Smith 1993, Kusch et al. 2004, Ferrari et al. 2010). However, much of this information is learned, hence knowing the predation history of the prey being tested is critical. Put simply, predator cues are more generic than conspecific alarm cues, and responses to them are much more variable. The predator may not be a threat to that particular species, or may not be hunting when the cue is detected by the prey. The strength and composition of predator cue also depends on the diet of the predator (Mathis and Smith 1993, Chivers and Mirza 2001), meaning that laboratory experiments will be less effective if the predator is not fed a relevant diet or is starved before use, as was the case in Clark et al.'s experiments. Previous studies by Welch, Ferrari and colleagues have routinely used alarm cue instead of predator cue because it elicits a highly specific and repeatable response. If replication was the goal of their study, it is unclear why Clark et al. did not use alarm cue in any of their experiments, especially considering that this cue is easily obtained.

Previous OA experiments using alarm cues have been conducted in the laboratory (i.e. indoors) where alarm cue is stable for several hours (Chivers et al. 2013). In their response, above, Clark et al. mislead readers by describing results from a paper showing that alarm cues quickly degrade, but fail to mention that this occurred when the cue was placed outside in the sun (i.e. in UV radiation). This is irrelevant, because past studies by Welch, Ferrari and others were done inside where alarm cue lasts much longer. Welch et al. (2014) was conducted in a temperature controlled laboratory at JCU and studies by Ferrari and colleagues were done inside a temperature controlled laboratory at Lizard Island. High CO_2 causes alarm cue to degrade more quickly, but it is still highly potent for 15-20 minutes, which is within the time duration of experimental trials in previous studies. In their responses, Clark et al. have made the erroneous assumption that Welch et al. (2014) conducted 9 x 11 minute trials in a row with a single batch of alarm cue in a 10 l header-tank. This is not correct - alarm cue was replenished after every second trial in Welch et al. (2014). Pilot studies demonstrated the cue remained potent for this duration for this species and experimental setup. Therefore, Clark et al.'s conclusion that alarm cue would have been degraded in 7 of 9 trials in elevated CO_2 is incorrect. We acknowledge that replenishment after every second trial was not stated in the original paper, but it could have easily been ascertained by contacting the corresponding author. Moreover, assuming that Welch et al. (2014) ran 9 x 11 minute trials in a row, at a flow rate of 100 ml min^{-1}, is a nonsensical assumption by Clark et al., because it would drain all the water for the 10 l header tank, making it impossible to retain the necessary gravity-fed flow rate.

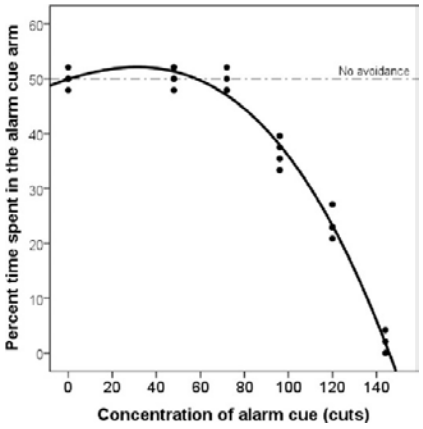


Figure 2. Mean (±SE) proportion of time damselfish spent in the alarm cue arm of the flume when exposed to various concentrations of alarm cues in experiment 1 (n = 5/treatment).

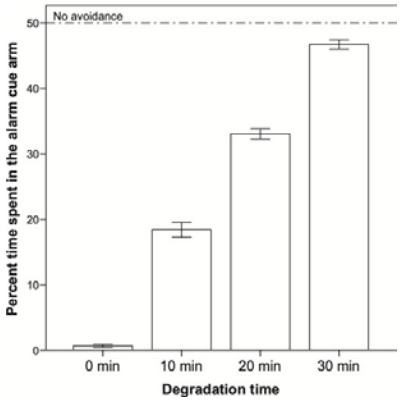


Figure 3. Mean (±SE) proportion of time damselfish spent in the alarm cue arm of the flume in experiment 2. Experiments were undertaken when the alarm cues had aged in ocean water for 0, 10, 20, or 30 min (n = 70–72 /degradation time).

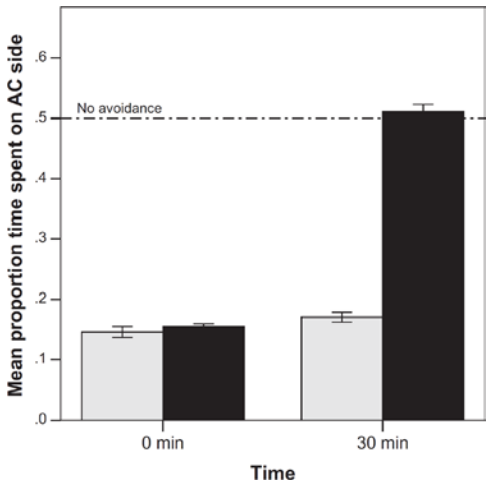


Fig. 1. Mean (±SE) proportion of time spent in the alarm cue arm of the choice flume for fish exposed to ambient (grey bars) or elevated CO_2 (black bars). Experiments were completed late in the day when UVR is at a minimum.

Fig. S1 Figures and captions reproduced from Chivers et al. (2013) (first and second figures) and Chivers et al. (2014b) (third figure). These papers present results from choice flume experiments showing large effect sizes and low within-treatment variation (note that the data points in the first figure are essentially identical on the Y-axis [percent time] across several concentrations). Notably, the authors claim in the first figure that applying 72 cuts to the skin of larval donor fish (6 donor fish x 12 cuts per fish) is insufficient to yield a viable concentration of alarm cue; avoidance responses in conspecifics are only observed once 96 or more cuts are made to the larval donor fish (6 donor fish x 16 or more cuts per fish). Many papers by Munday and colleagues (including authors of Chivers et al. (2013)) have used far fewer cuts to the donor fish but have still reported very clear results (e.g., Ferrari et al., 2011a; Chivers et al., 2014a). Moreover, Chivers et al. (2013) claim that conspecific alarm cue degrades after 30 min in seawater when stored outside (second figure; Y-axis value of 0 represents 100% cue efficacy, Y-axis value of 50 represents 0% cue efficacy), and Chivers et al. (2014b) report that alarm cue degradation occurs even faster (within 15 min) in seawater with elevated CO_2 (~905 µatm) under lab conditions (third figure). Despite this, papers by Munday and colleagues using different CO_2 levels have reported major effects of alarm cue on fish behaviour throughout experiments that took much longer than 15 min following the collection of the cue (see point 5).

Munday et al. reply

Clark et al. have suggested that the findings of Chivers et al (2013) indicate that the concentration of alarm cues used in previous experiments should have been insufficient to induce avoidance response. This is a naïve assumption, or one that intentionally ignores the specific methodologies of the various experiments. Prey fish show finely tuned responses to alarm cues. They respond more strongly to alarm cues from conspecifics than alarm cue from other closely related species. They also respond more strongly to alarm cues produced from individual donors that are of similar size to themselves (Mirza and Chivers 2002, Lonnstedt and McCormick 2011a) and more strongly to donors that are in better body condition (Lonnstedt and McCormick 2011b). Some species show more intense responses than others (Smith 1992, Ferrari et al. 2010). This means that it is not surprising that the concentration of cues that elicits a response in prey is variable across experiments both within and between species. However, even more important to explaining concentration effects, are the specific methodologies used in the previous experiment. The previous studies (Ferrari et al. 2011; Chivers et al. 2014a) cited by Clark et al. have not used a flume. Rather than having alarm cues evenly mixed in a flume, they injected concentrated alarm cues into a tank using a syringe. The cues then disperse through the tank with the aid of an airstone. Depending on where the fish is located in relation to the injection hose, the concentration of alarm cues the fish first detects is variable, but by definition is higher (often cases much higher) than when the cues are evenly dispersed. Their concentration arguments just do not fly, or in this case swim.

6. "*The juvenile Acanthochromis polyacanthus used by Clark et al.[1] were from a highly inbred population of a public aquarium (ReefHQ) that is routinely exposed to very high CO_2 (pCO_2 > 1,100µatm with pH_{NBS} 7.9, total alkalinity 3.1 milliequivalents l^{-1}, 26°C) and this population may therefore have adapted to ocean acidification conditions.*"

Clark et al. response

Only the juvenile *A. polyacanthus* used in some experiments in 2015 were obtained from a public aquarium (ReefHQ); in the other two years of our study, this species was wild-caught at the Lizard Island Research Station. There is no evidence in our dataset that the *A. polyacanthus* from 2015 were adapted to OA conditions, as Munday et al. suggest. Additionally, Sundin et al. (2019) found no difference in behaviour between wild-caught *A. polyacanthus* and those obtained from ReefHQ when testing for effects of OA. It is notable that previous papers by Munday and colleagues do not support the argument they make in their Matters Arising letter, as they have reported that behavioural impairments under OA remain in fish even after transgenerational acclimation (Welch et al. (2014) specifically tested *A. polyacanthus*), and when living in acidified conditions

around CO_2 seeps (with pCO_2 of 998 µatm and pH 7.72) (Munday et al., 2014). If the founder population of *A. polyacanthus* at ReefHQ had limited genetic diversity but their progeny in Clark et al. (2020) were still unimpaired by CO_2, this suggests that behavioural resilience to OA must be common in the genome of the species.

