Replication alert: behavioural lateralisation in a detour test is not repeatable in fishes

Dominique G. Roche[1], Mirjam Amcoff[2], Rachael Morgan[3], Josefin Sundin[3,4], Anna H. Andreassen[3], Mette H. Finnøen[3], Michael J. Lawrence[5], Eleanor Henderson[6], Tommy Norin[7,8], Ben Speers-Roesch[9], Culum Brown[10], Timothy D. Clark[11], Redouan Bshary[1], Brian Leung[12], Fredrik Jutfelt[3], Sandra A. Binning[1,12]

Correspondence to: dominique.roche@mail.mcgill.ca

[1] Éco-Éthologie, Institut de Biologie, Université de Neuchâtel, Neuchâtel, Switzerland
[2] Department of Zoology/Functional Zoomorphology, Stockholm University, Stockholm, Sweden
[3] Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway
[4] Department of Neuroscience, Uppsala University, Uppsala, Sweden
[5] Fish Ecology and Conservation Physiology Lab, Carleton University, Ottawa, Ontario, Canada
[6] School of Life Sciences, University of Technology Sydney, Sydney, New South Wales, Australia
[7] Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, Scotland, UK
[8] National Institute of Aquatic Resources, DTU Aqua, Kgs. Lyngby, Denmark
[9] Department of Biological Sciences, University of New Brunswick, Saint John, New Brunswick, Canada
[10] Department of Biological Sciences, Macquarie University, Sydney, Australia
[11] Deakin University, Geelong, Victoria, Australia, School of Life and Environmental Sciences
[12] Department of Biology, McGill University, Montréal, Québec, Canada
[13] Département de sciences biologiques, Université de Montréal, Montréal, Québec, Canada

Abstract

Behavioural lateralisation, defined as the asymmetric expression of cognitive functions, is reported to enhance key fitness-relevant traits such as predator escape performance, multitasking abilities, and group coordination. Therefore, studies reporting negative effects on lateralisation in fish due to environmental stressors such as ocean acidification, hypoxia, and pollutants are worrisome. However, such studies have focussed on population-level measures, without validating whether lateralisation is consistent within individuals across time. We conducted a multi-species, international assessment of the repeatability (R) of lateralisation in four previously studied fish species using the common detour test, and re-analysed a published dataset (on guppies) using new statistical methods. We expected the three shoaling species to exhibit greater within-individual consistency in lateralisation than their non-shoaling counterparts given previous reports of stronger lateralisation in group-living fishes. However, both absolute and relative lateralisation scores were highly non-repeatable in all five species ($0.01<R<0.08$). Thus, the commonly used detour test does not appear to be appropriate for quantifying behavioural lateralisation in fishes, calling into question inferences drawn by many published studies, including our own. As a consequence, potential anthropogenic effects on lateralisation as a proxy for adaptive brain functioning need to be assessed with alternative paradigms.

Keywords: Behavioural plasticity, *Ctenolabrus rupestris*, *Danio rerio*, laterality, lateralization, *Neopomacentrus azysron*, *Poecilia reticulata*, *Pomacentrus amboinensis*, repeatability, T-maze

Introduction

Behavioural lateralisation, the asymmetric expression of cognitive functions, has emerged as an important fitness correlate in numerous taxa including invertebrates [e.g. 1] and vertebrates [reviewed in 2]. Indeed, cerebral lateralisation is believed to offer advantages in terms of enabling multiple stimuli to be processed simultaneously by different sides of the brain [2]. Fitness benefits associated with high degrees of lateralisation are thought to include increased cognitive performance [3, 4], multitasking [5, 6], spatial learning [7], predator recognition learning [8], schooling performance [9], coordination of group behaviours [2], prey capture success [10], foraging efficiency [11], and escape performance [12, 13]. Despite these apparent advantages, most studies report considerable inter- and intra-specific variation in the strength and direction of lateralisation. Several studies also report a high degree of within-individual variation in the strength of lateralisation in individuals measured repeatedly across different contexts, suggesting that this trait may be more labile than previously believed [e.g. 8, 14]. Yet, no studies have systematically measured lateralisation several times per individual, across multiple species, without a change in context, to establish the baseline repeatability of this behaviour (but see [15] for a single-species study, the data for which are re-analyzed here). Such a validation of the repeatability of lateralisation is crucial to assess the usefulness of this trait, particularly when differences in individual-level lateralisation strength between ecologically relevant treatments are used to infer effects on fitness.

Behavioural lateralisation is commonly measured in the laboratory using a detour test (e.g. T-maze), wherein an individual moves along a runway until it faces a barrier forcing it to make a choice between turning left or right. This quick and simple test is now widely used in studies of fish behaviour, evolutionary ecology and ecotoxicology (Electronic Supplementary Material [ESM], Table S1), and is also used to assess lateralisation in various other taxa including birds [16], reptiles [17], and molluscs [1]. In fishes, lateral bias in a detour test is believed to arise from asymmetries in eye-use, a phenomenon widely documented in animals with laterally-placed eyes and low binocular overlap [18]. For example, goldbelly topminnows (*Girardinus falcatus*) reported to exhibit a strong side-bias in a detour test also display an eye preference for viewing a neutral *versus* threatening stimulus [19]. Similarly, left-right asymmetries in *G. falcatus* assessed in a detour test appear to be consistent with lateral bias measured using other methods [20]. Furthermore, lateralisation score in a detour test is reported to be heritable [21, 22], although lateralisation strength decreases rapidly across generations in artificial selection lines [23, 24]. These results have led researchers to assert that the detour test does, indeed, assess inherent asymmetry in an individual's brain function that influences fitness-relevant behavioural strategies.

Given likely fitness-relevant effects of lateralisation and the relative ease of use of the detour test, various studies have evaluated how environmental and anthropogenic stressors such as pollutants and climate change affect lateralisation in fishes, several of which have produced worrying results (ESM, Table S1). Most notably, ocean acidification is reported to decrease lateralisation across a range of marine fishes, including tropical [25-27] and temperate species [28-30]. Such effects are concerning given, for example, the important benefits of lateralisation reported for fishes under high predation risk [e.g. 8, 31, 32]. Despite this strong interest in behavioural lateralisation and growing use of the detour test in the ecological, cognitive and behavioural sciences, there is a notable absence of studies assessing the repeatability of individual lateralisation scores. Establishing the consistency of lateralisation within individuals is essential for determining the relevance of this trait for assessing fitness as well as the responses of animals to exogenous stressors.

To evaluate the repeatability of behavioural lateralisation in fishes, we studied four species from tropical and temperate environments, including marine and freshwater habitats: two coral reef damselfishes, the Ambon damsel (*Pomacentrus amboinensis*) and yellowtail demoiselle (*Neopomacentrus azysron*), the tropical freshwater zebrafish (*Danio rerio*), and the temperate marine goldsinny wrasse (*Ctenolabrus rupestris*). All of these species have previously been used in published studies assessing behavioural lateralisation using a detour test (T-maze) similar to the one used here [8, 25, 26, 33, 34]. We also used new (more appropriate) statistical methods to re-analyse a published dataset that included multiple lateralisation measurements for the tropical freshwater guppy (*Poecilia reticulata*) [15]. We predicted a significant repeatability in lateralisation scores comparable in magnitude to that observed for other behavioural traits [35]. Population-level lateralisation strength was also predicted to be higher in the shoaling species (*N. azysron* and *D. rerio*) compared to territorial species (*P. amboinensis* and *C. rupestris*) as this should help enhance school cohesion [36]. We had no *a priori* prediction about the direction (left or right) of lateralisation, as among-species differences in response to the same stimulus have been shown previously [36]. We also examined for an effect of habituation to the test arena across the four test trials because (1) repeatedly measuring the same individuals in a similar apparatus can decrease measurement error and thus be predicted to increase individual repeatability [37], and conversely (2) habituation or desensitisation to a stimulus may reduce repeatability by generating different results between trials within an individual [37].

Materials and methods

Fish collection and husbandry

Sixty *C. rupestris* (1.38-46.52 g) were collected with baited minnow traps from the wharfs at the Sven Lovén Centre for Marine Infrastructure, Kristineberg, Sweden. Experiments on *C. rupestris* were conducted between 22 June and 1 July 2017 (water temperature range in the holding tanks and experimental setup: 14.9-16.1°C). Sixty *N. azysron* (0.74-6.40 g) and 60 *P. amboinensis* (1.0-14.3 g) were collected by SCUBA divers with a barrier net and hand nets on coral reefs nearby the Lizard Island Research Station, Northern Queensland, Australia. Experiments on these two species were conducted between 24 July and 11 August 2017 (water temperature range: 24.3-25.5°C). Sixty *D. rerio* (0.26-1.07 g) were collected in West Bengal, India, in August to September 2016 using hand nets and kept in the Norwegian University of Science and Technology's (NTNU) aquarium facility, Trondheim, until the experiments were conducted between 10 and 17 October 2017 (water temperature range: 26.9-28.9°C).

Fish from each species were equally divided into three to four holding tanks (density ~5.5 individuals L^{-1}), provided with artificial plants and/or cut PVC pipes as shelter, and fed *ad libitum* twice daily. *C. rupestris* were fed blue mussels (*Mytilus edulis*) and bloodworms. *Neopomacentrus azysron* and *P. amboinensis* were fed commercial tropical fish flakes. *Danio rerio* were fed TetraPro fish flakes. Holding tanks were supplied with flow-through water from the fish's natural habitat for *C. rupestris*, *N. azysron*, and *P. amboinensis*. *D. rerio* were kept in flow-through aquaria at NTNU, with a water exchange rate of ~20% per day, and water quality (conductivity and temperature) monitored daily. Light cycles followed natural conditions at each location. Fish were individually marked using two-colour combinations of Visible Implant Elastomer (Northwest Marine Technology Inc., Shaw Island, WA, USA) implanted postero-dorsally, on both sides of the dorsal fin a minimum of two days prior to the first test.

Lateralisation test

We used a standard detour test to assess behavioural lateralisation [12, 30, 38]. Experiments were conducted in the morning between 9:00-12:30 and fish were returned to their respective holding tank between test days. A single fish was introduced into a double-sided, opaque T-maze, consisting of a tank with a runway down the middle and a barrier at both ends (dimensions in Fig. S1). *Ctenolabrus rupestris* and *D. rerio* were tested in one maze and *N. azysron* and *P. amboinensis* in another (see ESM Fig. S1). A symmetrical neutral stimulus was affixed to the centre of the barrier to focus the eye-gaze of the fish. This neutral stimulus was changed between each of the four trials (see below) to prevent habituation to the set-up (a cross, two parallel black bars, a cross with a solid circle above it, a cross with a horizontal bar below it). The water height was 10 cm. Experiments were conducted in a closed room and care was taken to ensure that the surrounding environment and lighting were as uniform as possible. We emptied and re-filled the maze with new water between each fish tested to standardise the temperature, maintain normal levels of dissolved oxygen, and to avoid any effect of conspecific stress hormones (e.g. cortisol) on fish behaviour.

To start the experiment, a fish was placed at one end of the runway (maintained in this position by an acrylic divider blocking the entrance to the middle runway) for one minute (ESM Fig. S1). The starting side was determined by flipping a coin and the fish was transferred from its holding tank to the maze in a water-filled container. The divider was then lifted from behind the tank; if the fish did not advance on its own, it was gently pushed with an acrylic paddle to initiate movement down the runway [30, 39]. At the end of the runway, the fish faced an opaque barrier forcing it to turn left or right. We ran 10 consecutive trials per fish, recording its turning direction each time. One experimenter conducted any given series of 10 trials, standing behind the fish, in a centred position, and walked from one side of the maze to the other between trials when the fish was behind the barrier (Fig. 1). The side of the maze walked on was random. Another experimenter recorded the observations on paper except for *D. rerio*, where observations were typed directly into a computer. Experiments were recorded on video with the exception of *C. rupestris*, which was the first species tested and where no video camera was available. Four series of 10 trials were repeated for each individual at intervals of 48 h. Four full series could not be obtained for some individuals as fish occasionally did not swim down the runway for 10 consecutive runs.

We calculated the relative lateralisation index (L_R) for each individual for each series of 10 trials, where L_R = ((turns to the right − turns to the left) / (turns to the right + turns to the left)) × 100 [18]. A score of −100 indicates that the fish turned left 10 times out of 10; a score of 100 indicates that the fish turned right 10 times out of 10. L_R can be examined at both the individual and the population (i.e. mean) level, informing whether individuals and/or populations have a side bias, respectively. Since several studies conduct analyses on the absolute lateralisation index (L_A), rather than L_R, we also calculated each individual's L_A for each series, where $L_A = |L_R|$ (ESM Fig. S3). Scores between 60 and 100 (i.e. 8 and 10 turns to one direction, respectively) are indicative of strong lateralisation.

Re-analysis of published data

To our knowledge, the only published study testing the consistency of individual lateralisation scores was performed on 40 female guppies (*P. reticulata*) [15]. Females were individually marked and tested in a detour test with 24 h or more between trial series. Fish were tested three times in each of two conditions: a barrier presenting a neutral stimulus (an empty aquarium behind a barrier of vertical bars) or a social stimulus (a conspecific in an aquarium behind a barrier of vertical bars) (Fig. 1 in [15]). For lack of a better statistical approach at the time, the data were analysed using six Spearman rank correlations to assess the consistency of L_R scores [15]. We re-analysed these data and computed a single intra-class correlation coefficient (R) for each of the two test conditions (see *Statistical analysis*).

Statistical analysis

We tested population-level lateralisation with generalised (binomial) linear random-effects models, setting the intercept equal to the grand mean of the sample. Individual-level lateralisation was examined with a chi-square test comparing the observed variance (numerator) to the expected variance (denominator) assuming a normal approximation to the binomial distribution. This is analogous to testing for overdispersion (i.e. are there more observations in the tail ends of the distribution than expected by chance). See the ESM text and Table S3 for details and an explanation of issues with tests of lateralisation employed in previous studies.

We computed the repeatability (R) in the number of left and right turns (in a series of 10 trials) across trial series for each species using the 'rpt' function in the package 'rptR' [40]. This analysis on binomial data amounts to testing the repeatability of L_R. R values were computed for primary data collected in this study and for the published data on *P. reticulata* [15]. R ranges from 0 (non-repeatable) to 1 (fully repeatable) and provides a standardised measure of the consistency of phenotypes across time or contexts [41]. We specified a binomial error distribution and 1000 bootstrapping and permutation iterations to calculate

95% confidence intervals (CIs). We also computed an 'adjusted repeatability' by specifying trial series (1 to 4), fish mass, and start-side of the maze as fixed effects in the models to account for any confounding effects of these variables on R [41]. The repeatability of L_A was calculated by specifying a Poisson error distribution. Finally, we tested the effect of experimenter ID and arena start side on L_R using a generalised linear mixed-effects model (GLMM) in the package 'lme4'; species, start side, and their interaction were specified as fixed factors in the model while fish ID and experimenter ID were included as random factors. The importance of experimenter ID was tested with a likelihood ratio (LR) test. Models were validated via diagnostic checks with the package 'DHARMa' [42]. Statistical analyses were performed in R3.4.3 [43].

Results

Fish exhibited statistically significant individual-level lateralisation in two-thirds of trial series, i.e. 15 of 22 trial series across the four species tested and the species for which data were re-analysed (ESM Table S2 and Fig. S5-S9). Population-level lateralisation was statistically apparent in five of the 22 trial series; two species exhibited a population side bias, but this bias changed with test day for *D. rerio*, and stimulus type for *P. reticulata* (ESM Table S2, Fig. S5-S9).

Importantly, however, L_R was highly variable within individuals and individual identity explained less than 6% of the variance in relative lateralisation across the four species tested, with repeatability estimates ranging from R=0.006 to R=0.028 (Table 1, Fig. 2, ESM Fig. S2). Controlling for the order of trial series, fish mass, and start-side did not change these results (Table 1); fixed factors accounted for less than 4% of the variance in L_R. Similar results were obtained when considering L_A (ESM Fig. S3). The identity of the experimenter (LR test; $\chi^2_{(1)}$=0, P=0.999) and the starting side of a series in the experimental arena had no effect on a fish's preference to turn left or right (GLMM; $\chi^2_{(1)}$=0.01, P=0.917), irrespective of the species (GLMM; species × start-side interaction; $\chi^2_{(3)}$=2.00, P=0.575); see also Table 1.