Munday et al. reply

Clark et al.'s conclusions about the potential for adaptation to high CO_2 in an aquarium population of A. polyacanthus are puzzling. A. polyacanthus from ReefHQ have been bred for many generations in an environment that regularly experiences high CO_2 (>1100 uatm CO_2 as shown in our Matter Arising). This provides ample opportunity for selection of genotypes with enhanced CO_2 tolerance in this population. Moreover, we already know that there is genetic variance in behavioural resilience to OA in this species (Welch and Munday 2017). This signature of genetic variance in behavioural resilience to elevated CO_2 is even detectable in the brain of juvenile A. polyacanthus. Juveniles of behaviourally tolerant parents exhibit a different brain transcriptome in high CO_2 compared with juveniles of behaviourally sensitive parents when both lineages are reared in a common environment (Shunter et al. 2016). So, yes, there is genetic variance associated with behavioural resilience to OA in A. polyacanthus, and ReefHQ provides an ideal environment for selection of this trait over a number of generations in a closed population (i.e. no gene swamping from other populations). Finally, there is increasing evidence that transgenerational plasticity may depend on the early developmental exposure of parents, or even grandparents, to altered environmental conditions (Donelson et al. 2018). Welch et al. (2014) only exposed breeding adults to elevated CO_2, therefore the absence of transgenerational plasticity in that population does not indicate that it will not occur when parents (and grandparents) develop their entire lives in high CO_2, as occurs at ReefHQ. Consequently, the Welch et al. (2014) study does not provide evidence against either transgenerational plasticity or adaption to high CO_2 in the ReefHQ population of A. polyacanthus as proposed here by Clark et al.

Contrary to Clark et al.'s assumptions above, CO_2 seeps are not conducive to adaption to high CO_2 in fish. The very small spatial scale of natural CO_2 seeps is mismatched with the highly dispersive larval stage of reef fishes, meaning that offspring from CO_2 seep populations disperse to other non-seep areas rather than being retained in the seep population. At the same time, seep populations are almost exclusively replenished by larvae from non-seep reefs, thus swamping any potential for local scale adaptation to high CO_2 (Munday et al. 2013b).

7. "*Lizard Island was affected by a heatwave in 2016 and the water temperature of the experimental treatments was 1-2°C above average during a large part of the experiment in 2016 described by Clark et al.*[1] *(Supplementary Information and Supplementary Table 2). Increased temperature is known to diminish or reverse the effects of elevated CO_2 on risk-assessment behaviour*[14]*, and A. polyacanthus is one of the most thermally sensitive coral reef fish species, which could explain the absence of significant effects for this species in 2016, whereas they observed significant effects in 2014.*"

Clark et al. response

Our experiments at Lizard Island in 2016 were conducted in mid-late January (austral summer). Specifically, wild fish were collected from the reef during 12-13 January, with experiments taking place during 18-31 January. Reef water temperatures during the experimental period ranged from a daily average of 28.5°C to 30.2°C (data obtained from the Australian Institute of Marine Science Data Centre). Fig. S2 below shows the daily reef temperature record during our 2016 experiments as well as 5 years prior, illustrating that the temperatures during our 2016 experimental period (28.5-30.2°C) are experienced by coral reef organisms at Lizard Island nearly every year. As Munday and colleagues should know from their many previous studies at the Lizard Island Research Station, the water temperature can increase slightly as water is pumped from the reef to the aquarium facilities where all tank-based experiments are conducted. Nevertheless, our in-tank experimental

temperatures averaged 29.5°C during our 2016 experiments, and thus our experiments were not conducted during heatwave conditions.

The heatwave conditions in 2016 commenced after our experiments, with sensitive *Acropora* corals starting to bleach in February and peak bleaching occurring in April (Wismer et al., 2019). Interestingly, Wismer et al. (2019) found that fish communities were quite resilient to the heatwave, despite the dramatic loss of live coral cover. Given that Munday et al. (2010) reported extreme behavioural impairments when conducting experiments at 30 ± 0.5°C (mean ± SD), and a recent study by Jarrold and Munday (2018) concluded that elevated temperature does not modify the behavioural effects of CO_2, Munday et al.'s claim of diminished effects of CO_2 at high temperatures does not align with their previous reports.

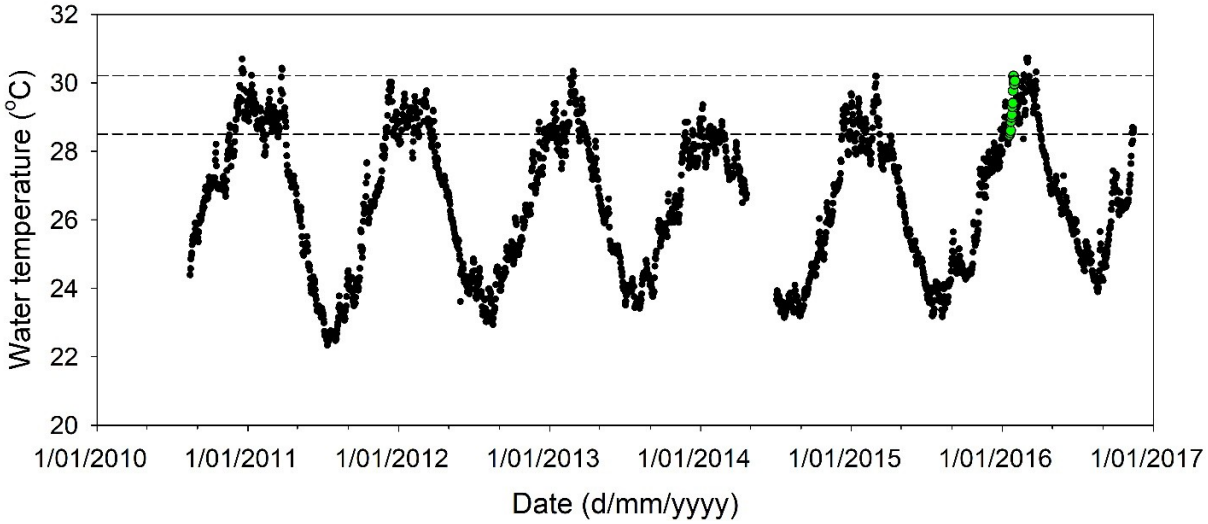


Fig. S2 Mean daily water temperature (0.6 m depth) at Lizard Island from 14 August 2010 to 14 November 2016 (sourced from the Australian Institute of Marine Science Data Centre). Experiments in Clark et al. (2020) were conducted in August-September 2014 and mid-late January 2016. Temperatures throughout the experimental period in 2016 are highlighted (green circles), and horizontal dashed lines illustrate the minimum (28.5°C) and maximum (30.2°C) temperatures experienced during that period to aid comparison with other years.

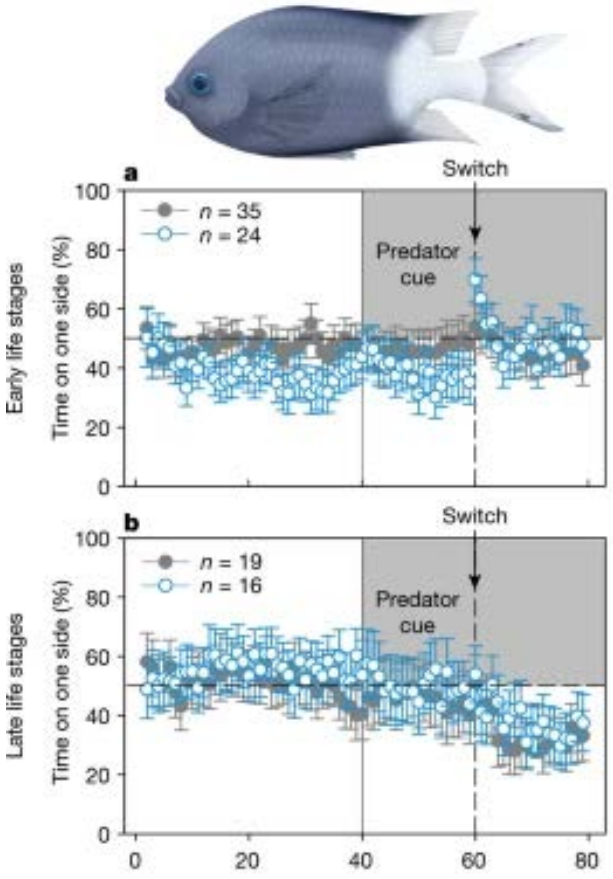
Munday et al. reply

Clark et al. have missed the point of our criticism here. Our point is that the temperatures in their experimental treatments were higher than the average summer water temperature at Lizard Island. Average water temperature at 0.6 m recorded by automated buoys at Lizard Island was 28.7°C for 15 January to 1 February in the 5 years 2011-2015. By contrast, water temperature in Clark et al.'s experiments exceed 29.7°C in 41% of their reported observations. This is detailed in Supplementary Table 2 in our Matters Arising. Clark et al. are correct that the temperature of bulk water for experiments at Lizard Island can be substantially higher than the surrounding ocean temperature, because the water is pumped into external header tanks exposed to the sun and ambient air temperature before being used in experiments. Consequently, water temperature being delivered to aquariums can rise to problematic levels during warm summer days at Lizard Island Research Station. We are well aware of this problem, which is why previous experiment by Munday and colleagues at Lizard Island have always been conducted in the temperature controlled rooms at this facility. In previous experiments we have used the room air conditioners to buffer the excessively high temperature of seawater being delivered to experiments on many summer days. Clark et al. (2020) do not provide temperature records for all days of the experiments in their supplementary material, but the data they do provide suggest they were not able to counter this artificial warming in their experimental treatments at Lizard Island.