Similarly to the four species tested here, individual differences among female *P. reticulata* [15] accounted for very little of the variance in L_R, both when fish were tested using a neutral (R=0.045 [0.015-0.088]) and a social stimulus (R=0.076 [0.028-0.128]) (ESM Fig. S4).

The 95% CIs for the R estimates of L_R in two species (*N. azysron*, *D. rerio*) overlapped zero; in the other three species, 95% CIs were narrow and close to zero, indicating high confidence in very small (albeit non-zero) values of R (Table 1).

Discussion

Numerous studies on fishes and other taxa report that behavioural lateralisation is linked to fitness-enhancing behaviours and can be severely impacted by environmental stressors such as predation, parasitism, pollutants, drugs, and various abiotic parameters (see ESM Table S1 for examples in fishes). We assessed whether lateralisation is consistent in fish species with differing ecologies to establish the baseline repeatability of this behavioural trait and broaden our understanding of its ecological and evolutionary importance. Our validation exercise used robust sample sizes, four species, multiple measurements through time (four repeats at 48 h intervals), and included a re-analysis of data on a fifth species collected seven years ago.

Our results show that behavioural lateralisation as assessed by a detour test is not a repeatable trait in fishes (Fig. 2, ESM Fig. S2-S4), and thus should be interpreted with great care. Even though many individuals from all species displayed a strong side-bias (individual-level lateralisation present in 68% of trial series; ESM Table S2, Fig. S5-S9), this turning preference varied markedly across days (Fig. 2, ESM Fig. S4). Of the 69 individuals across all five species displaying an absolute lateralisation (L_A) score of 80 or 100 (nine or 10 turns in either direction) in at least one of the trial series, 52 also had at least one L_A score of 20 (four or six turns in either direction) or zero (five turns each way) when tested again (ESM Fig. S3, S4). Overall, across the five species examined, individual identity explained only a small proportion of the phenotypic variance in lateralisation strength, ranging between 2.8 and 7.6%. These values contrast markedly with the results of a recent meta-analysis reporting that individual differences account for a much higher proportion (approximately 35%) of the variance in most behavioural traits [35]. A re-analysis of existing data (ESM Fig. S4) supports our findings, and further suggests that individual decisions to turn left or right in the detour test are random and do not provide accurate or precise estimates of eye-preference or cerebral asymmetries in fishes. Our experiments were conducted by two or more researchers at any one time and trials were video recorded except for *C. rupestris*. These recordings, lab notebooks, and the raw data are publicly available [44]. The fact that multiple labs collaborated to conduct this study across several geographic locations (Australia, Norway, Sweden) further strengthens the robustness of our results [45].

Only two of the five species exhibited a mean L_R significantly different from zero in at least one trial series, indicative of population-level lateralisation (ESM Table S2, Fig. S5-S9). This population-level side-bias varied between days in the case of *D. rerio*, and stimulus type in the case of *P. reticulata* (see methods in [15]). We anticipated a greater repeatability of L_R and stronger evidence for a positive or negative mean L_R in the three shoaling species examined (*N. azysron, D. rerio* and *P. reticulata*), as individual- and population-level lateralisation are reported to help social individuals coordinate group behaviours and enhance school cohesion [9, 36]. Bisazza & Dadda [9] used a detour test and reported that lateralised poecilids (*Girardinus falcatus*) exhibit greater school cohesion and coordination than non-lateralised conspecifics; however, schools were composed of only two individuals, which were females from a multi-generation laboratory strain. Bisazza et al. [36] also reported significant population-level lateralisation in 10 of the 16 fish species they examined (*n* per species: 7 to 18), yet only six of these were shoaling. Other studies have also reported such population-level side biases in fishes [e.g. 15, 18, 19, 38, 46], yet their association with group living appears ambiguous. For instance, Domenici et al. [26] and Lopes et al. [28] failed to observe population-level lateralisation in shoaling *N. azysron* juveniles and *Atherina presbyter* larvae, respectively, despite reporting L_A scores higher than random in their control groups. Similarly, Chivers et al. [32] reported high L_R scores in some shoals of the schooling fusilier *Caesio teres*, but lateralisation strength varied substantially both among and within the four groups tested. Taken together, these results suggest that that the detour test is not adequate for assessing lateralisation in fishes and that the benefits of lateralisation for shoaling species might be overstated.

Implications

Given the large body of literature reporting significant effects of environmental stressors on lateralisation in fishes (ESM Table S1), our results raise several questions that we address below.

Could methodological differences between ours and previous studies explain the lack of repeatability in L_R?

Published studies have used a range of different obstacles to elicit eye-use preference when fish arrive at the end of the runway in a detour test (Fig. S1). Some studies use a neutral obstacle, such as an opaque barrier [e.g. 12, 14, 20, 26, 46] or a barrier of vertical bars [e.g. 15, 18, 47, 48]. Preferential eye-use is said to occur because fish must explore the unknown space to the side or behind the barrier [19]. Other studies have used a non-neutral obstacle with a stimulus, such as a conspecific or an object resembling a predator placed behind a barrier of vertical bars [e.g. 5, 15, 19, 38]. The neutral stimuli (e.g. a cross on an opaque barrier) used in our experiments may not have been valent enough to result in a strong, consistent lateralised response in individuals. Indeed, some research suggests that lateralisation direction and strength is stimulus-dependent [18, 49]. It is possible that a predatory stimulus would have increased repeatability of measurements in this test since a consistent behavioural response to a predator may be under stronger directional selection than a neutral stimulus. However, previous studies using different methods for assessing lateralisation in fishes (i.e. mirror tests or circular arena tests) have reported correlations in the strength of individual lateralisation among tests, including between novel/neutral, predator, and social stimuli [20, 50]. These results suggest that the specific stimulus used to focus gaze should not significantly affect the strength of repeatability in the test. Our results provide evidence for this: we found that L_R was not repeatable across time when experiments were conducted with two different neutral stimuli (this study and re-analysis of [15]) as well as with a non-neutral stimulus (re-analysis of [15]). The occurrence of high individual L_R scores in all species in trial series 2, 3, and 4 of the experiment indicates that habituation to the test arena is unlikely (Fig. 2). Importantly, studies have also reported a significant population-level side-bias in species when only an opaque barrier with no stimulus was used [e.g. 31, 32, 38], suggesting that a neutral stimulus should not impede a strongly lateralised response in a detour test. Interestingly, Ferrari et al. [8] found considerable variability in the lateralisation scores of *P. amboinensis* tested twice on the same day using opaque barriers with no stimuli. These results were interpreted as representing adaptive plasticity in lateralisation strength due to different predation risk scenarios [8], yet these experiments lacked a control group establishing the consistency of lateralisation strength in untreated individuals. Our results suggest that such variability is likely due to the inherent randomness of turning choice displayed by tested individuals rather than any adaptive behavioural decision. Similarly, a recent reanalysis of a well-cited study on honeybee magnetoreception also revealed random patterns rather than adaptive behaviour as originally suggested [51].

Numerous other methodological variations exist across published studies that are worth considering in the context of our results. For example, studies differ in their maze design (e.g. single T in [14]; Y- *versus* T-entry in [15, 30, 52]) and dimensions (ESM Table S1), acclimation time before beginning a series (3 min in [14, 20]; 1 min in [39]), wait time between trials within a series (3 min in [14]; no wait time in [34, 39]), and method for encouraging fish to enter runway (e.g. [33] *versus* [34]), to name a few. Although subtle, these differences in protocol or experimental apparatus could considerably influence the results. For instance, Clark et al. (unpublished data) examined lateralisation behaviour in juvenile *Acanthochromis polyacanthus* in a double T-maze and found that a slight asymmetry in the barrier position at one end of their maze induced a strong side-bias in their tested fish. This side-bias was not observed at the other end of the arena where the barrier was centrally placed. Similarly, Sundin et al. (unpublished data) tested wild *P. reticulata* in a T-maze arena, which they later discovered had a small crack in one corner, causing this area of the tank to be slightly darker than the other side. Individuals consistently turned in the direction of this darker area, but no consistent side bias was apparent at the other end of the maze, or in the same individuals when tested in other identically constructed arenas. These observations illustrate the care that must go into the construction of experimental apparatus designed to assess side biases in individuals, as slight construction asymmetry or variations in protocol may dramatically influence the results obtained. Such side biases generated through very slight differences in arena construction or lighting may go overlooked, particularly in analyses of absolute lateralisation.

Could differences in species, sex, and/or life stage influence the repeatability of L_R?

Twenty-seven fish species have so far been tested in 31 published studies using a detour test to either relate lateralisation to fitness-relevant traits or assess the effects of environmental stressors on lateralisation strength (ESM Table S1). Several other fish species also feature in studies simply examining whether individuals or populations are lateralised [e.g. 36, 53]. The five species examined here (four original and one re-analysed) are phylogenetically diverse, originating from tropical, temperate, marine, and freshwater habitats. Therefore, they constitute a representative sample allowing our results to be generalised with a reasonable degree of confidence.

We did not include sex in our analyses of repeatability. Some studies suggest that sex should always be considered in studies of cerebral lateralisation because male and female brains are organised differently [54]. A meta-analysis also suggests that there are important sex-differences in the repeatability of many behavioural traits [35]. Indeed, several studies have reported sex-specific differences in lateralisation strength in fishes (e.g. [15, 38, 55-57]), although others have not (e.g. [33, 39, 47, 57]). While this is worth exploring in future studies, it is unlikely that sex differences account for the dramatically low repeatability in lateralisation strength measured in all four of our study species (R=0.006 to R=0.028). With the exception of the protogynous *P. amboinensis*, where our sample is likely to be female-biased based on the size distribution of the collected individuals [58], the remaining species tested have distinct sexes and are not strongly sexually dimorphic. As a result, we assume a roughly even sex-ratio in our samples. If lateralisation in one sex were highly repeatable, we would still expect an R measure considerably higher than observed, with consistency in a turning direction observed in approximately half of the population. This is clearly not observable in our data (Fig. 2).

The effect of life stage on the strength of lateralisation is another consideration that was not explicitly addressed in our study. Although Bell et al. [35] found no difference in the repeatability of behaviours between juveniles and adults in general, they note that, among ectotherms, juvenile behaviour is significantly more repeatable. We included a range of sizes in our tested species, which, for *C. rupestris* and *D. rerio,* included juveniles and sub-adults. We also included fish body mass as a fixed factor in our analyses. We found no effect of mass on the strength of lateralisation in any of our tested species. Published studies of detour tests in fishes have tested a range of life stages from pre-settlement larval fish to adults (ESM Table S1). Although there does not seem to be a consistent trend in lateralisation strength owing to life stage, this should be tested more systematically in future studies.

What explains positive results in previous studies?

The high intra-individual variation in lateralisation we observed across test days in all five species examined (Fig. 2, ESM Fig. S1) suggests that individual L_R scores are random at any point in time. Therefore, why do numerous studies report significant relationships between lateralisation strength and other phenotypic traits or environmental stressors? Unfortunately, many lateralisation studies (including previous publications by authors involved in this study) suffer from low sample sizes (e.g. 20 or fewer individuals per group [8, 14, 25, 28, 30, 36, 39, 57]; ESM Table S3), which considerably increases the likelihood of spurious results [59-61]. In addition, all 35 tests (31 studies) identified in ESM Table S3 employed inadequate statistics to test for the presence of lateralised individuals (see *Supplemental materials and methods* in the ESM). Confirmation bias and poor research practices such as p-hacking and selective reporting also contribute to false positives, which are published more readily than negative results (i.e. the publication bias or file-drawer effect) [62, 63]. A recent survey of over 800 researchers revealed that such practices are rife in ecology and evolution, contributing to the ongoing reproducibility crisis [64]. Improving our confidence in, and ability to replicate, lateralisation studies requires the implementation of validated methodologies, appropriate statistics, high powered designs [61], double-blinded protocols [65], video recordings [66], open data [67], and other transparency measures advocated by the recent Transparency and Openness Promotion (TOP) guidelines [68, 69].

Conclusion

Behavioural lateralisation is likely to be an ecologically important trait that should continue to be tested and measured. However, the method with which to assess eye-use preference as a proxy of cerebral asymmetry must be thoroughly validated for a species of interest before drawing conclusions about the fitness-consequences of changes in lateralisation direction or strength. Drawing inferences from non-repeatable measurements (or studies) not only obfuscates our understanding of a species' evolutionary ecology but also risks misleading policy and conservation efforts. Our study shows that the detour test as it has traditionally been implemented does not provide accurate, precise, or repeatable estimates of behavioural lateralisation in fishes. However, numerous other methods of assessing eye-use and side preference, including mirror tests, swimming, feeding or attack direction preference, and flume tests, have been applied in a range of species in both field and laboratory settings (see [3, 10, 70-76]). Measurements using these methods and their cross-context repeatability should be validated in accordance with TOP guidelines [69] to establish reproducible protocols that inspire confidence.

Animal ethics

Field collections and experiments were approved by the Swedish Board of Agriculture (Dnr 103-2014), Great Barrier Reef Marine Park Authority (G14/36625.1), the James Cook University Animal Ethics Committee in association with AIMS (A2314), and the Norwegian Animal Research Authority (Permit Number:8578).

Data, code, and materials

The data, script, notebook copies, and videos for this study are archived in the repository figshare following best practices [67] and were made available to editors and reviewers upon initial submission: https://doi.org/10.6084/m9.figshare.6881489.v1 [44].

Data re-use

Published data [15] were re-used for this study. The original authors were invited to participate and offered co-authorship.

Competing interests

We have no competing interests.

Author contributions

DGR, MA, JS, TDC, FJ, TN, BSR, and SAB designed the experiments; DGR, MA, RM, JS, AHA, MHF, FJ, MJL, EH, and SAB performed the experiments. JS, FJ and RB contributed vital equipment and funding for the experiments. CB provided the data used in the re-analysis. DGR analysed the data. DGR, MA, and SAB wrote the manuscript with input and approval from all authors.

Acknowledgements

We thank Melvyn Staiff, Antoine Baud, Eleonor Irving, and the staff of the Sven Lovén Centre for Marine Infrastructure, as well as Lizard Island Research Station and the Norwegian University of Science and Technology for assistance with the experiments.

Funding

This work was funded by the Royal Swedish Academy of Sciences (JS: FOA14SLC027; JS, FJ, BSR, DGR, SAB, MA, TDC: FOA17SLC), the Swedish Research Council VR (MA: 637-2014-449), the Swedish Research Council Formas (JS: 2013-947), the Swiss National Science Foundation (DGR, RB, SAB), the Natural Sciences and Engineering Research Council of Canada (BSR, SAB), and the Danish Council for Independent Research (TN: DFF-4181-00297). TDC is the recipient of an Australian Research Council Future Fellowship (project number FT180100154) funded by the Australian Government.

References

1. 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 mollusc. *J. Exp. Biol.* **220**, 667-676.
2. Vallortigara G., Rogers L.J. 2005 Survival with an asymmetrical brain: advantages and disadvantages of cerebral lateralization. *Behav. Brain Sci.* **28**, 575-589. (doi:10.1017/S0140525X05000105)
3. Bibost A.-L., Brown C. 2014 Laterality influences cognitive performance in rainbowfish Melanotaenia duboulayi. *Anim. Cogn.* **17**, 1045-1051.
4. Magat M., Brown C. 2009 Laterality enhances cognition in Australian parrots. *Proc. R. Soc. B* **276**, 4155.
5. Dadda M., Bisazza A. 2006 Lateralized female topminnows can forage and attend to a harassing male simultaneously. *Behav. Ecol.* **17**, 358-363.
6. Rogers L.J., Zucca P., Vallortigara G. 2004 Advantages of having a lateralized brain. *Proc. R. Soc. B* **271**, S420.
7. Sovrano V.A., Dadda M., Bisazza A. 2005 Lateralized fish perform better than nonlateralized fish in spatial reorientation tasks. *Behav. Brain Res.* **163**, 122-127.
8. Ferrari M.C.O., McCormick M.I., Mitchell M.D., Allan B.J.M., Gonçalves E.J., Chivers D.P. 2017 Daily variation in behavioural lateralization is linked to predation stress in a coral reef fish. *Anim. Behav.* **133**, 189-193. (doi:10.1016/j.anbehav.2017.09.020)
9. Bisazza A., Dadda M. 2005 Enhanced schooling performance in lateralized fishes. *Proc. R. Soc. B* **272**, 1677-1681.

10. Kurvers R.H., Krause S., Viblanc P.E., Herbert-Read J.E., Zaslansky P., Domenici P., Marras S., Steffensen J.F., Svendsen M.B., Wilson A.D. 2017 The evolution of lateralization in group hunting sailfish. *Curr. Biol.* **27**, 521-526.

11. Güntürkün O., Diekamp B., Manns M., Nottelmann F., Prior H., Schwarz A., Skiba M. 2000 Asymmetry pays: visual lateralization improves discrimination success in pigeons. *Curr. Biol.* **10**, 1079-1081. (doi:10.1016/S0960-9822(00)00671-0)

12. Dadda M., Koolhaas W.H., Domenici P. 2010 Behavioural asymmetry affects escape performance in a teleost fish. *Biol. Lett.* **6**, 414-417.

13. Lucon-Xiccato T., Chivers D.P., Mitchell M.D., Ferrari M.C. 2016 Prenatal exposure to predation affects predator recognition learning via lateralization plasticity. *Behav. Ecol.*, arw155.

14. Roche D.G., Binning S.A., Strong L.E., Davies J.N., Jennions M.D. 2013 Increased behavioural lateralization in parasitized coral reef fish. *Behav. Ecol. Sociobiol.* **67**, 1339-1344. (doi:10.1007/s00265-013-1562-1)

15. Irving E., Brown C. 2013 Examining the link between personality and laterality in a feral guppy *Poecilia reticulata* population. *J. Fish Biol.* **83**, 311-325.

16. Vallortigara G., Regolin L., Pagni P. 1999 Detour behaviour, imprinting and visual lateralization in the domestic chick. *Cogn. Brain Res.* **7**, 307-320. (doi:10.1016/S0926-6410(98)00033-0)

17. Csermely D., Bonati B., Romani R. 2010 Lateralisation in a detour test in the common wall lizard (*Podarcis muralis*). *Laterality* **15**, 535-547. (doi:10.1080/13576500903051619)

18. Bisazza A., Pignatti R., Vallortigara G. 1997 Detour tests reveal task- and stimulus-specific behavioural lateralization in mosquitofish (*Gambusia holbrooki*). *Behav. Brain Res.* **89**, 237-242. (doi:10.1016/S0166-4328(97)00061-2)

19. Facchin L., Bisazza A., Vallortigara G. 1999 What causes lateralization of detour behavior in fish? Evidence for asymmetries in eye use. *Behav. Brain Res.* **103**, 229-234.

20. Bisazza A., Sovrano V.A., Vallortigara G. 2001 Consistency among different tasks of left–right asymmetries in lines of fish originally selected for opposite direction of lateralization in a detour task. *Neuropsychologia* **39**, 1077-1085. (doi:10.1016/S0028-3932(01)00034-3)

21. Bisazza A., Facchin L., Vallortigara G. 2000 Heritability of lateralization in fish: concordance of right–left asymmetry between parents and offspring. *Neuropsychologia* **38**, 907-912. (doi:10.1016/S0028-3932(00)00018-X)

22. Brown C., Western J., Braithwaite V.A. 2007 The influence of early experience on, and inheritance of, cerebral lateralization. *Anim. Behav.* **74**, 231-238.

23. Bisazza A., Dadda M., Facchin L., Vigo F. 2007 Artificial selection on laterality in the teleost fish *Girardinus falcatus*. *Behav. Brain Res.* **178**, 29-38.

24. Facchin L., Argenton F., Bisazza A. 2009 Lines of *Danio rerio* selected for opposite behavioural lateralization show differences in anatomical left-right asymmetries. *Behav. Brain Res.* **197**, 157-165. (doi:10.1016/j.bbr.2008.08.033)

25. Nilsson G.E., Dixson D.L., Domenici P., McCormick M.I., Sorensen C., Watson S.A., Munday P.L. 2012 Near-future carbon dioxide levels alter fish behaviour by interfering with neurotransmitter function. *Nat. Clim. Change* **2**, 201-204. (doi:10.1038/nclimate1352)

26. Domenici P., Allan B., McCormick M.I., Munday P.L. 2012 Elevated carbon dioxide affects behavioural lateralization in a coral reef fish. *Biol. Lett.* **8**, 78-81. (doi:10.1098/rsbl.2011.0591)

27. Welch M.J., Watson S.-A., Welsh J.Q., McCormick M.I., Munday P.L. 2014 Effects of elevated CO_2 on fish behaviour undiminished by transgenerational acclimation. *Nat. Clim. Change* **4**, 1086.

28. Lopes A., Morais P., Pimentel M., Rosa R., Munday P., Gonçalves E., Faria A. 2016 Behavioural lateralization and shoaling cohesion of fish larvae altered under ocean acidification. *Mar. Biol.* **163**, 243.

29. Maulvault A.L., Santos L.H.M.L.M., Paula J.R., Camacho C., Pissarra V., Fogaça F., Barbosa V., Alves R., Ferreira P.P., Barceló D., et al. 2018 Differential behavioural responses to venlafaxine exposure route, warming and acidification in juvenile fish (*Argyrosomus regius*). *Sci. Total Environ.* **634**, 1136-1147. (doi:10.1016/j.scitotenv.2018.04.015)

30. Jutfelt F., Bresolin de Souza K., Vuylsteke A., Sturve J. 2013 Behavioural disturbances in a temperate fish exposed to sustained high-CO_2 levels. *PLoS ONE* **8**, e65825.

31. Ferrari M.C.O., McCormick M.I., Allan B.J.M., Choi R.B., Ramasamy R.A., Chivers D.P. 2015 The effects of background risk on behavioural lateralization in a coral reef fish. *Funct. Ecol.* **29**, 1553-1559. (doi:10.1111/1365-2435.12483)

32. Chivers D.P., McCormick M.I., Allan B.J., Mitchell M.D., Gonçalves E.J., Bryshun R., Ferrari M.C. 2016 At odds with the group: changes in lateralization and escape performance reveal conformity and conflict in fish schools. *Proc. R. Soc. B* **283**, 20161127.

33. Vossen L.E., Jutfelt F., Cocco A., Thörnqvist P.-O., Winberg S. 2016 Zebrafish (*Danio rerio*) behaviour is largely unaffected by elevated pCO_2. *Conserv. Physiol.* **4**, cow065.

34. Sundin J., Jutfelt F. 2016 9–28 d of exposure to elevated pCO_2 reduces avoidance of predator odour but had no effect on behavioural lateralization or swimming activity in a temperate wrasse (*Ctenolabrus rupestris*). *ICES J. Mar. Sci.* **73**, 620-632.

35. Bell A.M., Hankison S.J., Laskowski K.L. 2009 The repeatability of behaviour: a meta-analysis. *Anim. Behav.* **77**, 771-783.

36. Bisazza A., Cantalupo C., Capocchiano M., Vallortigara G. 2000 Population lateralisation and social behaviour: a study with 16 species of fish. *Laterality* **5**, 269-284.

37. Martin J.G.A., Réale D. 2008 Temperament, risk assessment and habituation to novelty in eastern chipmunks, *Tamias striatus*. *Anim. Behav.* **75**, 309-318. (doi:doi.org/10.1016/j.anbehav.2007.05.026)

38. Bisazza A., Facchin L., Pignatti R., Vallortigara G. 1998 Lateralization of detour behaviour in poeciliid fish: the effect of species, gender and sexual motivation. *Behav. Brain Res.* **91**, 157-164.

39. Sundin J., Jutfelt F. 2018 Effects of elevated carbon dioxide on male and female behavioural lateralisation in a temperate goby. *R. Soc. Open Sci.* **5**, 171550. (doi:10.1098/rsos.171550)

40. Stoffel M.A., Nakagawa S., Schielzeth H. 2017 rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol. Evol.* **8**, 1639-1644. (doi:10.1111/2041-210X.12797)

41. Nakagawa S., Schielzeth H. 2010 Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol. Rev.* **85**, 935-956.

42. Hartig F. 2017 DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models. R package version 0.1.5. https://CRAN.R-project.org/package=DHARMa.

43. R Core Team. 2017 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

44. Roche D.G., Amcoff M., Morgan R., Sundin J., Andreassen A.H., Finnøen M., Lawrence M.J., Henderson E., Norin T., Speers-Roesch B., et al. 2019 Data and script for "Replication alert: behavioural lateralization in a detour test is not repeatable in fishes". https://doi.org/10.6084/m9.figshare.6881489.v1.

45. Voelkl B., Vogt L., Sena E.S., Würbel H. 2018 Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biol.* **16**, e2003693. (doi:10.1371/journal.pbio.2003693)
46. Domenici P., Allan B.J.M., Watson S.-A., McCormick M.I., Munday P.L. 2014 Shifting from right to left: the combined effect of elevated CO_2 and temperature on behavioural lateralization in a coral reef fish. *PLoS ONE* **9**, e87969. (doi:10.1371/journal.pone.0087969)
47. Reddon A.R., Hurd P.L. 2009 Individual differences in cerebral lateralization are associated with shy-bold variation in the convict cichlid. *Anim. Behav.* **77**, 189-193. (doi:10.1016/j.anbehav.2008.09.026)
48. Dadda M., Bisazza A. 2016 Early visual experience influences behavioral lateralization in the guppy. *Anim. Cogn.* **19**, 949-958.
49. Sovrano V.A. 2004 Visual lateralization in response to familiar and unfamiliar stimuli in fish. *Behav. Brain Res.* **152**, 385-391. (doi:https://doi.org/10.1016/j.bbr.2003.10.022)
50. Brown C., Gardner C., Braithwaite V.A. 2004 Population variation in lateralized eye use in the poeciliid *Brachyraphis episcopi*. *Proc. R. Soc. B* **271**, S455-S457.
51. Baltzley M.J., Nabity M.W. 2018 Reanalysis of an oft-cited paper on honeybee magnetoreception reveals random behavior. *J. Exp. Biol.* **221**, jeb185454. (doi:10.1242/jeb.185454)
52. Vila Pouca C., Gervais C., Reed J., Brown C. 2018 Incubation under climate warming affects behavioral lateralisation in Port Jackson sharks. *Symmetry* **10**, 184. (doi:10.3390/sym10060184)
53. Bisazza A., Pignatti R., Vallortigara G. 1997 Laterality in detour behaviour: interspecific variation in poeciliid fish. *Anim. Behav.* **54**, 1273-1281. (doi:10.1006/anbe.1997.0522)
54. Bianki V.L., Filippova E.B. 2001 *Sex Differences in Lateralization in the Animal Brain*. London, CRC Press.
55. Reddon A.R., Hurd P.L. 2008 Aggression, sex and individual differences in cerebral lateralization in a cichlid fish. *Biol. Lett.* **4**, 338-340.
56. Reddon A.R., Hurd P.L. 2009 Sex differences in the cerebral lateralization of a cichlid fish when detouring to view emotionally conditioned stimuli. *Behav. Processes* **82**, 25-29.
57. Byrnes E.E., Pouca C.V., Brown C. 2016 Laterality strength is linked to stress reactivity in Port Jackson sharks (*Heterodontus portusjacksoni*). *Behav. Brain Res.* **305**, 239-246.
58. McCormick M.I. 2016 Protogyny in a tropical damselfish: females queue for future benefit. *PeerJ* **4**, e2198. (doi:10.7717/peerj.2198)
59. Halsey L.G., Curran-Everett D., Vowler S.L., Drummond G.B. 2015 The fickle P value generates irreproducible results. *Nat. Methods* **12**, 179. (doi:10.1038/nmeth.3288)
60. Colquhoun D. 2014 An investigation of the false discovery rate and the misinterpretation of p-values. *R. Soc. Open Sci.* **1**, 140216.
61. Button K.S., Ioannidis J.P., Mokrysz C., Nosek B.A., Flint J., Robinson E.S., Munafò M.R. 2013 Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365.
62. Nuzzo R. 2015 How scientists fool themselves-and how they can stop. *Nature* **526**, 182-185.
63. Parker T.H., Forstmeier W., Koricheva J., Fidler F., Hadfield J.D., Chee Y.E., Kelly C.D., Gurevitch J., Nakagawa S. 2016 Transparency in ecology and evolution: real problems, real solutions. *Trends Ecol. Evol.* **31**, 711-719.
64. Fraser H., Parker T., Nakagawa S., Barnett A., Fidler F. 2018 Questionable research practices in ecology and evolution. *PLOS ONE* **13**, e0200303. (doi:10.1371/journal.pone.0200303)

65. Holman L., Head M.L., Lanfear R., Jennions M.D. 2015 Evidence of experimental bias in the life sciences: why we need blind data recording. *PLoS Biol.* **13**, e1002190. (doi:10.1371/journal.pbio.1002190)
66. Clark T.D. 2017 Science, lies and video-taped experiments. *Nature* **542**, 166.
67. Roche D.G., Kruuk L.E.B., Lanfear R., Binning S.A. 2015 Public data archiving in ecology and evolution: How well are we doing? *PLoS Biol.* **13**, e1002295. (doi:10.1371/journal.pbio.1002295)
68. Clark T.D., Binning S.A., Raby G.D., Speers-Roesch B., Sundin J., Jutfelt F., Roche D.G. 2016 Scientific misconduct: The elephant in the lab. A response to Parker *et al. Trends Ecol. Evol.* **31**, 899-900. (doi:10.1016/j.tree.2016.09.006)
69. Nosek B.A., Alter G., Banks G.C., Borsboom D., Bowman S.D., Breckler S.J., Buck S., Chambers C.D., Chin G., Christensen G., et al. 2015 Promoting an open research culture. *Science* **348**, 1422-1425. (doi:10.1126/science.aab2374)
70. Brown C., Bibost A.-L. 2014 Laterality is linked to personality in the black-lined rainbowfish, *Melanotaenia nigrans*. *Behav. Ecol. Sociobiol.* **68**, 999-1005.
71. Dadda M., Domenichini A., Piffer L., Argenton F., Bisazza A. 2010 Early differences in epithalamic left–right asymmetry influence lateralization and personality of adult zebrafish. *Behav. Brain Res.* **206**, 208-215.
72. Bibost A.-L., Kydd E., Brown C. 2013 The effect of sex and early environment on the lateralization of the rainbowfish *Melanotaenia duboulayi*. In *Behavioral Lateralization in Vertebrates* (eds. Csermely D., Regolin L.), pp. 9-24, Springer.
73. Takeuchi Y., Hori M., Myint O., Kohda M. 2010 Lateral bias of agonistic responses to mirror images and morphological asymmetry in the Siamese fighting fish (*Betta splendens*). *Behav. Brain Res.* **208**, 106-111.
74. Forsatkar M.N., Dadda M., Nematollahi M.A. 2015 Lateralization of aggression during reproduction in male Siamese fighting fish. *Ethology* **121**, 1039-1047.
75. Dadda M., Bisazza A. 2006 Does brain asymmetry allow efficient performance of simultaneous tasks? *Anim. Behav.* **72**, 523-529.
76. Broder E.D., Angeloni L.M. 2014 Predator-induced phenotypic plasticity of laterality. *Anim. Behav.* **98**, 125-130.
77. Gagliano M., Depczynski M. 2013 Spot the difference: mimicry in a coral reef fish. *PLoS ONE* **8**, e55938. (doi:10.1371/journal.pone.0055938)

Table 1. Sample size (*n*), mass range (g), total length range (TL range, cm), and statistics for five species of fish tested to examine the repeatability of behavioural lateralisation. *P. reticulata* were tested with a neutral (^{1}) and a social (^{2}) stimulus (see [15]). Estimates are presented for agreement and adjusted repeatability of relative lateralisation (L_R) with 95% CIs in parentheses. Statistics and *P* values are presented for the effect of trial series (1 to 4), body size (total length for *P. reticulata* and mass for all other species), and start-side of the maze on L_R.