Inspection of Clark et al.'s 2016 data in their Figure 2 suggests that neither size class of A. polyacanthus had a meaningful response to the introduction of predator cue into the flume in the 2016 experiment – the lines are essentially flat for the period before and after predator cue is introduced (see Figure below). Most importantly, control fish did not exhibit a negative reaction to the introduction of predator after 40 minutes acclimation, or over the next 20 minutes of the trial. This indicates that A. polyacanthus was not responding to predator cue in any meaningful way in the 2016 experiment, possibly because of the higher temperature in experiments that year. Clark et al. (2020) dismiss their significant effect of elevated CO_2 on response to predator odour in A. polyacanthus in 2014 because they did not find an effect in 2016. We suggest that it's the 2016 result that should be dismissed instead, because control fish exhibit no meaningful response to predator odour in the flume in 2016. It is simply not possible to test for an effect of high CO_2 on the response to predator cue in this case, because the fish are not responding to it.

In their response, above, Clark et al. make an incorrect and misleading statement about the water temperature in past experiments by Munday and colleagues at Lizard Island. They say "Munday et al. (2010) reported extreme behavioural impairments when conducting experiments at 30 ± 0.5°C (mean ± SD)" when discussing water temperature in experiments at Lizard Island. However, this temperature was for warm-acclimated clownfishes, the species not tested by Clark et al., reared at James Cook University in Townsville. The temperature in experiments done at Lizard Island was 27.6 ± 1.5°C (mean ± SD), which is clearly stated in Munday et al. (2010).

Clark et al. are correct that Jarrold and Munday (2018) did not find an effect of higher temperature (31°C) on the altered lateralization response of juvenile A. polyacanthus in elevated CO_2 treatments. However, there is good evidence from other studies that elevated temperature can ameliorate, or reverse, the behavioural effects of elevated CO_2 observed at ambient temperatures in reef fishes (Domenici et al. 2014, Ferrari et al. 2015).



Time spend in a water stream with predator cue in a two-channel flume for two life stage of *Acanthochromis polyacanthus* tested by Clark et al. (2020) in 2016 at Lizard Island. Fish were acclimated to the flume for the first 40 minutes, after which time predator cue was introduced to one side of the flume. Grey dots are control fish and blue dots are high CO_2 fish. It is apparent from this graph that control fish (grey dots) did not respond negatively to the introduction of predator cue at 40 min, or for the 20 min thereafter. The absence of a meaningful response to predator cue in control fish shows that either the concentration of the predator cue was insufficient, or some other factor(s) impacted the experiment.

Figure 2 from Clark et al. (2020) Ocean acidification does not impair the behaviour of coral reef fishes. Nature, https://doi.org/10.1038/s41586-019-1903-y

8. "*Clark et al.[1] did not include at least ten other studies that used alternative methods that support previous Y-maze studies, most of which were carried out by observers who were blinded to treatment (Fig. 1). Notably, the altered behaviour and increased mortality rate of field-transplanted fish, which was recorded by blinded observers, strongly support the results of Y-maze and other laboratory experiments (Fig. 1).*"

Clark et al. response

We believe that Munday et al. are referring to the 14 studies listed in their fig. 1 under 'Supporting studies using other methods' and 'Supporting field-based experiments', of which Munday is an author on all but one. In fact, of the 44 studies on coral reef fishes listed in Munday et al.'s Supplementary Table 1, 41 document an effect of OA on some aspect of fish behaviour, and Munday is an author on 37 of those 41. If we examine the studies listed in Munday et al.'s fig. 1 and Supplementary Table 1 that were not conducted in coral reef systems (i.e., 60% (66/110) of the studies listed), none document the extreme behavioural impairments that Munday and colleagues have reported in coral reef fishes. As such, our *a priori* assumption when we began our experiments was that coral reef fishes were the most sensitive fish group to OA, which is why we chose to investigate them. It is likely that the dramatic results reported in multiple papers by Munday and colleagues on coral reef fishes led to some degree of confirmation and publication biases around the world (see our main document). This phenomenon, although often the result of actions that are subconscious and not ill-intentioned, is unfortunately common in science (e.g., Young et al., 2008; Ioannidis et al., 2015; Nissen et al., 2016; Parker et al., 2016; Fanelli et al., 2017).

We encourage researchers conducting experiments and analyses in a blinded manner to provide details of how such blinding was achieved, as well as videos and other supporting evidence to bolster these important claims. Properly blinding experiments can be difficult, particularly in situations with small teams where everyone must assist with multiple aspects of the project. For example, in the preliminary stages of our research, we discovered that there were many subtle hints (like solenoid valves firing to inject CO_2 into header tanks) that would reveal the treatment group of the test fish and prevent proper blinding. Consequently, in Clark et al. (2020) we opted to video our experiments and include a descriptive note at the commencement of each trial to remove biases (also see Clark, 2017).

Munday et al. reply

The evidence that elevated CO_2 can affect the behaviour of coral reef fishes, other fishes, and invertebrates is overwhelming. As noted in our Matters Arising, at least 85 research papers from 56 different lead authors, with over 180 coauthors, and involving more than 90 different institutions, report significant effects of elevated CO_2 on the behaviour of fish from coral reefs and other habitats, including at least 8 papers by the authors of Clark et al. (Supplementary Table 1). Clark et al. cannot seriously be arguing that the majority of these studies are wrong, or somehow are all the subject to confirmation bias. We have greater faith in the ability of other researchers to conduct robust and unbiased studies than Clark et al. seem to have. Effects of elevated CO_2 on fish behaviour has been reported in species as diverse as salmon, sharks, seabass, barramundi and eels, so this is not a phenomenon unique to coral reef fishes. There are also dozens of studies showing significant effects of elevated CO_2 on the behaviour of invertebrates (Clements and Hunt 2016, Thomas et al. 2020).

Focusing on the studies with coral reef fishes, there are more than 40 peer-reviewed papers, authored by more than 20 different lead authors and over 45 different co-authors, from over 30 different international institutions, showing significant effects of elevated CO_2 on fish behaviour. Munday is one of many authors in these studies, and in the majority of cases provided resources and expertise in OA research rather than being directly involved in the collection of behavioural data. To argue that studies by different lead authors, working independently from each other, should not be considered as independent research because they have one author in common is an extraordinary strict definition of independent research. Taking the same approach to OA papers from Clark and colleagues would mean that they all lack independence because Fredrick Jutfelt is an author on all but one of them.

Given Clark et al.'s concern about researcher independence, it is worth noting that Paula et al. (2019) report significant effects of elevated CO_2 on the behaviour of the coral reef cleaner fish (Labrioides dimidiatus) that closely match the patterns reported by Munday and colleagues in previous studies. Paula and colleagues show that the cognitive ability of the cleaner wrasse was significantly reduced at both 750 and 980 µatm CO_2. Yet, there was more variation in the response at 750 µatm compared with 980 µatm CO_2, with some individuals still able to solve the cognitive task at that CO_2 level. By contrast, no individuals were considered to have solved the task at 980 µatm CO_2. This study independently replicates the dose dependent effects of CO_2 on reef fish behaviour reported previously by Munday and colleagues (e.g. Munday et al. 2010, 2012, 2013, Ferrari et al. 2011) and generally supports the findings of cognitive impairment in reef fishes at elevated CO_2 in previous studies.

With regards to the experimental treatment blinding done in previous experiments by Munday and colleagues, this was achieved by one group of researchers doing the experimental CO_2 manipulation and fish husbandry in one laboratory room and another group of researchers doing the behavioural assays in another laboratory room, or in the field. The researchers doing the behavioural assays were given fish with a numerical code or tag that did not contain clues as to the treatment. This separation of space and people ensured that blinding was not compromised.

Flume methodology

9. "*Previous studies concentrated the predator odour by turning off the water flow for several hours; by contrast Clark et al.[1] used an open flow and did not concentrate predator odour. The water turnover in the predator tank was 10 min (60 l, 6 l min^{-1}) at LIRS in 2014 and 2016, even when using a flume flow of 135 ml min^{-1}. This causes the predator cue to be greatly diluted compared with previous studies. The weak response of control fish to predator odour in Clark et al.[1] (mean 40% and mode 50% of time in predator cue: see extended data figure 4 of Clark et al.[1]) is indicative of weak cue strength. The strength of avoidance to predator odour is proportional to its concentration[15], so using weak cues elicits weaker and more variable responses.*"

Clark et al. response

There are three main issues with the methodology used in the previous papers by Munday and colleagues, which compromise the reproducibility and reliability of their findings. We list them below and expand on them in the subsequent paragraphs.

i. Insufficient information is provided in many previous papers for independent parties to reproduce predator cue concentrations similar to the ones used by Munday and colleagues.

ii. A concentrated predator cue (or alarm cue) obtained as described by Munday et al. is not ecologically relevant and may lose efficacy throughout trials. Moreover, using a depleting water source in the flume header tank means that head pressure and mixing characteristics in the flume constantly change throughout trials.

iii. Isolating a predator in a bucket has the confounding effect of elevating molecules such as CO_2, ammonia and cortisol in the water, which may alter the response of the focal fish in the flume. This approach also means that the temperature in the bucket is likely to increase or decrease depending on the ambient air temperature, which can cause issues of mixing in the flume (Gouraguine et al., 2019).

The salient point here is that there was clear evidence of predator cue avoidance by fishes across years in Clark et al. (2020), showing that our predator cue concentration and flume design were effective at generating the expected control behavioural response (see figs. 1 and 2 in Clark et al. 2020). Thus, the previously reported effects of CO_2 on predator cue avoidance should have been detectable using our experimental setup.