	C. rupestris	***N. azysron***	***P. amboinensis***	***D. rerio***	***P. reticulata***^{1}	***P. reticulata*** ^{2}
n	57	52	60	49	40	40
mass range	1.61-14.9	0.90-6.40	1.0-10.0	0.43-1.07	NA	NA
TL range	4.00-45.24	4.16-8.27	3.77-7.90	2.68-3.58	1.74-3.51	1.74-3.51
R (agreement)	0.028 (0.004-0.052)	0.012 (0.000-0.033)	0.027 (0.008-0.046)	0.006 (0.000-0.023)	0.045 (0.015-0.088)	0.076 (0.028-0.128)
R (adjusted)	0.028 (0.006-0.053)	0.012 (0.000-0.028)	0.022 (0.004-0.038)	0.007 (0.000-0.023)	0.046 (0.011-0.081)	0.077 (0.030-0.126)
trial series	$\chi^2_{(1)} = 0.123$ $P = 0.725$	$\chi^2_{(1)} = 2.379$ $P = 0.123$	$\chi^2_{(1)} = 0.663$ $P = 0.415$	$\chi^2_{(1)} = 3.078$ $P = 0.080$	$\chi^2_{(1)} = 2.757$ $P = 0.097$	$\chi^2_{(1)} = 1.696$ $P = 0.193$
body size	$\chi^2_{(1)} = 0.084$ $P = 0.772$	$\chi^2_{(1)} = 0.890$ $P = 0.345$	$\chi^2_{(1)} = 0.458$ $P =0.499$	$\chi^2_{(1)} = 0.002$ $P = 0.963$	$\chi^2_{(1)} = 3.050$ $P = 0.081$	$\chi^2_{(1)} = 0.942$ $P = 0.332$
start side	$\chi^2_{(1)} = 0.204$ $P = 0.651$	$\chi^2_{(1)} = 0.849$ $P = 0.357$	$\chi^2_{(1)} = 0.354$ $P = 0.552$	$\chi^2_{(1)} = 0.020$ $P = 0.887$	N/A	N/A

Fig. 1 The four fish species from temperate, tropical, marine, and freshwater habitats tested to determine the repeatability of lateralisation: *Ctenolabrus rupestris* (credit: F. Jutfelt), *Neopomacentrus azysron* (credit: picture.world, https://goo.gl/mTLphF), *Pomacentrus amboinensis* [77], and *Danio rerio* (credit: P.H. Olsen, NTNU).

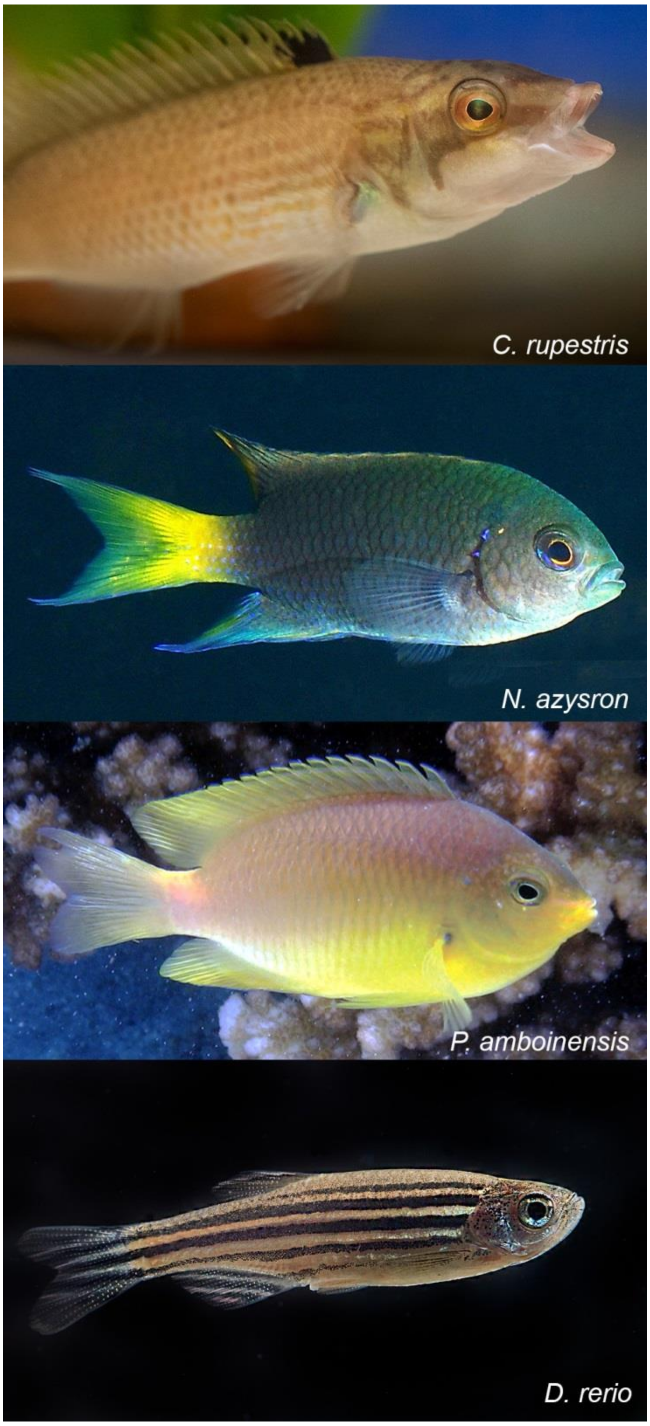
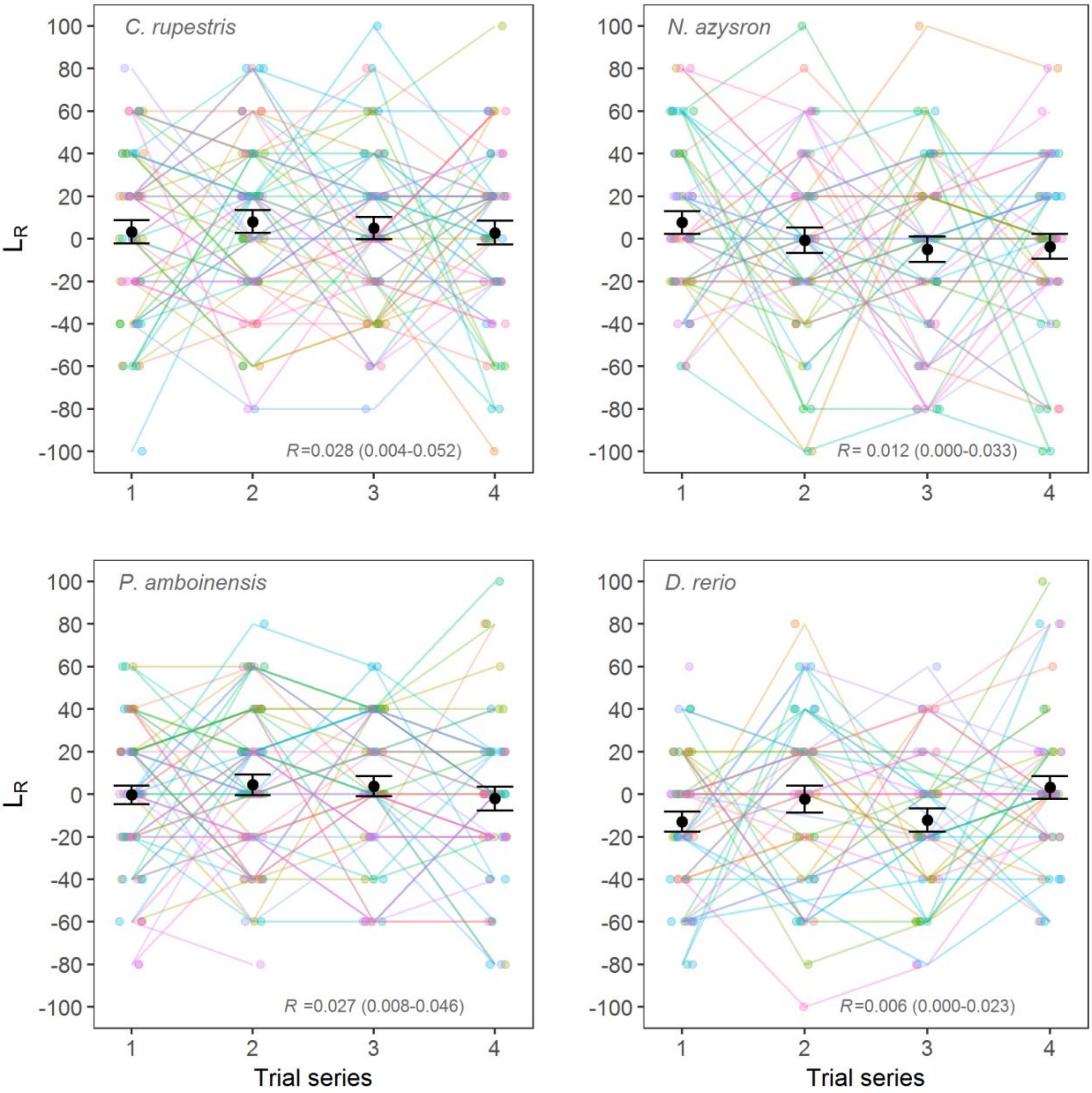


Fig. 2 Relative lateralisation index (L_R) across repeated trials (2 to 4) per individual for four fish species: *Ctenolabrus rupestris* (*n*=57), *Neopomacentrus azysron* (*n*=52), *Pomacentrus amboinensis* (*n*=60), and *Danio rerio* (*n*=49). Individual-level L_R is indicated by the coloured dots connected by lines; the mean (i.e. population-level) L_R and its 95% CI are indicated by black dots and error bars. The repeatability index (*R*) of L_R and its 95% CI are indicated for each species. Four series of 10 trials were conducted. Data points are jittered along the x-axis for presentation purposes.



SUPPLEMENTARY MATERIAL

Replication alert: behavioural lateralisation in a detour test is not repeatable in fishes

Dominique G. Roche[1], Mirjam Amcoff[2], Rachael Morgan[3], Josefin Sundin[3,4], Anna H. Andreassen[3], Mette H. Finnøen[3], Michael J. Lawrence[5], Eleanor Henderson[6], Tommy Norin[7,8], Ben Speers-Roesch[9], Culum Brown[10], Timothy D. Clark[11], Redouan Bshary[1], Brian Leung[12], Fredrik Jutfelt[3], Sandra A. Binning[1,12]

Correspondence to: *dominique.roche@mail.mcgill.ca*

[1] Éco-Éthologie, Institut de Biologie, Université de Neuchâtel, Neuchâtel, Switzerland
[2] Department of Zoology/Functional Zoomorphology, Stockholm University, Stockholm, Sweden
[3] Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway
[4] Department of Neuroscience, Uppsala University, Uppsala, Sweden
[5] Fish Ecology and Conservation Physiology Lab, Carleton University, Ottawa, Ontario, Canada
[6] School of Life Sciences, University of Technology Sydney, Sydney, New South Wales, Australia
[7] Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, Scotland, UK
[8] National Institute of Aquatic Resources, DTU Aqua, Kgs. Lyngby, Denmark
[9] Department of Biological Sciences, University of New Brunswick, Saint John, New Brunswick, Canada
[10] Department of Biological Sciences, Macquarie University, Sydney, Australia
[11] Deakin University, Geelong, Victoria, Australia, School of Life and Environmental Sciences
[12] Department of Biology, McGill University, Montréal, Québec, Canada
[13] Département de sciences biologiques, Université de Montréal, Montréal, Québec, Canada

Supplementary materials and methods

Issues with tests of behavioural lateralisation in the literature

We systematically reviewed the statistical tests employed by studies in Table S1 to determine whether groups of fish exhibit population- and/or individual-level lateralisation. We found that authors often use an arbitrary cut-off or eyeball the distribution of lateralisation scores obtained to determine whether there is evidence of population- or individual-level lateralisation (e.g. see Tables S1 and S3 for a full list of studies). Many studies employ linear models to compare the mean lateralisation scores of control and treatment groups but without testing whether control and treatments groups actually exhibit lateralisation (at the population- or individual-level) (Table S3). This is problematic because populations might have significantly different lateralisation scores without statistically exhibiting lateralisation. Worryingly, the few studies that relied on statistics to explicitly test for lateralisation employed inadequate tests, such as a one-sample t-test or a G-test of goodness-of-fit (this includes previous studies by the authors) [e.g. 1, 2-6] (Table S3). Unfortunately, goodness-of-fit tests such as the G-test result in an inflated type-I error rate when the total sample size is less than 500 and/or an expected number (e.g. the number of right of left turns taken by an individual) is less than five ([7 p. 86-89], page 86-89). Given that, in a detour test, each individual is subjected to only 10 decisions whether to turn left or right, the probability of obtaining an expected number of right or left turns smaller than five is 90%. Therefore, using such tests considerably increases the probability of false positives (see [see 7 p. 87], page 87).

Testing for lateralisation is not straightforward because it involves multiple binomial experiments with structure. Since a binomial trial (a decision to turn left or right) is repeated 10 times per individual in the detour test, trials conducted on multiple individuals are non-independent. This data structure is different from a simpler situation, such as in a standard binomial experiment, such as a coin toss or a trial involving a decision to turn left or right, repeated multiple times by/on a single individual. Here, each trial can be treated as independent.

We have developed and examined the sensitivity of two statistical approaches for testing population- and individual-level lateralisation. A test for detecting lateralisation at the population level requires examining the mean lateralisation score across all individuals in the sample since population-level lateralisation is present when a group of individuals collectively exhibits a side-bias. Specifically, this test involves assessing whether the mean number of turns to the right (or to the left) is significantly different from five, given 10 trials per individual. In contrast, a test for detecting individual-level lateralisation requires examining the sample variance since individual-level lateralisation is present when more individuals exhibit a side-bias than expected by chance (irrespective of whether it is to the left or to the right). For explanations and examples of these two concepts, see Bisazza et al. [8], Domenici et al. [1], and Roche et al. [3].

A robust test to detect population-level lateralisation

Population-level lateralisation can be tested with a generalised linear random-effects model (glmer function in R) that sets the intercept equal to the grand mean of the data. This model can be written as:

```
g <- glmer(X~1+(1|ind), data=dat, family="binomial")
```

where ‘dat’ is a data frame containing three columns: individual ID (‘ind’); trial number (1 to 10); and turning side (column ‘X’) as right (1) or left (0). Specifying the error family as “binomial” sets the probability of success (p) to 0.5. Here, a right turn is considered a success. One *expects* a mean number of 5 turns to the right across all individuals if p = 0.5 (i.e. there is a 50% chance of turning left or right). Testing whether the *observed* mean number of turns to the right differs from the *expected* mean under a binomial error distribution with p = 0.5 simply requires extracting the p-value for the model intercept as such:

```
summary(g)$coefficients[4]
```

If $P < 0.05$, there is statistical evidence that the population is lateralised (i.e. has a side-bias).

A robust test to detect individual-level lateralisation

Individual-level lateralisation can be tested with a chi-square test comparing the observed variance (numerator) to the expected variance (denominator) assuming a normal approximation to the binomial distribution. This is analogous to testing for overdispersion (i.e. are there more observations in the tail ends of the distribution than expected by chance), and is achieved as such:

```
chi_sq=((N-1)*var(X1)/(n*.5*.5))
pchisq(chi_sq,df=(N-1),lower.tail=F)
```

where N is the total number of individuals tested; n is the number of trials per individual (i.e. 10); and X1 is a column in a data frame corresponding to the number of right (or left) turns per individual (each row of the data frame represents one individual).

If $P < 0.05$, there is statistical evidence that some individuals are lateralised (i.e. more individuals have an extreme lateralisation score than expected by chance based on a normal approximation to the binomial distribution with p = 0.5).

Sensitivity analysis

We ran simulations to examine the tests' sensitivity and probability of type I and type II errors. Both tests were robust to a changes in N, n, and p across a wide range of values. The R code to carry out the tests and run the sensitivity analysis are publicly available on figshare: https://figshare.com/s/b8a70d0f083e7741508f.

Alternatives

Note that: 1) individual-level lateralisation can also be assessed with a GLMM rather than a chi-square test but the GLMM approach is less powerful; and 2) population-level lateralisation can also be assessed with a Z-test rather than a GLMM – both have similar power but the Z-test does not account for structure in the data (i.e. it ignores individual ID). We provide these alternatives but recommend the two tests detailed in the paragraphs above.

Table S1. Fish studies documenting links between phenotypic traits or environmental stressors and lateralisation assessed using a detour test. Studies that used a detour test only to assess whether individuals or populations are lateralised are not included. Similarly, studies that used methods other than a detour test to assess lateralisation are excluded. Information is provided on: the type of environmental stressor or phenotypic trait examined (stressor / trait) for each species; life stage; body size (T = treatment group, C = control group); sample size (*n*); environment (env.; FW = fresh water, SW = salt water); whether or not individuals are shoaling/schooling (shoal / school; yes or no); habitat (temperate or tropical); sex (♀ = females, ♂ = males, ♀♂ = both sexes); whether population-level lateralisation was assessed (pop. lat.; R = right biased, L = left biased); the effect of the environmental stressor on the strength of lateralisation or the relationship between lateralisation and a given trait (effect; '+' = increase or positive, '–' = decrease or negative); whether relative (L_R) and/or absolute (L_A) lateralisation were examined (L_R / L_A); the dimensions of the T-maze in cm (dimensions; L = length of the runway, W = width of the runway, B = length of the barrier); reference (ref.).

stressor / trait	**species**	**life stage**	**body size**	***n***	**env.**	**shoal / school**	**habitat**	**sex**	**pop. lat.**	**effect**	**L_R / L_A**	**dimensions**	**ref.**
aggression	*Archocentrus nigrofasciatus*	adult	NA	C=26 T=14	FW	no	tropical	♀♂	no	♀ + right ♂ + left	L_R + L_A	L=75, W=10, B=15	[9]
aquatic acidification	*Acanthochromis polyacanthus*	juvenile	NA	C=105 T=250	SW	yes	tropical	NA	yes (L)	–	L_R + L_A	L=25, W=3, B=12	[10]
	Argyrosomus regius	juvenile	6.8±0.6 cm 2.6±0.8 g	10	SW		temperate	NA	yes (L)	–	L_R + L_A	NA	[11]
	Atherina presbyter	larval	20.48±0.30 mm	C=46 T=60	SW	yes	temperate	NA	yes (L)	–	L_R + L_A	L=50, W=3, B=NA	[12]
	Ctenolabrus rupestris	adult	C: 94.2±8.46 93.7±8.25 mm	C&T= 17-24	SW	no	temperate	NA	no	none	L_R + L_A	L=33, W=9, B=NA	[13]
	Danio rerio	adult	NA	C&T= 7-11	FW	yes	tropical	♀♂	no	+	L_R + L_A	L=33, W=9, B=NA	[14]
	Gadus morhua	juvenile	5.8±0.3 to 13.0±2.4 g	C=21 T=17	SW	yes	temperate	NA	yes (L)	none	L_R + L_A	L=33, W=9, B=NA	[15]
	Gasterosteus aculeatus	adult	~1.6 - 1.9 g	C&T= 20-25	SW	yes	temperate	♀	no	–	L_A	L=33, W=9, B=NA	[16]

	Gasterosteus aculeatus	adult	46.5±2.2 mm 1.3±0.2 g	C=11-12 T=12-13	SW	yes	temperate	NA	no	–	L_A	L=33, W=9, B=NA	[17]
	Gasterosteus aculeatus	adult	C: 46.5±8.9 T: 46.2±6.8 mm	C=14 T=13	SW	yes	temperate	♀♂	yes (R)	–	$L_R + L_A$	NA	[18]
	Gobiusculus flavescens	adult	0.73-0.86 g	C=11-30 T=19-25	SW	yes	temperate	♀♂	yes (R)	none	$L_R + L_A$	L=33, W=9, B=NA	[19]
	Neopomacentrus azysron	larval	~12 mm	C=15-30 T=15-30	SW	yes	tropical	NA	NA	–	L_A	L=25, W=3, B=NA	[5]
	Neopomacentrus azysron	larval	12.1±0.0 mm	C=70 T=68	SW	yes	tropical	♀♂	no	–	$L_R + L_A$	L=25, W=3, B=12	[1]
	Pomacentrus wardi	juvenile	NA	24-58	SW	yes	tropical	♀♂	yes (R)	reversed	$L_R + L_A$	L=25, W=3, B=12	[2]
boldness	*Archocentrus nigrofasciatus*	adult	NA	100	FW	no	tropical	♀♂	no	+	$L_R + L_A$	L=75, W=10, B=15	[6]
	Heterodontus portusjacksoni	juvenile	~30 cm	17	SW	no	tropical	NA	no	none	$L_R + L_A$	NA	[4]
	Poecilia reticulata	adult	♀: 28.4±0.4 ♂: 19.2±0.2 mm	80	FW	yes	tropical	♀♂	yes (R)	none	$L_R + L_A$	L=25, W=10, B=7.5	[20]
cognition	*Cyprinus carpio*	adult	4.6-6.9 g 7.0-7.8 cm	C=8 T=8	FW	no	temperate	NA	no	none	L_R	L=30, W=12, B=30	[21]
	Girardinus falcatus	adult	NA	12-16	FW	yes	tropical	♀	no	+	L_R	NA	[22]
	Girardinus falcatus	adult	NA	6-9	FW	yes	tropical	♀	no	+	L_R	NA	[23]
	Gambusia holbrooki	adult	NA	4-7	FW	yes	subtropical	♀	no	+	L_R	L=40, W=7, B=NA	[24]

cortisol	*Pomacentrus amboinensis*	juvenile	NA	C=25 T=25	SW	no	tropical	NA	no	−	L_R	L=25, W=3, B=12	[25]
growth rate	*Geophagus brasiliensis*	adult	7.0-14.4 cm	26	FW	no	tropical	♀♂	no	+	$L_R + L_A$	L=75, W=10, B=NA	[26]
hypoxia	*Leptocottus armatus*	adult	12.4±0.3 cm (mean±SE)	C=42 T=42	SW	no	temperate	NA	yes (R)	−	$L_R + L_A$	L=75, W=10, B=20	[27]
parasitism	*Scolopsis bilineatus*	adult	T: 11.8±2.5 C: 14.8±1.5 cm (mean±SD)	C=25 T=16	SW	no	tropical	NA	no	+	$L_R + L_A$	L=70, W=15, B=25	[3]
pesticide exposure	*Acanthurus triostegus*	larval	NA[1]	C=10 T=10	SW	yes	tropical	NA	no	−	L_R	NA	[28]
Pharma-ceuticals	*Argyrosomus regius*	juvenile	C: 5.4-6.8 T: 5.4-6.6 cm	C=10 T=10	SW		temperate	NA	yes (L)	−	$L_R + L_A$	NA	[11]
predation	*Brachyrhaphis episcopi*	adult	~2.5 cm	T_1=32 T_2=27	FW	no	tropical	♀♂	no	+	$L_R + L_A$	L=47, W=10, B=NA	[29]
	Poecilia reticulata	adult	NA	C=65 T=68	FW	yes	tropical	♂	no	+	$L_R + L_A$	L=NA, W=8, B=15	[30]
	Caesio teres	juvenile	3.59±0.03 cm (mean±SE)	6-35	SW	yes	tropical	NA	yes (R)	+	$L_R + L_A$	L=25, W=3, B=12	[31]
	Pomacentrus amboinensis	juvenile	NA	T_1=10 T_2=15	SW	no	tropical	NA	no	+	$L_R + L_A$	L=25, W=3, B=12	[25]
	Pomacentrus chrysurus	juvenile	NA[2]	C=30 T=30	SW	no	tropical	NA	NA	+	L_A	L=25, W=3, B=12	[32]
schooling	*Girardinus falcatus*	adult	NA	17-37	FW	yes	tropical	NA	no	+	L_R	NA	[33]
temperature	*Pomacentrus wardi*	juvenile	NA	C=20 T=20	SW	yes	tropical	♀♂	yes (R)	−	$L_R + L_A$	L=25, W=3, B=12	[2]

	Argyrosomus regius	juvenile	C: 5.4-6.8 T: 7.0-8.6 cm	C=10 T=10	SW	temperate	NA	yes (L)	–	$L_R + L_A$	NA	[11]

[1] 0.21-0.25 cm at capture; [2] settlement stage

Table S2. Statistics for tests of individual- and population-level lateralisation for five species of fish: *Ctenolabrus rupestris* (*n*=57), *Neopomacentrus azysron* (*n*=52), *Pomacentrus amboinensis* (*n*=60), *Danio rerio* (*n*=49), and *Poecilia reticulata* (*n*=40). Individuals were tested on four occasions (trials 1-4), except for *P. reticulata*, which were tested three times using both a neutral and a social stimulus (see [20]). A chi-square statistic (ind. χ^2) and *P*-value (ind. *P*) are presented for tests of individual-level lateralisation; *P*<0.05 indicates lateralisation. A z-value (pop. z) and *P*-value (pop. *P*) are presented for tests of population-level lateralisation; *P*<0.05 indicates lateralisation. Statistical significance is indicated in bold (α=0.05).

Species	**trial**	**ind. χ^2**	**ind. *P***	**$\bar{X}$**	**pop. z**	**pop. *P***
C. rupestris	1	**89.42**	**0.002**	5.16	-0.60	0.546
	2	**91.88**	**0.002**	5.40	-1.53	0.126
	3	**81.80**	**0.011**	5.25	-0.99	0.322
	4	**93.14**	**0.001**	5.14	-0.52	0.603
N. azysron	1	**70.31**	**0.025**	4.98	-1.45	0.149
	2	**92.70**	**0.000**	5.22	0.15	0.885
	3	**93.50**	**0.000**	5.19	0.87	0.386
	4	**84.55**	**0.001**	4.90	0.64	0.521
P. amboinensis	1	68.39	0.189	5.38	0.08	0.939
	2	**80.05**	**0.029**	4.96	-0.93	0.354
	3	60.05	0.207	4.75	-0.82	0.412
	4	**73.80**	**0.010**	4.82	0.37	0.712
D. rerio	1	51.59	0.299	**4.35**	**2.71**	**0.007**
	2	**70.57**	**0.004**	4.88	0.39	0.696
	3	49.50	0.144	**4.39**	**2.24**	**0.025**
	4	58.36	0.072	5.16	-0.59	0.557
P. reticulata (neutral stimulus)	1	**78.04**	**0.000**	**3.65**	**-3.72**	**0.000**
	2	49.56	0.120	**4.05**	**-3.35**	**0.001**
	3	38.79	0.479	**4.23**	**-3.09**	**0.002**
P. reticulata (social stimulus)	1	**58.84**	**0.022**	4.65	-1.16	0.245
	2	**74.04**	**0.001**	5.15	0.47	0.642
	3	**78.55**	**0.000**	5.12	0.37	0.709

Table S3. Statistical analyses used by studies listed in Table S1.

[9] Reddon & Hurd 2008 – aggression Tested 40 adults, 20 of each sex. Tested effect of sex and aggressor status on relative and absolute lateralisation (stated as laterality index). Statistical methods are not described, but **R^2 and F-statistic are given**.
[10] Welch et al. 2014 – aquatic acidification Tested juveniles across nine treatment groups. Differences between treatment groups (parents and offspring exposed to control, intermediate and high CO2 in a full factorial design) in absolute lateralisation, was tested using a general **linear mixed-effects model** (**LMM**; LA ~ Parental treatment * Offspring treatment + various random effects). Auhors state that relative lateralisation was used to compare turning bias among the nine treatments using **Pearson's Chi-square test** but no test statistics are presented for L_R in the main text or the supplementary material.
[11] Maulvault et al. 2018 – aquatic acidification, warming, and pharmaceuticals The effects of venlafaxine (in water and in feed), high temperature, and CO2 in various combinations on absolute and relative lateralisation were tested using **generalised linear mixed-effects models (GLMMs)** with a Gaussian error distribution. Tank was specified as random factor (not specified given whether fish from 2 or 3 tanks were used in the test) and the authors employed a model selection approach using AIC.
[12] Lopes et al. 2016 – aquatic acidification Tested the effect of high CO2 exposure for 7 and 21 days on relative and absolute lateralisation with **GLMMs**. Treatment and time were entered as fixed effects and tank as a random effect. Model selection was performed using AIC.
[13] Sundin & Jutfelt 2016 – aquatic acidification Tested control and fish exposed to high CO2, at exposure day 9, 19, and 21 (all fish tested in control water for the last run). Used **generalised linear models (GLMs)** with a binomial error structure and relative and absolute lateralisation as response variables (number of turns to the left over total number of left and right turns (corresponding to relative lateralisation), and maximum number of turns to the preferred side over total number of left and right turns (corresponding to absolute lateralisation), and treatment, test run (for run 1 and 2, run 3 was analysed separately), and the interaction between them as fixed effects.
[14] Vossen et al. 2016 – aquatic acidification Tested eight treatment groups (control CO_2, no gabazine females; control CO_2, no gabazine males; high CO_2, no gabazine females; high CO_2, no gabazine males; control CO_2, gabazine females; control CO_2, gabazine males; high CO_2, gabazine females; and high CO_2, gabazine males). Relative lateralisation (proportion of right turns) analysed using a **GLMM with binomial errors** with CO_2 exposure level, gabazine treatment, and sex as fixed effects, and exposure tank and number of days exposed as random effects.
[15] Jutfelt & Hedgärde 2015 – aquatic acidification Tested juveniles after 29-30 days exposure to control or high CO_2. Used a **nested ANOVA** (tank nested under treatment) on relative and absolute lateralisation to test for differences among treatment groups.
[16] Jutfelt et al. 2013 – aquatic acidification

Tested control and high CO_2 treatments at 20 and 40 days of exposure. Among-treatment effects on relative and absolute lateralisation were tested at two time points using **independent samples t-test with unequal variances**.
[17] Lai et al. 2015 – aquatic acidification Tested control and high CO_2 treatments at 40 and 50 days of exposure (following 30 minutes of gabazine treatment). Absolute lateralisation between control and high-CO_2 fish on day 40 was analysed using a **two-tailed t-test**. A **two-way ANOVA** was used to examine the effects of high CO_2 and gabazine on day 50 followed by Sidak's multiple comparison test as the post-hoc test.
[18] Näslund et al. 2015 – aquatic acidification Tested control and high CO_2 at 20 days of exposure. Absolute lateralisation between the two groups was analysed using a **permutation test using Boss Permutations Tester**.
[19] Sundin & Jutfelt 2018 – aquatic acidification Tested the effect of CO_2 exposure on relative lateralisation across two years. Individuals were tested at 12-19 days of exposure. The effect of CO_2 on relative lateralisation was tested using a **GLMM with binomial errors**. **Wilcoxon-signed rank tests** were used to test against an expected value of 0 to investigate whether the fish were significantly left- or right-biased.
[5] Nilsson et al. 2012 – aquatic acidification Tested the effect of CO_2 and gabazine exposure in a full factorial design. It appears that separate within-treatment tests of absolute lateralisation were performed using **G-tests** (i.e. separate tests were performed for the group treated with control or CO_2 water). Differences before *versus* after gabazine treatment were tested using **t-tests** (also separately for each control/CO_2 treatment).
[1] Domenici et al. 2012 – aquatic acidification Tested the effect of four days of CO_2 exposure, using relative and absolute lateralisation. "In addition, a random simulation (RS) was generated based on 10 random binary choices (i.e. left or right) per individual (n = 70). This simulation was generated in order to test if any of the samples yielded left–right proportions that were not different from that expected by random choice." Within-group preference for left *versus* right turns was tested using **one sample t-tests**, among-group relative and absolute lateralisation was tested using **Kruskal–Wallis tests** (testing all three groups: control, CO_2, and hypothetical random choice group) followed by **Dunn's multiple comparison tests**. The distributions of the control and the elevated-CO_2 individuals were compared with a theoretical binomial distribution using a **G-test**.
[2] Domenici et al. 2014 – aquatic acidification Tested effect of seven days of CO_2 exposure and elevated temperature on fish from four treatments (control CO_2/control temperature, control CO_2/high temperature, high CO_2/control temperature, high CO_2/high temperature), using both relative and absolute lateralisation. Within-group preference for left *versus* right turn (relative lateralisation) was tested using **one sample t-tests**, among-group relative and absolute lateralisation was tested using **two-way ANOVAs** followed by **Tukey HSD tests**.
[6] Reddon & Hurd 2009 – boldness Tested for laterality index (LI) and absolute LI (corresponding to relative and absolute lateralisation). Both were analysed using **one-sample t-tests**.