Munday et al. point out that, in their previous landmark studies showing impairment of predator cue avoidance, predator cues were created by isolating a predator in a small bucket containing an airstone. This

water was then placed into one of the header tanks supplying the choice flume. Unfortunately, many methodological details regarding the procedures used to generate a particular chemical cue concentration are missing from these papers (e.g., volume of water in the predator isolation bucket, how many different predators were used individually across all flume trials, predator size, predator confinement time, what and when the predators were fed, etc.). This lack of methodological information obviously hampers replication efforts. We thought carefully about how to best generate chemical cues for our experiments given this lack of detail and the other issues caused by the approach described by Munday et al. As we mention above, confining a predator to a bucket with no water flow has the confounding effect of concentrating excreted products like CO_2, ammonia and cortisol, while also allowing water temperature to drift. It has also been reported that chemical alarm cues degrade very rapidly once released into seawater (Chivers et al., 2013; Chivers et al., 2014b) (Fig. S1), and nothing is known about the degradation dynamics of predator chemical cues. Thus, the method advocated by Munday et al. may lead to a rapid decline in cue efficacy as sequential trials continue to drain the header tank after the initial addition of the water containing the chemical cue (see point 5). All of these factors can impact the behavioural response of focal fish and inferences that can be drawn from the resulting flume experiments.

Rather than isolating a single predator in a bucket, we used several predators (higher biomass) in one of our flow-through header tanks to ensure that the predator cue concentration remained elevated and detectable. Flow rates and predator biomass varied across years depending on water supply needs (we used a range of flume sizes to accommodate varying fish sizes; see Clark et al. 2020). Importantly, we opted for the chosen methodology because we wanted to ensure the ecological relevance of the predator cue to which fish were exposed. In coral reef environments, there is typically a vast surrounding body of water and regular flow from tides and waves, meaning that predator cues are likely to be very dilute and not concomitant with elevated excretory products (as well as being combined with a cocktail of other chemical, auditory, and visual cues). We believe that the bucket method described by Munday and colleagues is less ecologically relevant to prey fishes for these reasons, as well as potentially being confounded by a decline in cue efficacy throughout trials.

Another advantage of our approach was that it allowed us to keep a constant head pressure in our 'predator cue' header tank, rather than having the tank water gradually deplete over the course of several trials as is the case with the method described by Munday et al. By providing a constant supply of predator cue in this way, we were able to conduct longer trials (e.g., 18-120 min, cf. 9-11 min in most previous studies), allowing sufficient time for fish to habituate to the experimental apparatus and sample the water chemical cues at will. We have found in our work that maintaining laminar flow in two-current choice flumes depends on having consistent, even flow of water of the same temperature in both channels throughout the trials (Jutfelt et al., 2017; Gouraguine et al., 2019). Gradually depleting water levels and head pressure in the header tanks (i.e., higher head pressure at the start vs. end of the experiment) makes maintaining constant laminar, unmixing flow a serious challenge with the flow meters that are used. Adopting the approach advocated by Munday et al. of making and using 'batches' of concentrated predator cue/metabolites would also have made it more difficult to maintain matching high- or ambient-CO_2 levels in both header tanks at different acclimation treatments. We encourage researchers to consult published guidelines that explore some of these methodological issues involved in using two-current choice flumes (Jutfelt et al., 2017).

Munday et al. reply.

Clark and colleagues responses, above, serve to exemplify that the methods used in their experiments were vastly different from those used in previous studies by Munday and colleagues, and thus they have not closely repeated the methods of previous studies as they claim to have done in Clark et al. 2020. Clark et al. used a much weaker concentration of predator cue in their flume experiments compared with earlier studies and this can explain why they got much weaker and more variable results than earlier studies. If Clark et al.'s goal was to repeat past experiments as closely as possible they could have contacted the corresponding author(s) of past studies at any time to get clarification about any methods they were unsure about, as Nosek and Errington (2020) recommend.

It is important to mention here that the Y-maze flume used in previous studies by our group has been successfully used by other research groups to test olfactory preferences of reef fishes and other species over a period of nearly 20 years (e.g. Atema et al. 2002, Gerlach et al. 2007, Miller-Sims et al. 2011, Gould et al. 2015, Ou et al. 2015, Coppock et al. 2016). This indicates that the issue operating this style of flume that are raised by Clark and colleagues are not shared by other research groups – as we do, these other groups are able to achieve laminar flow and maintain adequate flow rates in the flume.

Here, we deal briefly with each of the three specific issues that Clark et al. raise above.

(i) We acknowledge that more details of the fluming methods would have been useful in the earliest papers we published to facilitate replication; however, reaching out to the original authors for clarification would have been both the easiest and most productive way to resolve this issue. Word length restrictions meant we did not provide expansive details of the methods. However, we contend that if Clark et al.'s motivation was to repeat past experiments as closely as possible, as they repeatedly state in Clark et al. (2020), then they could have contacted the corresponding author(s) of past studies at any time to ask for clarification about the methods, which they did not do.

(ii) Clark et al. are correct that the predator odor used in papers by Munday et al. may be at a higher concentration than what is often ecologically relevant – we never tried to replicate natural occurring concentrations of predator cue and never claimed to have done so. Concentration of predator cue will vary dramatically within a coral reef system, and in relation to individual prey fish, dependent on a vast number of factors, such as, currents, tides, predator type, predator location on the reef, predator size, number and activity, and the proximity of individual prey fish. In our initial experiments to determine if ocean acidification conditions could affect fish behaviour we purposefully ignored natural concentrations of predator odour, because they are highly variable, context dependent, and unknown in most cases. Instead, we used concentrated predator odour to test if elevated CO_2 altered the ability of prey fish to detect chemical cues, or not. The aim of these flume studies was not to test how prey might react to natural concentrations of predator cue (which we could not know), only to test if elevated CO_2 affected the response to chemical cues. Having established that the response to concentrated predator cues was affected by elevated CO_2 (Dixson et al. 2010), we then transplanted small juveniles to natural reef habitat where they were exposed to precisely the predator cue strength they would experience in nature. We did not pretend to know what the concentration of predator cue might be on the reef, which Clark et al. cannot possibly know either, instead we transplanted juvenile fish to small reefs where they would experience the correct natural mix of predator and other chemical cues (Munday et al. 2010, 2012, Ferrari et al. 2011, McCormick et al. 2013, Chivers et al. 2014b). Juvenile fish that had been exposed to elevated CO_2 were more active, ventured farther away from shelter and were less responsive to threats than controls, which could be related to their impaired ability to detect chemical cues. As a result they experienced a much higher mortality than juvenile fish that had not been exposed to elevated CO_2. This is a more realistic ecological scenario than anything that Clark and colleagues have done in their experiments.

Importantly, observations of behaviour and survival in field-transplanted fish, done by fully blinded observers, matched the results obtained from the flume, confirming the reliability of the flume results. Individual fish that were found to be unaffected by elevated CO_2 when tested in the flume did not exhibit altered behaviour when transplanted to the field and did not suffer higher mortality (Munday et al. 2012). The reverse was true for fish found to be affected by elevated CO_2 when tested in the flume – they did exhibit altered behaviour in the field and suffered higher mortality (Munday et al, 2012). This is a clear validation that the results from the flume were meaningful and that the methodology could detect individual variation in tolerance to OA conditions.

Also mentioned previously, the header tanks were never left to empty completely during experimental trials. The header tanks are gravity driven and therefore once the water pressure falls below the point where the flow meters are restricting the flow (100 ml/min), the dynamics in the flume will change and potentially cause issues with laminar flow. Flow rates were constantly

monitored and dye tests conducted often to ensure laminar flow existed.

(iii) As we explain above, previous flume studies did not attempt to replicate natural concentrations of predator odour, which could not possibly be known. That was the purpose of field-based studies (Munday et al. 2010, 2012, Ferrari et al. 2011, McCormick et al. 2013), where predator concentrations were natural by the very design of the experiment. Concentrating predator cue may indeed have concentrated a range of metabolic products from predators, but this did not hinder our ability to show that elevated CO_2 affected the response of prey fish to those cues. Furthermore, Dixson et al. (2010) included comparisons of non-predator treated seawater vs. untreated seawater, and non-predator vs. predator treated seawater, made using identical methods as those described for generating the predator cue. Fish displayed contrasting behavioral choices towards these other stimuli, indicating the results of the predator cue trials were not due to some unknown metabolic contamination in the preparation of predator cue water. Finally, we are well aware of the problem that a temperature differential between water streams will have on the function of the flume, which is why predators tanks were kept in a temperature controlled water bath (all experiments at JCU) or in a temperature controlled room (Lizard Island), not in ambient outside air conditions.

10. "*Previous studies have always kept predator and alarm-cue donors in control water, because a low pH can reduce the efficacy of the cue[16]. By contrast, Clark et al.[1] kept the predator in low pH water for 4-18 days*."

Clark et al. response

sThis argument is not supported by previous work by Munday and colleagues. Munday et al. (2013) state "*One water source contained seawater with predator odor and the other contained untreated seawater. Water containing the chemical cue of a predator was created by soaking a single C. cyanostigma in 60 L seawater for 2 h prior to testing. Both water sources were at the same CO_2 level as the rearing treatment of the fish*" (this included a predicted end-of-century pH of 7.85 [939 µatm CO_2]). The predator cue attraction/avoidance and CO_2 effects reported in Munday et al. (2013) are as extreme as in other studies by Munday and colleagues, indicating no loss of cue efficacy (see Fig. 1c of our main document). Another study by Munday and colleagues (Welch et al., 2014) placed conspecific chemical alarm cues directly into treatment water with different pH levels (including a predicted end-of-century pH of 7.85 [912 µatm CO_2]) and found no indication of pH-dependent differences in cue efficacy (cf. Fig. S1).