[4] Byrnes et al. 2016 – boldness Tested for correlation between laterality index and laterality strength (corresponding to relative and absolute lateralisation) and personality. Population-level departures from random choice in laterality direction and overall non-lateralisation were examined using **two-tailed and one-tailed Wilcoxon signed rank tests**, respectively. General **linear models (LMs)** were used to examine the influence of sex, population, time in captivity, total length, boldness, and stress reactivity on lateralisation direction and strength.
[20] Irving & Brown 2013 – boldness Used **repeated measures ANOVA** to examine the effect of stimulus (neutral or control; repeated measure) and sex on relative lateralisation. **Regression** was used to examine relationships between relative (and absolute) lateralisation and boldness, activity, and sociability.
[21] Garina et al. 2016 – cognition Tested fish divided among three experimental groups (serotonin-modulating anticonsolidation protein and two controls). Fish were tested on four different days, five times per day. The median number of right turns out of all five turns was used to assess lateralisation. Changes in turning direction and latency were analysed using **Friedman's test and coefficient of concordance**. The median number of right turns was compared pair-wise between days within each experimental group, as well as on the same day between different experimental groups using **Mann–Whitney U tests**. The median values for each experimental group were calculated across days and compared pair-wise using the same test.
[22] Dadda & Bisazza 2006 – cognition Tested fish from three lines that had been selected for lateralisation (left detour fish, right detour fish, and non-lateralised fish). Data were analysed using **two-way ANOVAs** where laterality type and presence/absence of the male were included as the independent factors, and average time to retrieve food was included as the dependant variable. Lateralisation was not tested directly in the paper; therefore, no statistics on absolute or relative laterality are presented.
[23] Sovrano et al 2005 – cognition Tested fish from lines that had been selected for lateralisation: right detour (RD) fish, left detour (LD) fish, and non-lateralised (NL) fish. Data were analysed using **ANOVAs** firstly to see if laterality affected the proportion of correct choices and, secondly, to see if laterality affected the frequency of choice corner. Laterality type (RD, LD, and NL) was first included as a between-subjects factor, and trial as within-subjects factor. The RD and LD were then grouped into lateralised fish (L) and reanalysed using an **ANOVA.** Lateralisation was not tested directly in the paper; therefore, no statistics on absolute or relative laterality are presented.
[24] Dadda & Bisazza 2006 – cognition Tested fish from lines that had been selected for lateralisation (right detour fish, left detour fish, and non-lateralised fish). Data were analysed using **ANOVAs** (e.g. to test if eye-preference differed between lateralisation groups) and **t-tests** (e.g. to test eye-preference within each lateralisation group). Lateralisation was not tested directly in the paper; therefore, no statistics on absolute or relative laterality are presented.
[25] Ferrari et al. 2017 – cortisol

Tested absolute lateralisation using a **two-way blocked ANOVA**, testing the effect of cortisol (sham *versus* cortisol), and blocking for testing day (random factor).
[26] Reddon et al. 2009 – growth rate Tested fish for absolute and relative lateralisation. Population- and individual-level lateralisation were tested using **one sample t-tests**.
[27] Lucon-Xiccato et al. 2014 – hypoxia Tested relative and absolute lateralisation. Both measurements were analysed after an arcsine transformation. Relative lateralisation was tested within each group using **one sample t-tests**. **Independent sample t-tests** were used to compare mean relative and absolute lateralisation indices between treatment groups.
[3] Roche et al. 2013 – parasitism Tested two groups (unparasitised and parasitised) for relative and absolute lateralisation. Used **goodness-of-fit G-tests** to test within-group relative lateralisation and a **GLM** to test for differences between groups. Between-group absolute lateralisation was tested using a **Mann–Whitney U test** and between parasitised and parasite-removed fish, using a **Wilcoxon paired-sample test**. A **GLM** was used to test turning direction with respect to the side the parasite was attached on parasitised and parasite-removed fish.
[28] Besson et al. 2017 – pesticide exposure Tested eye-use in fish exposed to one of four stimuli, for relative lateralisation. Within-stimulus relative lateralisation was tested using two-tailed **Mann-Whitney U tests**, differences between stimuli were tested using non-parametric tests (**Wilcoxon rank sum test and Kruskall-Wallis test**). Additionally, fish were tested for the effect of lateralisation on predator avoidance, using **Kruskal-Wallis test** (lateralisation was tested and fish were divided into three groups: left tending, right tending, and unbiased – it is not specified which criterion was used). Finally, relative lateralisation was tested for fish divided among two treatments (chlorpyrifos and control) and four different stimuli using **Mann-Whitney U test** as above.
[11] Malvault et al. 2018 – pharmaceuticals See above.
[29] Brown et al. 2007 – predation Tested first generation descendants, from either a high-predation population, or a low-predation population, as well as wild-caught adult females from high-predation or low-predation population, as representatives of the respective wild populations. Fish were tested for relative and absolute lateralisation when sequentially presented with three different stimuli: control, novel object, and unfamiliar same-sex conspecific. **Repeated measures ANOVA** was used to test the response of wild-caught female fish to the three treatments, the difference between the wild-caught females and the female laboratory-reared offspring, and the differences between males and females in the laboratory-reared fish only.
[30] Broder & Angeloni 2014 – predation Tested males from four different populations (high predation populations and low predation populations) exposed to one of two treatments (predator or no predator exposure). Relative and absolute lateralisation were tested using ‘**mixed-effects**’ **ANOVAs**.
[31] Chivers et al. 2016 – predation Tested fish from two treatments (low-risk and high-risk). Relative and absolute lateralisation were tested using a **two-way nested ANOVA** where fish were nested within

tanks, and tanks within risk group. Lateralisation scores of the low- and high-risk fish were also compared to pre-treatment scores using a **one-way ANOVA** followed by post hoc **Tukey tests**.
[25] Ferrari et al. 2017 – predation Tested fish from two treatment groups differing in perceived predation risk pattern for relative and absolute lateralisation using a **two-way repeated measures ANOVA**, investigating the effect of risk peak time (noon *versus* evening) on the absolute lateralisation of fish tested both at noon and in the evening.
[32] Ferrari et al. 2015 – predation Tested fish divided among three treatment groups (no treatment [‘before’ treatment], high-risk environment, and low-risk environment) for absolute lateralisation using a **one-way ANOVA**.
[33] Bisazza & Dadda 2005 – schooling Tested fish originating from lines selected for lateralisation: left detour fish, right detour fish, a mixture of right and left detour fish, and non-lateralised fish. No additional tests on laterality are included in the paper; therefore, no statistics on absolute or relative laterality are presented.
[2] Domenici et al. 2014 – temperature See above.
[11] Malvault et al. 2018 – temperature See above.

Fig. S1 Schematic representation of a double T-maze used to test lateralisation in fishes. A focal fish is acclimated in one side of the arena for one minute behind a barrier (in grey). The barrier is then lifted and the fish is made to swim back and forth in the arena. As it approaches the end of the corridor, the fish faces a barrier and must choose to turn left or right. Decisions to turn left or right are recorded 10 times. Dimensions (mm) of the maze used for *Ctenolabrus rupestris* and *Danio rerio*: 500 (a), 500 (b), 200 (c), 330 (d), 80 (e), 100 (f), and 40 (g). Dimensions (mm) of the maze used for *Neopomacentrus azysron* and *Pomacentrus amboinensis*: 350 (a), 640 (b), 125 (c), 400 (d), 80 (e), 150 (f), and 50 (g).

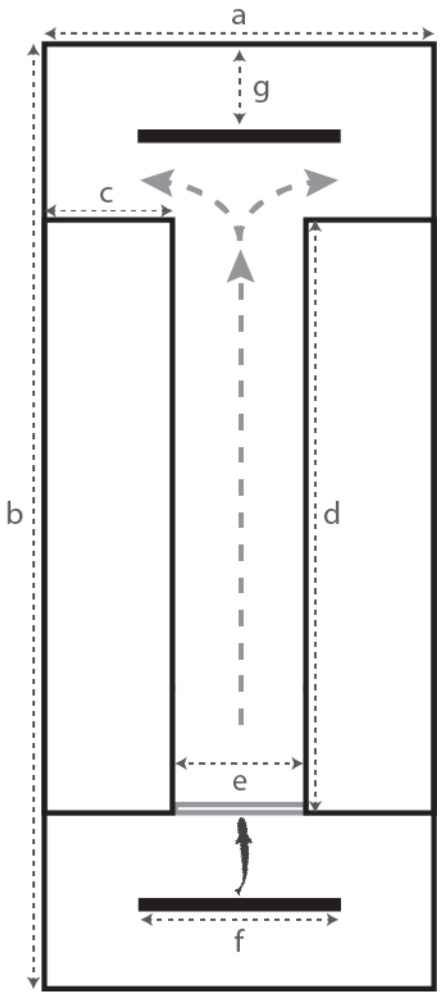


Fig S2. Relative lateralisation index (L_R) across repeated trials per individual for four fish species: *Ctenolabrus rupestris* (*n*=57), *Neopomacentrus azysron* (*n*=52), *Pomacentrus amboinensis* (*n*=60), and *Danio rerio* (*n*=49). Individuals are in ascending order of their mean L_R along the x-axis. The repeatability index (*R*) of L_R and its 95% CI are indicated for each species.

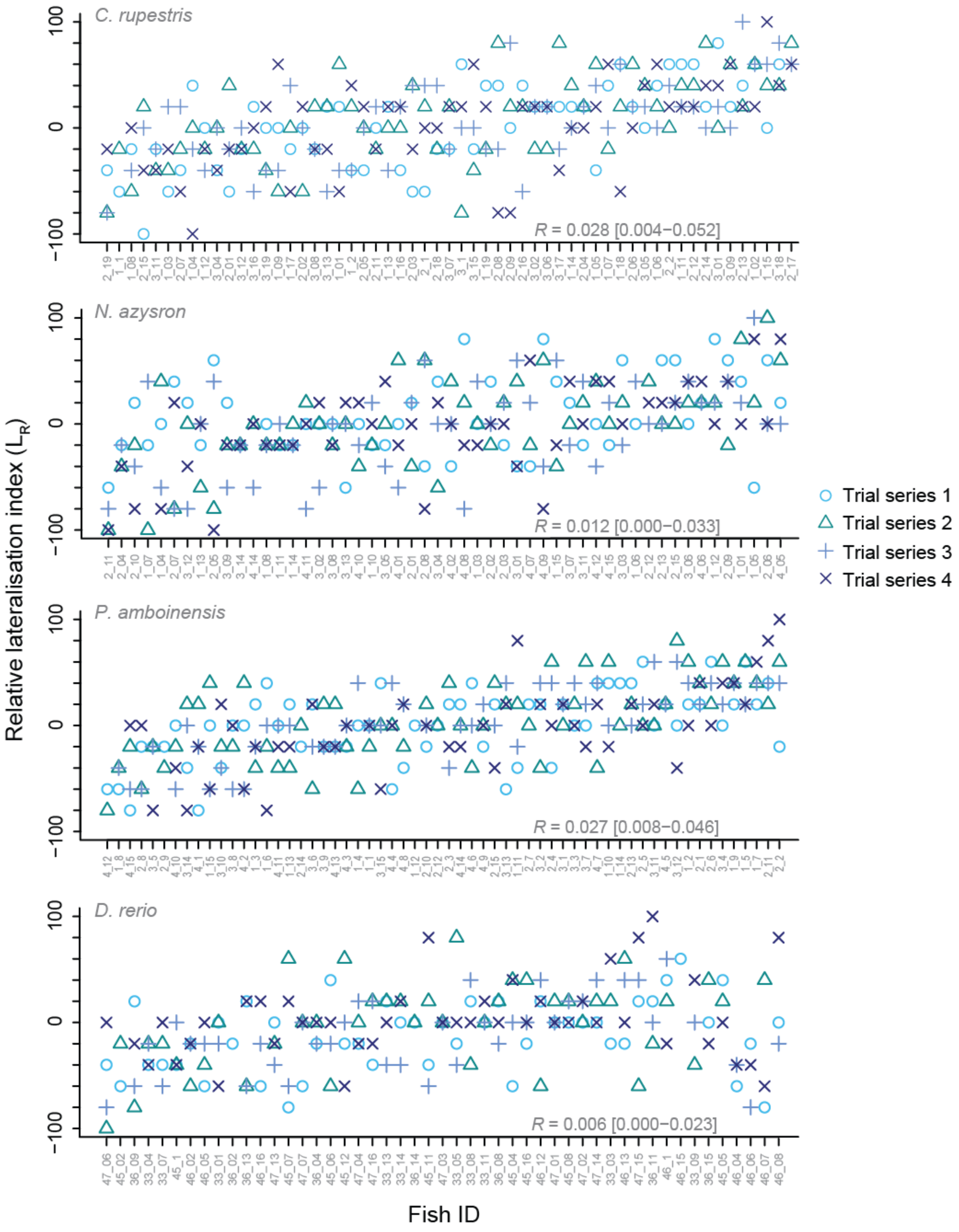


Fig. S3 Absolute lateralisation index (L_A) across repeated trials for four fish species: *Ctenolabrus rupestris* (*n*=57), *Neopomacentrus azysron* (*n*=52), *Pomacentrus amboinensis* (*n*=60), and *Danio rerio* (*n*=49). Individual-level L_A is indicated by the coloured dots connected by lines; the mean (i.e. population-level) L_A and its 95% CI are indicated by black dots and error bars. The repeatability index (*R*) of L_A and its 95% CI are indicated for each species. *R* was computed using the 'rpt' function in the package 'rptR' [34], with a Poisson error distribution. Four series of 10 trials were conducted. Data points are jittered along the x-axis for presentation purposes.

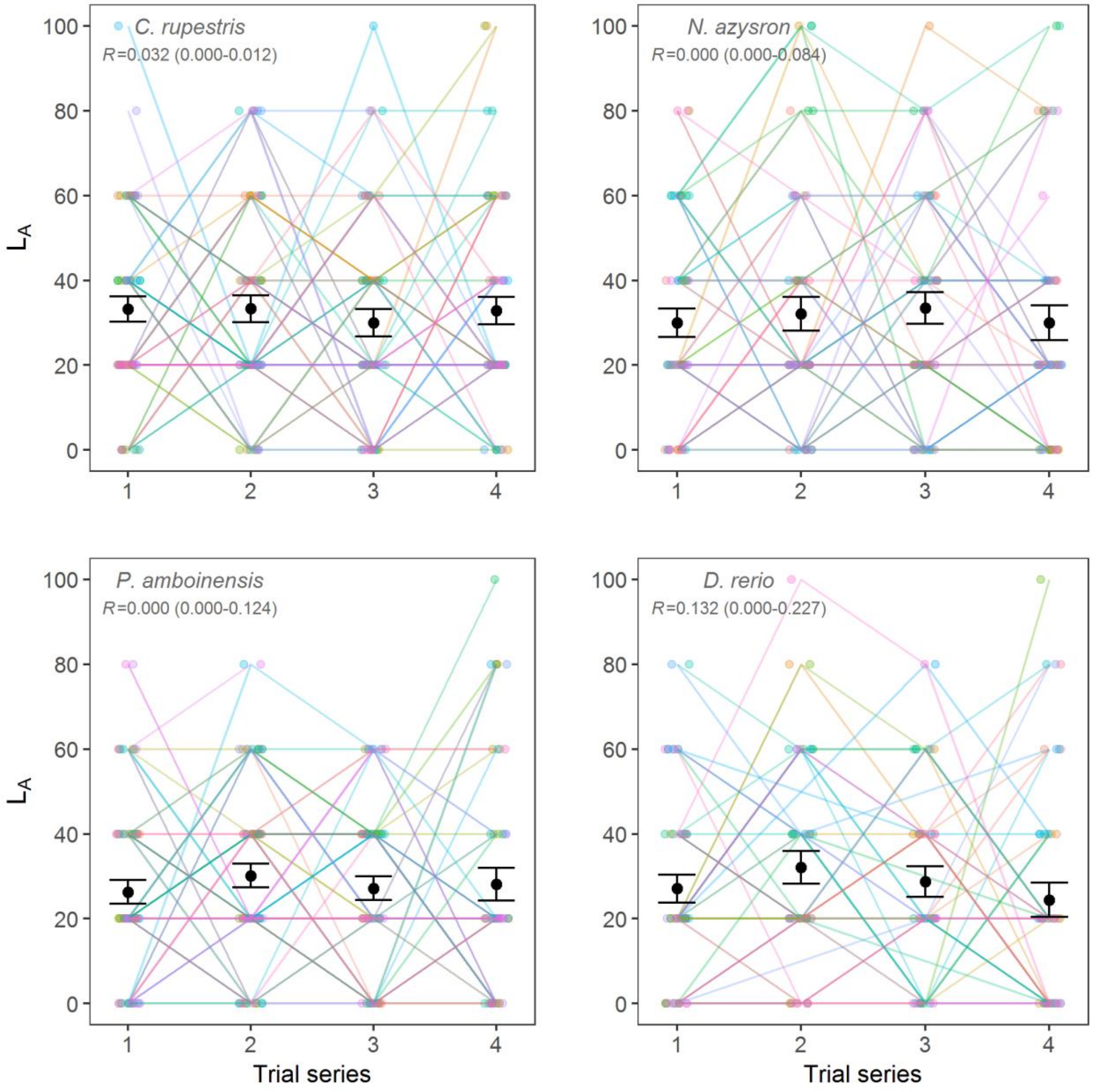


Fig. S4 Relative lateralisation index (L_R) of female guppies (*Poecilia reticulata*) (*n*=40) across three repeated trial series per individual in two different contexts: (**A**) the barrier of the T-maze consists of a neutral stimulus (an empty aquarium behind barrier of vertical bars); (**B**) the barrier of the T-maze consists of a social stimulus (an aquarium containing a conspecific behind a barrier of vertical bars); see Fig. 1 in [17]. Data are from Irving & Brown [20]. Individual-level L_R is indicated by the coloured dots connected by lines; the mean (i.e. population-level) L_R and its 95% CI are indicated by black dots and error bars. The repeatability index (*R*) of L_R and its 95% CI are indicated at the bottom of each panel. Three series of 10 trials were conducted on each individual. Data points are jittered along the x-axis for presentation purposes.

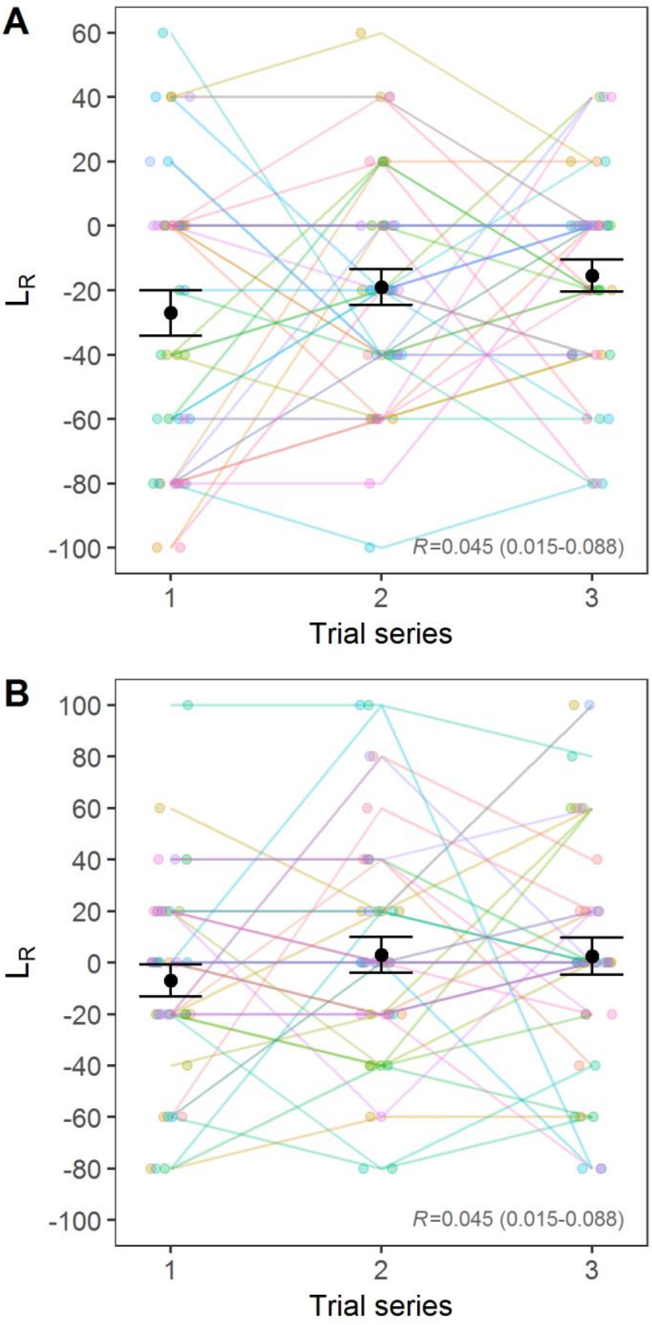


Fig. S5 Frequency distribution of the number of right turns taken by *Ctenolabrus rupestris* (*n*=57) in a detour test (total of 10 turns per individual). The same individuals were tested on four occasions (A, B, C, D) at 48 h intervals. The dashed line represents the mean number of right turns. See Table S2 for statistics.

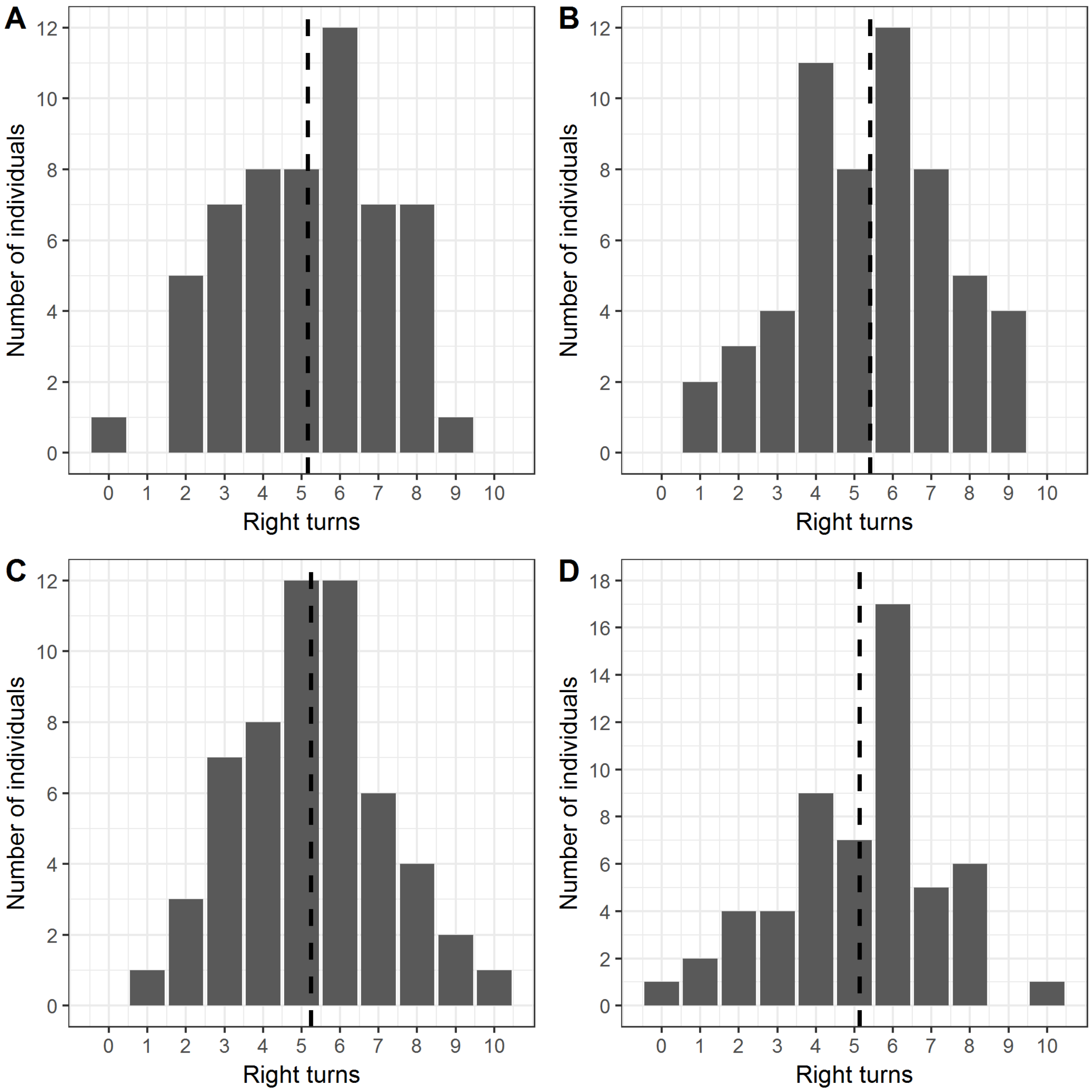


Fig. S6 Frequency distribution of the number of right turns taken by *Neopomacetrus azysron* (*n*=52) in a detour test (total of 10 turns per individual). The same individuals were tested on four occasions (A, B, C, D) at 48 h intervals. The dashed line represents the mean number of right turns. See Table S2 for statistics.

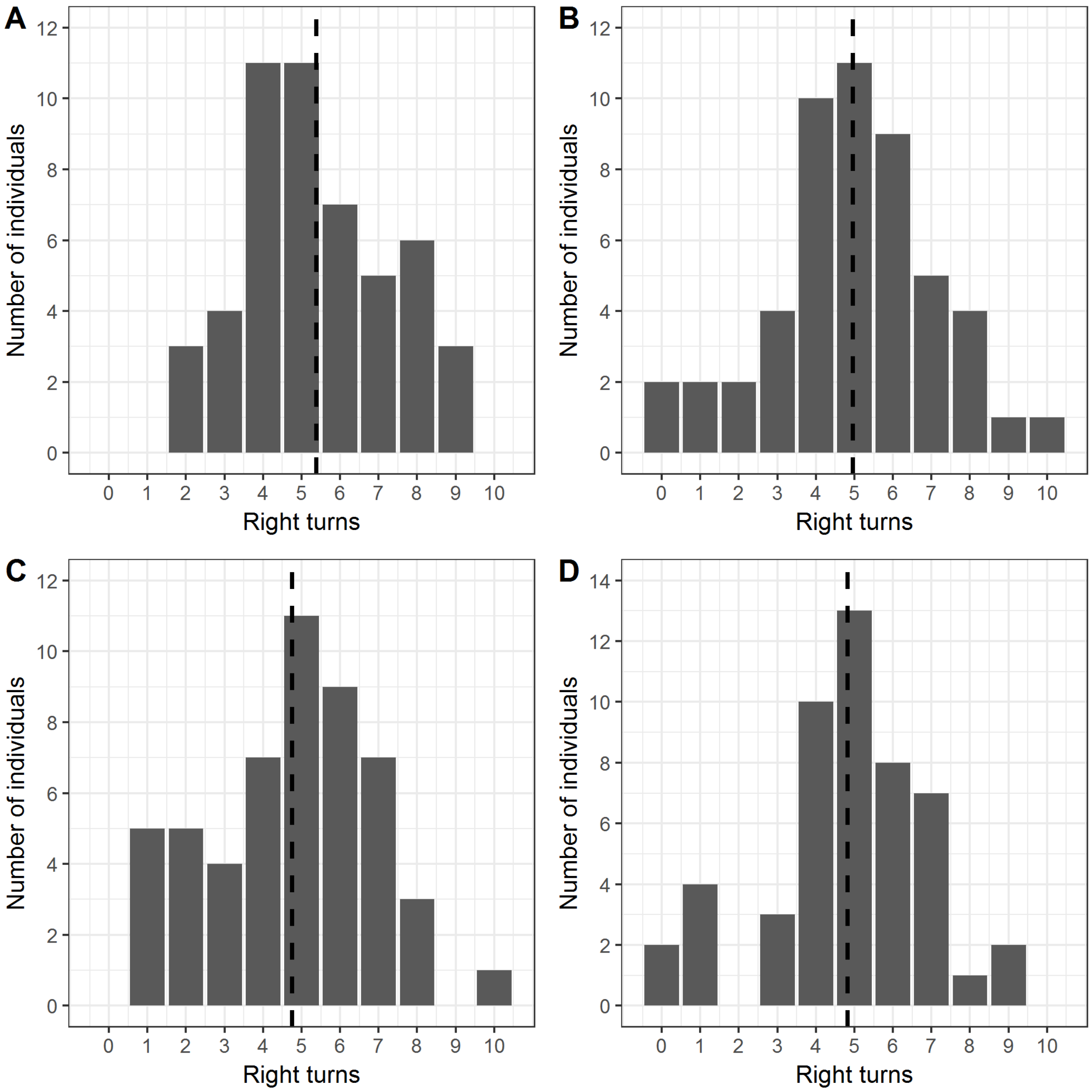


Fig. S7 Frequency distribution of the number of right turns taken by *Pomacentrus amboinensis* (*n*=60) in a detour test (total of 10 turns per individual). The same individuals were tested on four occasions (A, B, C, D) at 48 h intervals. The dashed line represents the mean number of right turns. See Table S2 for statistics.

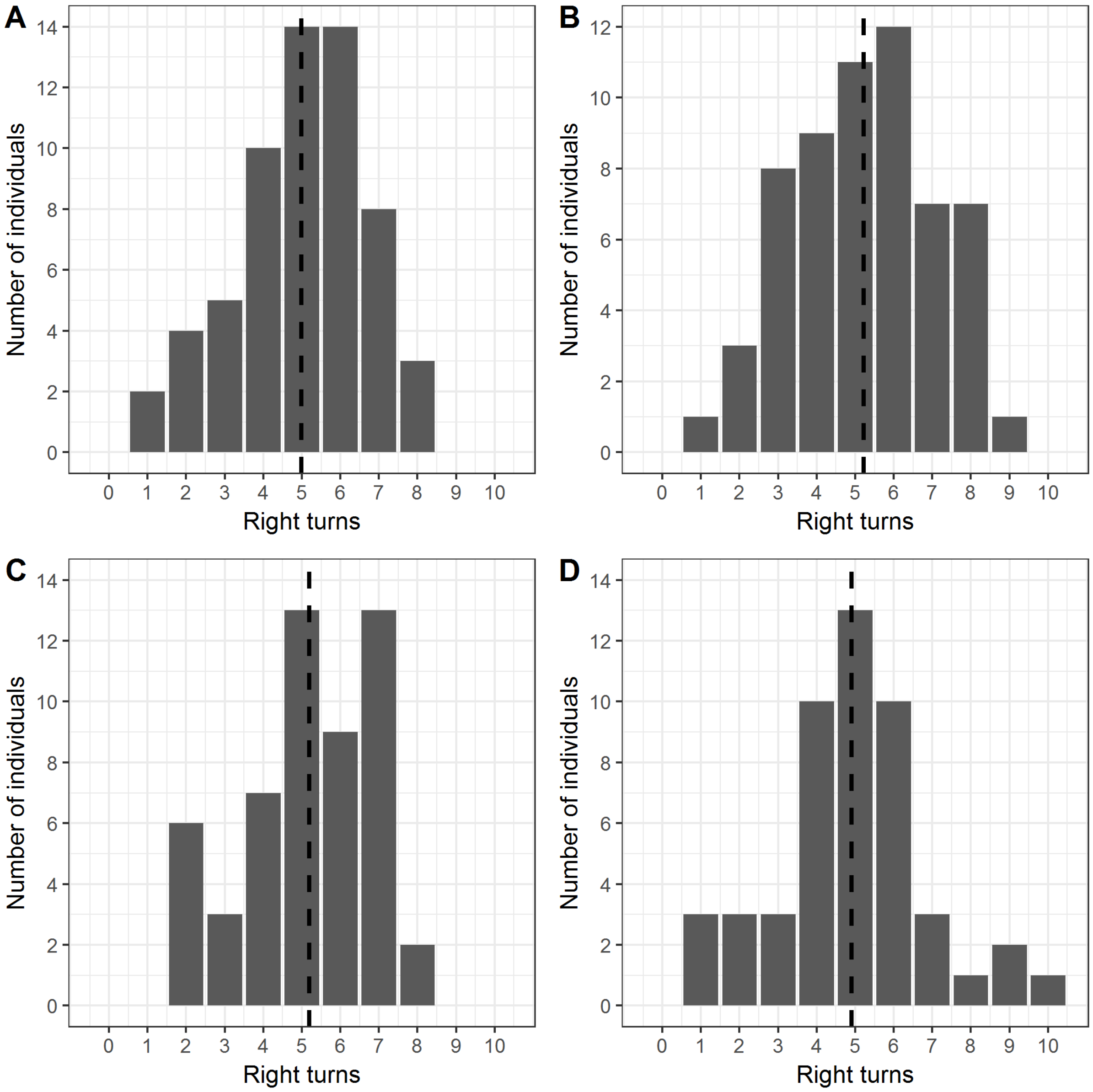


Fig. S8 Frequency distribution of the number of right turns taken by *Danio rerio* (*n*=49) in a detour test (total of 10 turns per individual). The same individuals were tested on four occasions (A, B, C, D) at 48 h intervals. The dashed line represents the mean number of right turns. See Table S2 for statistics.