In our research, we kept predators and test fish in matching acclimation conditions throughout experiments, so that predator avoidance could be tested in ecologically relevant CO_2 concentrations. As discussed above and as is clear in Clark et al. (2020), we detected clear predator cue avoidance responses in fish throughout our experiments, demonstrating that our cue treatment worked (figs. 1 and 2 in Clark et al. 2020). If high CO_2 decreases cue efficacy, as Munday et al. claim, then their previous experiments having maintained predators in present-day CO_2 conditions should be viewed with caution as their relevance to end-of-century OA scenarios is unclear.

Munday et al. reply

In their response, above, Clark and colleagues take a section of text from Munday et al. (2013) out of context of the entire paragraph in which it was embedded. The sentence "Both water sources were at the same CO_2 level as the rearing treatment of the fish" was in the paragraph describing the two-channel flume and meant that the water sources used in the flume were the at same CO_2 level as the rearing treatment of the fish. This was achieved by dosing the predator cue water with CO_2 to the desired level of pH immediately before it was used in the flume. We appreciate that our wording was unclear in the original paper, but this was clarified in the legend of Table 1 in our Matters Arising, where we say "Predator cue water was CO_2-dosed immediately before use when required (Munday et al. 2013)." Clark et al. kept predator in low pH water for 4-18 days,

which could affect predator physiology as well as the efficacy of cues. Again, this only serves to illustrate that Clark et al. (2020) did not closely repeat earlier experiments as they claim.

11. "*Previous studies used a compressed Y-maze design (commonly used in fish ecology at the time) in which the fish could enter either arm of the flume and remain there during the trial (for example, type C in ref. [17]). The fish was gently recentered in the middle of the rear part of the flume after the water switch to ensure it had equal opportunity to choose during the second half of the trial (communicated to F. Jutfelt on 18 July 2017). Clark et al.[1] modified the apparatus to prevent fish entering either arm of the maze (type H).*"

Clark et al. response

The suggestion by Munday et al. is that fish exposed to elevated CO_2 are strongly attracted to predator cues in two-current choice flumes of "type C" configuration, but that this effect disappears in two-current choice flumes of "type H" configuration. Unfortunately, they do not provide a rationale for this assertion, and we cannot think of a reason why a slight difference in the design of the flume would restore the ability of CO_2-exposed fish to detect and correctly respond to chemical cues. On the contrary, our modified design was chosen to overcome two important shortcomings in previous papers by Munday and colleagues (further detailed in Jutfelt et al. 2017):

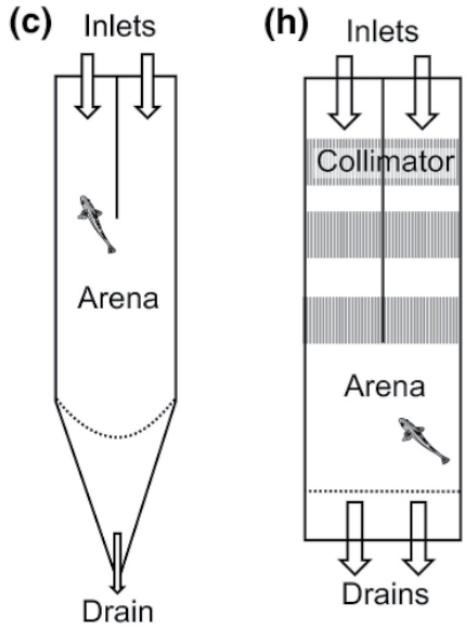
i. Because a fish will typically "enter either arm of the flume and remain there during the trial", type C flumes require manipulating and disturbing the test fish during behavioural trials. This is illustrated in the videos provided by Munday et al., where regular human interference clearly increases stress and modifies behaviour in the fish (doi: 10.4225/28/5add60af3a267; direct link: http://data.qld.edu.au/public/Q5842/2018_WelchMunday_RawDataOlfactoryResponse/). The communication to F. Jutfelt highlighting that "the fish was gently recentered in the middle of the rear part of the flume after the water switch to ensure it had equal opportunity to choose" is an important detail that should be provided in publications, not via a personal communication. Even a "gentle" repositioning likely induces a stress response in the focal fish, which may significantly influence their subsequent side choice. A fish pushed out from one arm of a type C choice flume may be less likely to re-enter that arm during the remainder of the trial.
The type H flume design used in Clark et al. (2020) eliminates the need to physically disturb the test fish during the behavioural trial, including during side-switching of water sources, and allows the fish to constantly sample and re-sample both water cues throughout the trial (see Jutfelt et al., 2017). Our methods involved leaving the fish alone and behind a visual barrier with a camera overhead, so that there were no visual disturbances or observer effects.

ii. Unlike Munday and colleagues, we used automated video tracking to assess the amount of time fish spent on either side of the flume, removing any potential observer biases. Observer bias is the well-described phenomenon in which the conscious or subconscious expectations of observers can influence the outcome of observational experiments (e.g., Young et al., 2008; Simmons et al., 2011; John et al., 2012; Ioannidis et al., 2015; Parker et al., 2016; Fanelli et al., 2017; Ihle et al., 2017). It is colloquially said that unblinded visual observations can be likened to operating a dowsing rod while knowing where the water pipe is buried. To avoid observer bias, it is imperative to use methods that minimise this source of bias, such as blinded observations (but see point 8) or video recording with automated tracking of animal movements (Holman et al., 2015; Kardish et al., 2015; Jutfelt et al., 2017).
Direct visual observation of the fish in the field of OA and fish behaviour has often been conducted with the observer visible to the fish (e.g., Munday et al., 2009; Dixson et al., 2010; Munday et al., 2010; Munday et al., 2014), introducing an additional stressor. Video recording, and ensuring procedures such as side switching of cues do not visually or physically (e.g., bench vibrations) disturb the fish, can eliminate these issues (Jutfelt et al., 2017).

Munday et al. reply

Above, Clark et al. describe the addition of barriers so that fish cannot move into the arms of the flume as a minor modification – it is not. Fish swim into the current in a flume. Without barriers across the entry to the lanes (arms), a fish can swim up into the preferred lane of the flume where it will not be able to sample the other water stream unless it turns and exits the lane. By contrast, in a flume with a barrier preventing entry to either lane, the fish can move laterally into either water stream at any time. It is easy to comprehend that once a preference is made in an unobstructed flume design (type C) that a stronger preference is likely to be registered compared with a flume with a barrier preventing entry into the arms (type H).

As Clark et al. describe above, they make other changes to the operation of the flume, including the time that the water sources are switched, and using continuous data recording rather than sampling at set time points. We are not arguing the pros and cons of the different methods, the point we make is that Clark et al. did not closely repeat the methods of past experiments as they claimed. Instead they changed the methodology in so many ways, which are likely to generate weaker and more variable results, that a direct comparison of results from previous studies is not possible and comparison of data from previous studies with a multispecies data distribution derived from their own data is invalid.



Side by side comparison of the flumes used by (c) Munday and colleauges and (h) Clark et al. (2020).

Figure 1 from Jutfelt et al. (2016) Two-current choice flumes for testing avoidance and preference in aquatic animals. Methods in Ecology and Evolution. Doi.org/10.1111/2041-210X.12668

12. "*Previous studies had 2 x 2 min observation periods that were interspersed with a 3-min water switch and habituation period. Test durations of Clark et al.*[1] *were 5-20 times longer (10-40 min per side). The motivation of fish after 10-40 min of predator cue exposure without further predator reinforcement is likely to be different compared with trials with a short duration that simulate immediate risk. All of these differences could explain the much stronger responses and reduced variance that were observed in previous studies compared with those described by Clark et al.*[1]"

Clark et al. response

Our data unequivocally disprove the argument of Munday et al. – we demonstrated in our 2016 dataset that even 20 min of continuous exposure to predator cue down one side of the flume does not result in fish becoming ambivalent to the predator cue and dismissing it due to lack of further reinforcement (e.g., fig. 2, in in Clark et al. 2020). Even so, most of our trials were closer in length to the short (9-11 min) durations advocated by Munday et al. (i.e., 18 min total, with the cue on each side for 8 min and a 2 min 'switch' period). The longer trials we conducted in 2016 were motivated in part by an interest in assessing how fish behaviour might differ once they have recovered from mounting an acute stress response caused by handling and transfer into the

flume, something that had never been assessed previously in this field. Additional stress in focal animals may have been caused by Munday and colleagues' technique of "recentering" the fish in a type C flume (see point 11). Notably, the same flume methodology advocated by Munday et al. yielded results (Dixson et al., 2014) with coral larvae inexplicably swimming more than twice as fast as they are physically able to achieve (Baird et al., 2014; Hata et al., 2017). Something is clearly awry with the advocated techniques.

Munday et al. reply

Clark et al. are correct that their 2016 data set does not demonstrate fish becoming less responsive to predator cue through time in their experiments. However, this is within the context of their experiments using more diluted predator cues that previous studies, predators permanently exposed to elevated CO_2 and fundamental differences in flume design and operation. A side-by-side comparison with past studies is problematic when the duration of the trial and many other aspects of the approach are so fundamentally different.

There are two other important observations to make from Clark et al.'s 2016 data in their Figure 2. First, as described at Point 7 above, neither size class of A. polyacanthus appear to respond to the introduction of predator cue into the flume in the 2016 experiment. Most importantly, control fish do no exhibit a negative reaction to the introduction of predator at the start, or over the first 20 minutes of the trial. This indicates that in the 2016 experiment this species was not responding to predator cue in any meaningful way. Clark et al. (2020) dismiss their significant effect of elevated CO_2 on response to predator odour in A. polyacanthus in 2014 because they did not find an effect in 2016. We suggest that the 2016 result is the one that should be dismissed because there appears to be no meaningful response to predator odour in the flume. It is pointless to argue that long trials do not result in ambivalence if there is no response to predator cue in the first place, at least for this species. Second, for the other two species that did respond to predator odour in that experiment, Dascyllus aruanus and Dischistodus perspicillatus, a close examination of the time period immediately before and after the switch suggests that elevated CO_2 fish may indeed be responding differently to control fish. By pooling all the data from different species together in a multispecies data set (for comparison with previous studies) Clark et al. end up with a much less precise and vastly more variable data set than if they considered each species separately.

To be clear, we are not advocating one style of flume over another, they have different pros and cons (e.g. the flumes recommended by Clark et al. require large volumes of treatment water to operate) and each may be suited to different circumstances (e.g. the Atema style flume is small and requires limited equipment to operate making it suited to field research in remote locations). Our primary point is that Clark et al. (2020) did not closely repeat the methods of past experiments as they claimed. Instead they changed the methodology in so many different ways that a direct comparison of results from previous studies with a multispecies data distribution derived from their own data is invalid.

<u>Activity</u>

13. "*In contrast to their suggestion that elevated CO_2 does not affect the behaviour of coral reef fishes, Clark et al.[1] found that elevated CO_2 levels increased activity by 59-92% in small Dascyllus aruanus and 50% in A. polyacanthus. This is consistent with some other studies[4], despite differences in methodologies. Clark et al.[1] did not provide fish with a shelter in 2015 or 2016 and used a vertical wall in 2014, whereas previous studies used either natural coral habitat or a horizontal pipe that provides a cave-like shelter. More generally, the effects of elevated CO_2 on activity vary considerably among studies and it is not a reliable metric of the effects of ocean acidification in fish[2].*"

Clark et al. response

We agree with Munday et al. that the effects of elevated CO_2 on activity vary greatly among studies, making it

an unreliable metric of OA effects in fish. Yet, many studies have used this metric for exactly this purpose (14 studies by Munday and colleagues in their Supplementary Data Table 1 used activity as a metric of OA effects in fish). Notably, some pioneering papers reported very strong effects of CO_2 on activity (e.g., 9000% increase; Munday et al. (2013)), suggesting that these results should be easily replicable. The effects we found are discussed in our paper (pg. 3 of Clark et al. 2020) – they are not ignored – but since those effects were the exception rather than the rule, and were not repeatable, our overall conclusion was that CO_2 is not detrimental to activity.

The shelter we used in 2014 was a halved PVC pipe standing on its end, and the fish did indeed use it for shelter. If fish activity is affected by CO_2 only if individuals have access to a shelter, then we should have seen an effect in our fish in 2014, but not in 2015 or 2016. This was not the case. In any event, it is unclear what the ecological consequences are of inconsistent effects of CO_2 on activity, especially given that any CO_2-induced changes in activity can be transient (Sundin et al., 2019).

Munday et al. reply

A number of previous studies have reported effects of elevated CO_2 on the activity of fishes, including in some of the damselfishes considered by Clark et al. (2020). In their response, Clark and colleagues have chosen to highlight a paper on juvenile coral trout, presumably because it reported a large effects size, but they do not mention other papers with damselfishes (e.g. Munday et al. 2010) that report similar changes in activity levels to those observed by Clark et al. (2020). Ignoring a more relevant result in favour of an extreme example is misleading.

Presumably, Clark and colleagues converted the change in activity of juvenile coral trout reported in Munday et al. (2013) to 9000% because it seems unreasonably excessive that way. However, they do not give the full story that places the result in context. In Munday et al. (2013) juvenile coral trout that had been exposed to either ambient conditions or elevated CO_2 were placed in an aquarium that contained a cave-like shelter (a piece of plastic pipe). Control fish spend 90% of their time hidden in the shelter, and did not venture far from the shelter when they did emerge. In contrast, fish from the elevated CO_2 treatment spent 90% of their time outside the shelter and explored the entire aquarium. Activity was scored as the number of lines crossed on a grid placed over the front of the tank. It is hardly surprising that the number of lines crossed in fish swimming outside the shelter was much greater than for fish that hid within the shelter most of the trial. The full results are shown below so the reader can judge for themselves if reporting the difference as a 9000% increase in activity is misleading, or not.

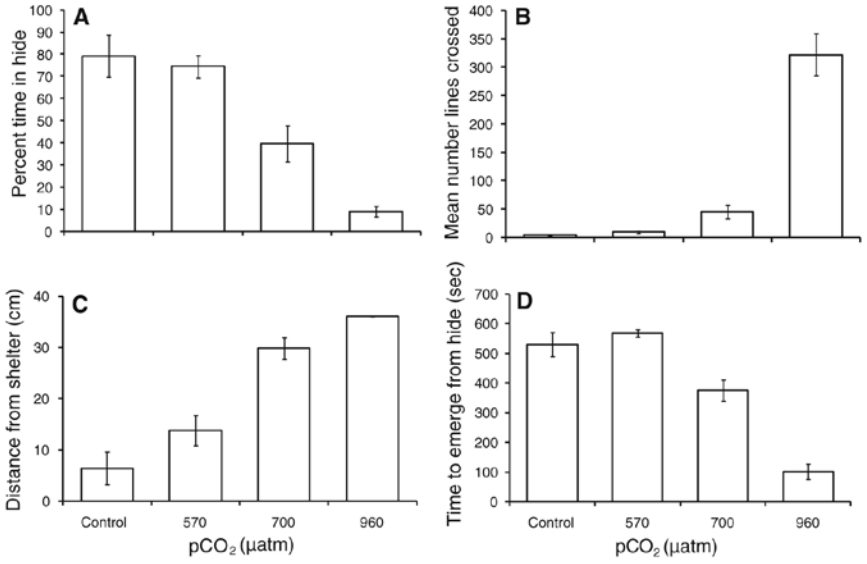


Figure 2 from Munday et al. (2013) showing percent time that juveniles coral trout from different CO_2 treatments spend in a cave-like shelter (hide), the distance ventured from shelter, and activity measured by number of lines crossed on a grid placed on the front of the tank

It also needs to be clear that the vertical half section of pipe used as a shelter in Clark et al. (2020), where the fish is still fully exposed from above, is in no way the same type of shelter as a horizontal pipe that forms a cave-like structure where the fish is fully enclosed from above and the sides.

Lateralization

14. "*Previous studies on the effects of elevated CO_2 on behavioural lateralization in coral reef fishes used species that were initially lateralized at the individual or population level[12,18-20]. There is no a priori prediction for environmental effects in non-lateralized populations. None of the four species tested by Clark et al.[1] in 2014 exhibited population-level lateralization, yet one species, Chromis atripectoralis, became significantly right lateralized in elevated levels of CO_2. Only one of the four species, Pomacentrus amboinensis, exhibited individual-level lateralization, and lateralization decreased at high CO_2 levels, as has been reported previously[7,12,18,20]. Therefore, in contrast to the conclusions by Clark et al.[1], this result supports the findings of previous studies. The effect in P. amboinensis was absent on retesting several days later, which suggests that lateralization is not stable over this timeframe, or that individuals learn about the apparatus, which alters their response in repeat trials. The repeat trial also had a lower number of fish (n = 15) than the first trial (n = 21) and the P value was marginally not significant (P = 0.068); therefore, the non-significant result of the repeat trial could simply be owing to a lower statistical power. Clark et al.[1] repeated the analysis with only the species that was significantly lateralized in the first trial, but not with any other species.*"

Clark et al. response

Munday et al. correctly point out that one of the six species tested in Clark et al. (2020) became lateralized at the population level after exposure to elevated CO_2 (*C. atripectoralis* in 2014), a result which runs counter to the predictions and findings presented in the pioneering studies listed in their fig. 1 (see Domenici et al., 2012; Domenici et al., 2014; Welch et al., 2014). Munday et al. also correctly point out that individual-level lateralization declined in one of the six species (*P. amboinensis*) after exposure to elevated CO_2, but this effect was not repeatable several days later.

Munday et al. reply

As we clearly say in our original point, above, previous studies did not consider effects on non-lateralized populations and did not have a-priori predictions for environmental effects in non-lateralized populations. Clark et al. experiments on a non-lateralized population cannot be considered to run counter the results of previous work, because previous work was not carried out on non-lateralized populations. Quite simply, studies on non-lateralized populations are a different type of work than what was carried out previously. In light of the literature suggesting that lateralization provides benefits to fish, we have difficulties to see the rationale for studying the effect of OA on lateralization in a non-lateralized population, as Clark et al. did.

Clark et al. response

We agree with Munday et al.'s interpretation of our findings. However, we also note that, although statistically significant, these results are unlikely to be biologically meaningful.

First, the difference (effect size) in both cases is small. For *C. atripectoralis* (Extended Data fig. 2b in Clark et al. 2020), the mean lateralization level (i.e., population lateralization) of CO_2-treated individuals was 5.76 turns to the right versus 5.04 turns to the right for the control group (Extended Data Table 4 in Clark et al. 2020). This result indicates only a slight deviation (less than one turn) from a random pattern of turning at the population level (i.e., 5 turns to the right and 5 turns to the left). Similarly, for *P. amboinensis*, only 3 out of 21 individuals made ≤1 or ≥9 turns to the right in the control group (the criterion required for individuals to be lateralized based on an exact binomial test), resulting in a p-value of 0.014 (Extended Data fig. 2d and Table 4 in Clark et al. 2020). Arguably, evidence for individual lateralization in this group, although "statistically significant", is weak. A

difference of two turns for one of these three fishes would have rendered these results "statistically non-significant".

Second, a large, recent study we conducted on five fish species showed that (a) behavioural lateralization based on the detour test is not a repeatable trait in fishes, and (b) previous studies did not use appropriate statistics to assess lateralization and compare treatment groups (Roche et al., 2020). The notebooks, videos, data, and analysis script for Roche et al. (2020) are publicly available: https://doi.org/10.6084/m9.figshare.6881489.v1.

Importantly, Roche et al. (2020) studied *P. amboinensis*, one of the species tested in Clark et al. (2020), and showed that, while this species can be lateralized at the individual level, this is not consistent through time: individual-level lateralization was measured on four different days and was apparent on only two of the four days (Table S2 in Roche et al. (2020)). We note that some of the other species tested by Roche et al. (2020) did display consistent individual-level lateralization across days; however, relative lateralization scores for individual fish of these species were highly variable from one day to the next (fig. 3 in Roche et al. (2020)), bringing into question the relevance/existence of consistent "individual-level lateralization" assessed using a detour test. Therefore, we agree with Munday et al.'s suggestion above that "*[behavioural] lateralization [measured using a detour test] is not stable over this timeframe*". The analysis of Roche et al. (2020) shows that the repeatability (R) of lateralization is close to nil ($0.01<R<0.08$).

Munday et al. reply

(a) Clark et al. have taken our comment about the repeatability of lateralization out of context. We repeat it in full here for the reader: "The effect in P. amboinensis was absent on retesting several days later, which suggests that lateralization is not stable over this timeframe, or that individuals learn about the apparatus, which alters their response in repeat trials. The repeat trial also had a lower number of fish (n = 15) than the first trial (n = 21) and the P value was marginally not significant (P = 0.068); therefore, the non-significant result of the repeat trial could simply be owing to a lower statistical power. Clark et al.[1] repeated the analysis with only the species that was significantly lateralized in the first trial, but not with any other species."

So we offer three alternative reasons why an effect of elevated CO_2 on lateralization was not found on retesting a few days later, not just the one explanation that Clark and colleagues chose to repeat. Given the borderline significance of the statistical test, the most likely reason Clark et al. failed to detect a significant effect on retesting was that they had insufficient statistical power because of the low sample size in the repeat trial. Why Clark et al. (2020) did not repeat test other species with borderline significance values in the first trial remains a mystery. In addition, oddly, Clark et al did not directly test if the lateralization of the elevated CO_2 fish differed from the controls, but only tested if the control and the elevated CO_2 samples were lateralized.

(b) Clark et al. claim that a recent analysis of Roche et al. (2020) shows that repeatability (R) of lateralization is close to nil ($0.01<R<0.08$). Roche's work was based on five species, including Poecilia reticulata, none of which show significant repeatability in their study. However, a recent study by McLean and Morell (2020) tested lateralization in Poecilia reticulata and found that relative lateralization is repeatable across context in both males and females, and absolute lateralization is repeatable across context in males. While this recent work was based on repeatability across context, arguably if repeatability within context is nil as suggested by Clark et al, one would not expect repeatability to occur across context either. Similarly, Domenici et al. (2017) showed that lateralization is repeatable after 6 months in a keystone marine mollusc, but not when exposed to elevated CO_2. Clearly, Roche's et al.'s generalization that lateralization is not repeatable is challenged by this recent work.

(c) It is important to point out that temporal repeatability of lateralization over several days is irrelevant to the question of whether elevated CO_2 affects behavioural lateralization, or not. Using a standard scientific experimental design, previous studies exposed one group of fishes to ambient conditions and one group of fishes to elevated CO_2. Several days later the fish were tested in a randomized order. The

two groups exhibited statistically significant differences in behavioural lateralization. The only conclusion that can be drawn from that result is that elevated CO_2 had a significant effect on behavioural lateralization in that group of fishes. If the same group of fish did not exhibit an effect of elevated CO_2 on lateralization several days later, that demonstrates something different altogether. Regardless, previous studies did not test for repeatability of lateralization.

15. "*In contrast to previous studies, Clark et al.[1] did not detect an effect of elevated CO_2 on lateralization in A. polyacanthus. However, they used an inbred population from a public aquarium that routinely experiences very high CO_2 levels (pCO_2 > 1,100µatm) and that could therefore be acclimated or adapted to high CO_2 levels.*"

Clark et al. response

See our response to point 6 regarding the population of fish from ReefHQ. We note that a study by Munday and colleagues reported that transgenerational acclimation to elevated CO_2 does not diminish the OA-induced impairments to lateralisation in *A. polyacanthus* (Welch et al. 2014). Therefore, this previous work runs counter to the above argument by Munday et al.

Munday et al. response

Our response at Point 6 is relevant here. There is ample reason to suspect that the ReefHQ population of A. polyacanthus could be adapted to high CO_2 because they have been bred for many generations as a closed population in a high CO_2 environment (> 1100 µatm CO_2). Furthermore, Welch et al. (2014) does not support the conclusion that transgenerational acclimation to elevated CO_2 does not occur in the Reef HQ population, because Welch et al. only exposed reproductive adults to elevated CO_2 during the breeding season, whereas adults in the Reef HQ population have lived their whole life in high CO_2 over several generations, which can be important to the induction of beneficial transgenerational plasticity (Donelson et al. 2018).

<u>Visual response</u>

16. "*Clark et al.[1] tested visual acuity by shortening one side of the barrier in the T-maze. This is not a recognized visual acuity test and is not a replication of a visual threat stimulu[21] or direct measurement of visual acuity by electroretinogram, as used in previous studies[22].*"

Clark et al. response

We took the opportunity in Clark et al. (2020) to determine whether the reported results from previous papers by Munday and colleagues about OA-impaired vision could be verified using what we believe is a functionally relevant test (detecting whether fish could visually determine the shortest path around an obstacle when being ushered down a channel). We found no evidence to support the claim of OA-induced visual impairment.

We note that Munday and colleagues took the unusual step of using a planktivorous (non-predatory) species (the damselfish *A. polyacanthus*) to investigate how other damselfish (*P. amboinensis*) respond to a visual "threat" stimulus (Ferrari et al., 2012b). The authors reported a strong reduction in "antipredator" responses of *P. amboinensis* exposed to elevated CO_2 when presented in a plastic bag to naturally co-occurring *A. polyacanthus*. Surprisingly, *P. amboinensis* exposed to high CO_2 were attracted to the other damselfish species (*A. polyacanthus*) and exhibited none of the responses typically adopted by juveniles of this species in response to threatening situations. If elevated CO_2 is so disruptive to visual acuity that it causes fish to swim towards threatening situations, we would have expected to see evidence of this visual impairment in the visual test we conducted in Clark et al. (2020).

Munday et al. reply

At Point 14, above, Clark and colleagues argue that the t-maze apparatus should not be used for testing behavioural lateralization, so it is puzzling that they should now decide it is useful to test visual acuity, when it was not designed for this purpose and has not previously been used for this purpose. Moreover, it is hard to imagine two more divergent methods for testing visual acuity than: (i) examining turning direction in a structurally alerted t-maze, as Clark et al. have done, compared with (ii) direct physiological measurement of retinal responses using an electroretinogram as the earlier study by Chung et al. (2014). These methods have no similarity whatsoever.

Clark and colleagues statement, above, about Munday and colleagues taking the "unusual step of using a planktivorous (non-predatory) species (the damselfish A. polyacanthus) to investigate how other damselfish (P. amboinensis) respond to a visual "threat" stimulus" is highly misleading, because it fails to point out the difference in size and life stages of the two groups of fish. The focal fish (P. amboinensis) were settlement-stage juveniles, approximately 14 mm long and naïve to larger reef resident fishes. The threat stimulus was an adult A. polyacanthus, approximately 124 mm long, placed in a plastic bag. As explained in the original study, the much larger fish would represent a potential threat to the much smaller naïve juvenile (Ferrari et al. 2012). The fact the large fish is planktivorous is irrelevant, it's the large size that causes that fish to be a potential threat to a small juvenile that has no experience of reef resident fishes (predators or otherwise).

<u>Concluding remarks</u>

Despite all of the variation in methods we have described above between our work and that of Clark et al. (2020), these authors fail to concede any possibility that this may be a significant source of the differences in outcomes of our studies, insisting yet again that "independent and transparent" studies must be conducted in the future. We find it more than a little ironic that their approach is likely to have done nothing more than impede this taking place. By taking an accusatory, antagonistic approach and failing to engage with us at the outset, Clark et al (2020) have devolved important issues of replication into little more than long lists of claim and counter-claim. As Nosek and Errington (2020) show, this does nothing to advance any field of research, nor does it confer on Clark et al. any moral high ground. In fact, given their willingness to accuse others, it has probably acted to retard progress in ocean acidification research. Why would any young researcher now want to embark on studies in this field when the consequences of doing so might be veiled accusations of misconduct? Nosek and Errington (2020) provide a roadmap for engagement of researchers on opposite sides of a debate involved in replicating past studies. We urge Clark et al to pay attention to it so that real progress on this important scientific issue can be made.

References used in Munday et al. reply

Atema, J., Kingsford, MJ, Gerlach, G. (2002). Laval reef fish could use odour for detection, retention and orientation to reefs. Marine Ecology Progress Series. 241: 151-160.

Clark et al. (2020). Ocean acidification does not impair the behaviour of coral reef fishes. Nature, https://doi.org/10.1038/s41586-019-1903-y.

Clements, J.C. and Hunt, H.L. (2015). Marine animal behaviour in a high CO_2 ocean. Marine Ecology Progress Series, 536 :259-279.

Coppock, A. G., Gardiner, N.M., Jones, G. P. (2016). Sniffing out the competition? Juvenile coral reef damselfishes use chemical cues to distinguish the presence of conspecific and heterospecific aggregations. Journal of Behavioral Processes 125: 43-50

Cripps, I.L., Munday, P.L., McCormick, M.I. (2011). Ocean acidification affects prey detection by a predatory reef fish. PLoS ONE, 6: e22736.

Chivers, D.P. and Mirza, R.S. (2001). Predator diet cues and the assessment of predation risk by aquatic vertebrates: A review and prospectus. Chemical Signals in Vertebrates, 9: 277-284.

Chivers, D.P., et al. (2013). Degradation of chemical alarm cues and assessment of risk throughout the day. Ecology and Evolution, 3: 3925–3934.

Chivers, D. P., et al. (2014a). Background level of risk influences learning of predators and non-predator by coral reef damselfish. Proceedings of the Royal Society B, 281: 20140355.

Chivers, D.P., et al. (2014b). Impaired learning of predators and lower prey survival under elevated CO_2: a consequence of neurotransmitter interference. Global Change Biology, 20: 512-522.

Chung, W-S., et al. (2014). Ocean acidification slows retinal function in a damselfish through interference with GABA-A receptors. Journal of Experimental Biology, 217: 323-326.

Dixson D.L., Munday P.L., Jones G.P. (2010). Ocean acidification disrupts the innate ability of fish to detect predator olfactory cues. Ecology Letters, 13: 68-75.

Domenici, P., et al. (2014). Shifting from right to left: the combined effect of elevated CO2 and temperature on behavioural lateralization in a coral reef fish. PLoS One, 9: e87969. doi:10.1371/journal.pone.0087969.

Domenici, P., Torres, R., Manríquez, P.H. (2017). Effects of elevated carbon dioxide and temperature on locomotion and the repeatability of lateralization in a keystone marine mollusk. Journal of Experimental Biology 220, 667-676.

Donelson, J.M. et al. (2018). Transgenerational plasticity and climate change experiments: where do we go from here? Global Change Biology, 24: 13-34.

Ferrari, M.C.O., Wisenden, B.D., Chivers, D.P. (2010). Chemical ecology of predator-prey interactions in aquatic ecosystems: a review and prospectus. Canadian Journal of Zoology, 88: 698-724.

Ferrari. M.C.O., et al. (2011). Intrageneric variation in antipredator responses of coral reef fishes affected by ocean acidification: implications for climate change projections on marine communities. Global Change Biology, 17: 2980-2986.

Ferrari, M.C.O. et al. (2012). Effects of ocean acidification on visual risk assessment by coral reef fishes. Functional Ecology, 26: 553-558.

Ferrari, M.C.O., et al. (2015). Interactive effects of ocean acidification and rising sea temperatures alter prey mortality and predator selectivity in reef fish communities. Global Change Biology, 21: 1848-1855.

Ferrari, M.C.O., et al. (2017). Predation in high CO_2 waters: prey fish from high-risk environments are less

susceptible to ocean acidification. Integrative and Comparative Biology, 57: 55-62.

Gerlach, G., et al. (2007). Smelling home can prevent dispersal of reef fish larvae. Proceeding of the National Academy of Science, 104: 858-863

Gould, A.L. Saki, H.V., Dunlap, P.V. (2015). Cues from the reef: olfactory preferences and a symbiotically luminous cardinalfish. Coral Reefs, 34: 673-677.

Hannan, K.D., et al. (2020). Diel pCO_2 variation among coral reefs and microhabitats at Lizard Island, Great Barrier Reef. Coral Reefs. doi: 10.1007/s00338-020-01973-z.

Heuer, R., et al. (2016). Altered brain ion gradients following compensation for elevated CO_2 are linked to behavioural alterations in a coral reef fish. Scientific Reports, 6: 33216. doi: 10.1038/srep33216.

Jarrold, M.D., Humphrey, C., McCormick, M.I., Munday, P.L. (2017). Diel CO_2 cycles reduce severity of behavioural abnormalities in coral reef fish under ocean acidification. Scientific Reports, 7: 10153. doi:10.1038/s41598-017-10378-y.

Jarrold, M.D. and Munday, P.L. (2018). Elevated temperature does not substantially modify the interactive effects between elevated CO_2 and diel CO_2 cycles on the survival, growth and behaviour of a coral reef fish. Frontiers in Marine Science, 5: 2018.00458. doi: 10.3389/fmars.2018.00458.

Kusch, R.C., Mirza, R.S., Chivers, D.P. (2004). Making sense of predator scents: investigating the sophistication of predator assessment abilities of fathead minnows. Behavioral Ecology and Sociobiology, 55: 551-555.

Lönnstedt OM, McCormick MI (2011a). Chemical alarm cues inform prey of predation threat: the importance of ontogeny and concentration in a coral reef fish. Animal Behaviour, 82: 213-218.

Lönnstedt OM, McCormick MI (2011b). Growth history and intrinsic factors influence risk assessment at a critical life transition for a fish. Coral Reefs, 30: 805-812.

Mathis, A. and Smith, R.J.F. (1993). Fathead minnows, Pimephales promelas, learn to recognize northern pike, Esox lucius, as predators on the basis of chemical stimuli from minnows in the pike's diet. Animal Behaviour, 46: 645-656.

McCormick, M.I., Watson, S-A., Munday, P.L. (2013). Ocean acidification reverses competition for space as habitats degrade. Scientific Reports, 3: 03280. doi:10.1038/srep03280.

Miller-Sims, V.C. et al. (2011). How stable are the reef odor preferences of settling fish larvae? Marine and Freshwater behavior and physiology, 44: 133-141.

Mirza, R.S. and Chivers, D.P. (2002). Brook Char (Salvelinus fontinalis) can differentiate chemical alarm cues produced by different age/size classes of conspecifics. Journal of Chemical Ecology, 28: 555–564.

Mitchell, M.D., et al. (2015). Learning to distinguish between predators and non-predators: understanding the critical role of diet cues and predator odours in generalisation. Scientific Reports, 5: 15537.

McLean, S. and Morrell, L. J. (2020). Consistency in the strength of laterality in male, but not female, guppies across different behavioural contexts. Biology Letters, 1620190870.

Munday, P.L. et al. (2009). Ocean acidification impairs olfactory discrimination and homing ability of a marine fish. Proceedings of the National Academy of Sciences USA, 106: 1848-1852.

Munday P.L., et al. (2010). Replenishment of fish populations is threatened by ocean acidification. Proceedings of the National Academy of Sciences USA, 107: 12930-12934.

Munday, P.L., et al. (2012). Selective mortality associated with variation in CO_2 tolerance in a marine fish. Ocean Acidification, 1: 1-5. doi: 10.2478/oac-2012-0001.

Munday, P.L., et al. (2013). Elevated CO_2 affects the behaviour of an ecologically and economically important

coral reef fish. Marine Biology, 160: 2137-2144.

Munday, P.L., et al. (2013b). Predicting evolutionary responses to climate change in the sea. Ecology Letters, 16: 1488-1500.

Munday, P.L., et al. (2014). Behavioural impairment in reef fishes caused by ocean acidification at CO2 seeps. Nature Climate Change, 4: 487-492.

Munday, P.L., Welch, M.J., Allan, B.J.M., Watson, S-A., McMahon, S.J., McCormick, M.I. (2016). Effects of elevated CO_2 on predator avoidance behaviour by reef fishes is not altered by experimental test water. PeerJ, 4: e2501.

Nagelkerken et al. (2016). Ocean acidification alters fish populations indirectly through habitat modification. Nature Climate Change, 6: 89–93.

Nilsson, G.E, et al. (2012). Near-future CO_2 levels alter fish behaviour by interference with neurotransmitter function. Nature Climate Change, 2: 201-204

Nosek, B.A. and Errigton, T.M. (2020). The best time to argue about what a replication means? Before you do it. Nature, 583: 518-520. doi: 10.1038/d41586-020-02142-6.

Ou et al. (2015). Responses of pink salmon to CO2-induced aquatic acidification. Nature Climate Change, 5: 950-957.

Paula, et al. (2019). The past, present and future of cleaner fish cognitive performance as a function of CO_2 levels. Biology Letters, 15: 20190618.

Schunter, C., et al. (2016). Molecular signatures of transgenerational response to ocean acidification in a reef fish. Nature Climate Change, 6: 1014-1018.

Simpson, S.D., et al. (2011). Ocean acidification erodes crucial auditory behaviour in a marine fish. Biology Letters, 7: 917-920.

Smith, R.J.F. (1992). Alarm signals in fishes. Reviews in Fish Biology and Fisheries, 2: 33-63.

Thomas, J.T., Munday, P.L., Watson, S-A. (2020). Toward a mechanistic understanding of marine invertebrate behavior at elevated CO2. Frontiers in Marine Science, 7: 345. doi:10.3389/fmars.2020.00345.

Watson, S.-A. et al. (2017). Quantifying pCO_2 in biological ocean acidification experiments: a comparison of four methods. PLoS One, 12: e0185469.

Welch, J.M., et al. (2014). Effects of elevated CO_2 on fish behaviour undiminished by transgenerational acclimation. Nature Climate Change, 4: 1086-1089.

Welch, M.J. and Munday, P.L. (2017). Heritability of behavioural tolerance to high CO_2 in a coral reef fish is masked by non-adaptive phenotypic plasticity. Evolutionary Applications, 10: 682-693.