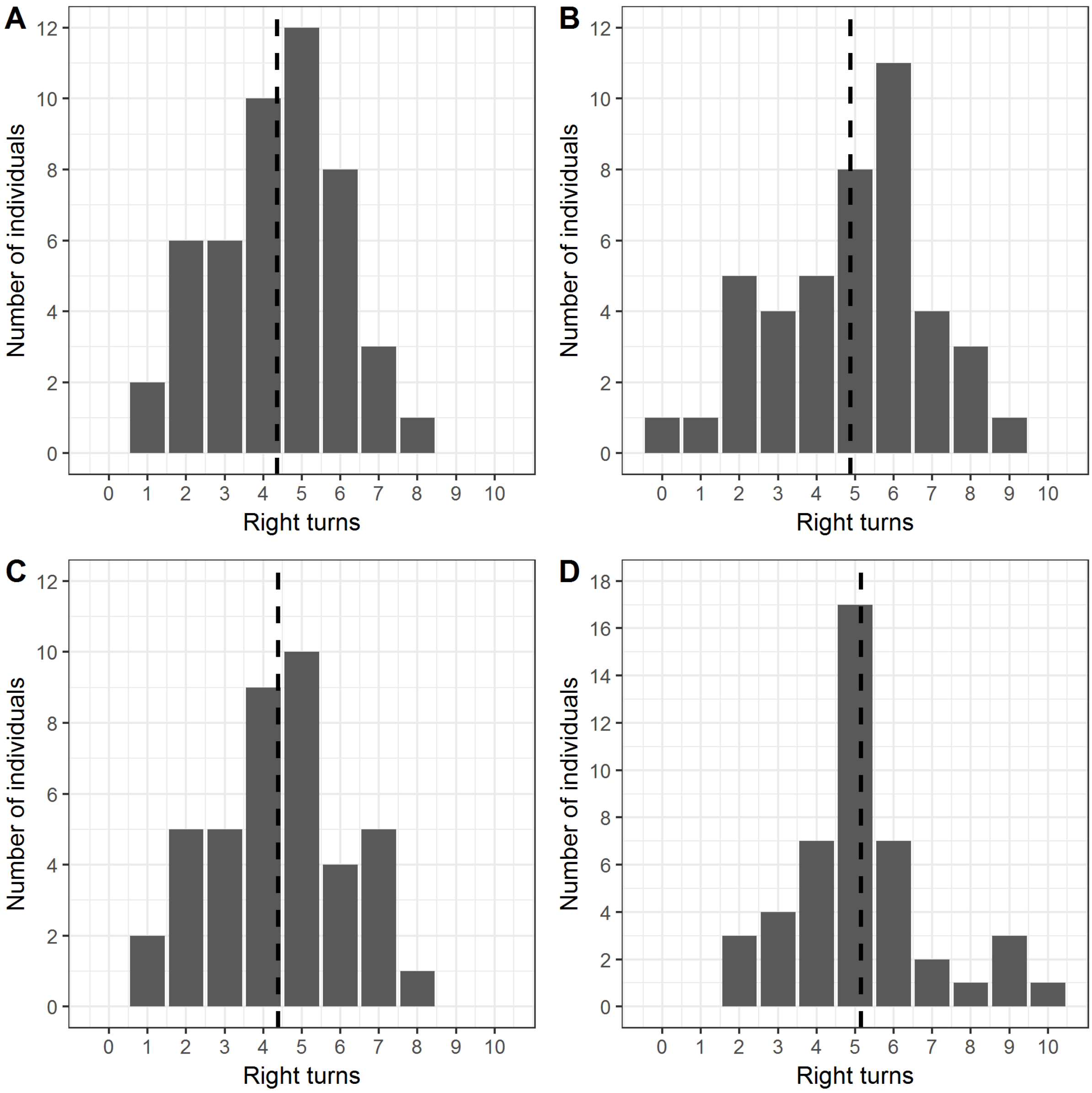
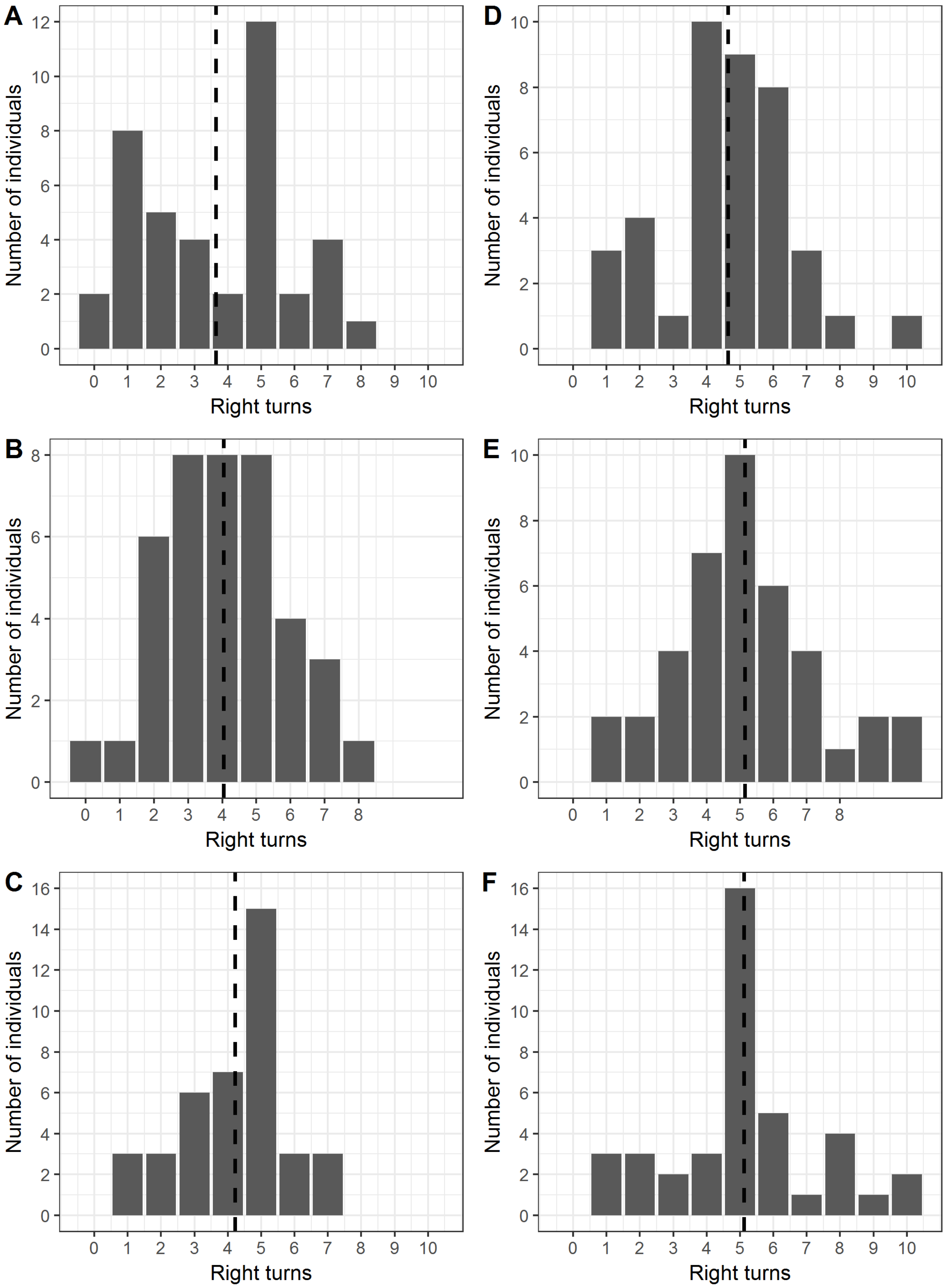


Fig. S9 Frequency distribution of the number of right turns taken by *Poecilia reticulata* (*n*=40) in a detour test (total of 10 turns per individual). The same individuals were tested on three occasions using either a neutral (A, B, C) or a social (D, E, F) stimulus. The dashed line represents the mean number of right turns. See Table S2 for statistics.



References

1. Domenici P., Allan B., McCormick M.I., Munday P.L. 2012 Elevated carbon dioxide affects behavioural lateralization in a coral reef fish. *Biol. Lett.* **8**, 78-81. (doi:10.1098/rsbl.2011.0591)
2. Domenici P., Allan B.J.M., Watson S.-A., McCormick M.I., Munday P.L. 2014 Shifting from right to left: the combined effect of elevated CO_2 and temperature on behavioural lateralization in a coral reef fish. *PLoS ONE* **9**, e87969. (doi:10.1371/journal.pone.0087969)
3. Roche D.G., Binning S.A., Strong L.E., Davies J.N., Jennions M.D. 2013 Increased behavioural lateralization in parasitized coral reef fish. *Behav. Ecol. Sociobiol.* **67**, 1339-1344. (doi:10.1007/s00265-013-1562-1)
4. Byrnes E.E., Pouca C.V., Brown C. 2016 Laterality strength is linked to stress reactivity in Port Jackson sharks (*Heterodontus portusjacksoni*). *Behav. Brain Res.* **305**, 239-246.
5. Nilsson G.E., Dixson D.L., Domenici P., McCormick M.I., Sorensen C., Watson S.A., Munday P.L. 2012 Near-future carbon dioxide levels alter fish behaviour by interfering with neurotransmitter function. *Nat. Clim. Change* **2**, 201-204. (doi:10.1038/nclimate1352)
6. Reddon A.R., Hurd P.L. 2009 Individual differences in cerebral lateralization are associated with shy-bold variation in the convict cichlid. *Anim. Behav.* **77**, 189-193. (doi:10.1016/j.anbehav.2008.09.026)
7. McDonald J.H. 2014 Handbook of Biological Statistics (3rd ed.). *Sparky House Publishing, Baltimore, Maryland. 305 pp.*
8. Bisazza A., Pignatti R., Vallortigara G. 1997 Detour tests reveal task- and stimulus-specific behavioural lateralization in mosquitofish (*Gambusia holbrooki*). *Behav. Brain Res.* **89**, 237-242. (doi:10.1016/S0166-4328(97)00061-2)
9. Reddon A.R., Hurd P.L. 2008 Aggression, sex and individual differences in cerebral lateralization in a cichlid fish. *Biol. Lett.* **4**, 338-340.
10. Welch M.J., Watson S.-A., Welsh J.Q., McCormick M.I., Munday P.L. 2014 Effects of elevated CO_2 on fish behaviour undiminished by transgenerational acclimation. *Nat. Clim. Change* **4**, 1086.
11. Maulvault A.L., Santos L.H.M.L.M., Paula J.R., Camacho C., Pissarra V., Fogaça F., Barbosa V., Alves R., Ferreira P.P., Barceló D., et al. 2018 Differential behavioural responses to venlafaxine exposure route, warming and acidification in juvenile fish (*Argyrosomus regius*). *Sci. Total Environ.* **634**, 1136-1147. (doi:10.1016/j.scitotenv.2018.04.015)
12. Lopes A., Morais P., Pimentel M., Rosa R., Munday P., Gonçalves E., Faria A. 2016 Behavioural lateralization and shoaling cohesion of fish larvae altered under ocean acidification. *Mar. Biol.* **163**, 243.
13. Sundin J., Jutfelt F. 2016 9–28 d of exposure to elevated pCO_2 reduces avoidance of predator odour but had no effect on behavioural lateralization or swimming activity in a temperate wrasse (*Ctenolabrus rupestris*). *ICES J. Mar. Sci.* **73**, 620-632.
14. Vossen L.E., Jutfelt F., Cocco A., Thörnqvist P.-O., Winberg S. 2016 Zebrafish (*Danio rerio*) behaviour is largely unaffected by elevated pCO_2. *Conserv. Physiol.* **4**, cow065.
15. Jutfelt F., Hedgärde M. 2015 Juvenile Atlantic cod behavior appears robust to near-future CO_2 levels. *Front. Zool.* **12**, 11.
16. Jutfelt F., Bresolin de Souza K., Vuylsteke A., Sturve J. 2013 Behavioural disturbances in a temperate fish exposed to sustained high-CO_2 levels. *PLoS ONE* **8**, e65825.
17. Lai F., Jutfelt F., Nilsson G.E. 2015 Altered neurotransmitter function in CO_2-exposed stickleback (*Gasterosteus aculeatus*): a temperate model species for ocean

acidification research. *Conserv. Physiol.* **3**, cov018. (doi:doi.org/10.1093/conphys/cov018)

18. Näslund J., Lindstrom E., Lai F., Jutfelt F. 2015 Behavioural responses to simulated bird attacks in marine three-spined sticklebacks after exposure to high CO_2 levels. *Mar. Freshwat. Res.* **66**, 877-885. (doi:10.1071/mf14144)
19. Sundin J., Jutfelt F. 2018 Effects of elevated carbon dioxide on male and female behavioural lateralisation in a temperate goby. *R. Soc. Open Sci.* **5**, 171550. (doi:10.1098/rsos.171550)
20. Irving E., Brown C. 2013 Examining the link between personality and laterality in a feral guppy *Poecilia reticulata* population. *J. Fish Biol.* **83**, 311-325.
21. Garina D.V., Nepomnyashchikh V.A., Mekhtiev A.A. 2016 Does serotonin-modulating anticonsolidation protein (SMAP) influence the choice of turning direction in carps, *Cyprinus carpio*, in a T-maze? *Fish Physiol. Biochem.* **42**, 1137-1141. (doi:10.1007/s10695-016-0203-9)
22. Dadda M., Bisazza A. 2006 Lateralized female topminnows can forage and attend to a harassing male simultaneously. *Behav. Ecol.* **17**, 358-363.
23. Sovrano V.A., Dadda M., Bisazza A. 2005 Lateralized fish perform better than nonlateralized fish in spatial reorientation tasks. *Behav. Brain Res.* **163**, 122-127.
24. Dadda M., Bisazza A. 2006 Does brain asymmetry allow efficient performance of simultaneous tasks? *Anim. Behav.* **72**, 523-529.
25. Ferrari M.C.O., McCormick M.I., Mitchell M.D., Allan B.J.M., Gonçalves E.J., Chivers D.P. 2017 Daily variation in behavioural lateralization is linked to predation stress in a coral reef fish. *Anim. Behav.* **133**, 189-193. (doi:10.1016/j.anbehav.2017.09.020)
26. Reddon A.R., Gutiérrez-Ibáñez C., Wylie D.R., Hurd P.L. 2009 The relationship between growth, brain asymmetry and behavioural lateralization in a cichlid fish. *Behav. Brain Res.* **201**, 223-228. (doi:10.1016/j.bbr.2009.02.015)
27. Lucon-Xiccato T., Nati J.J., Blasco F.R., Johansen J.L., Steffensen J.F., Domenici P. 2014 Severe hypoxia impairs lateralization in a marine teleost fish. *J. Exp. Biol.* **217**, 4115-4118.
28. Besson M., Gache C., Bertucci F., Brooker R.M., Roux N., Jacob H., Berthe C., Sovrano V.A., Dixson D.L., Lecchini D. 2017 Exposure to agricultural pesticide impairs visual lateralization in a larval coral reef fish. *Sci. Rep.* **7**, 9165.
29. Brown C., Western J., Braithwaite V.A. 2007 The influence of early experience on, and inheritance of, cerebral lateralization. *Anim. Behav.* **74**, 231-238.
30. Broder E.D., Angeloni L.M. 2014 Predator-induced phenotypic plasticity of laterality. *Anim. Behav.* **98**, 125-130.
31. Chivers D.P., McCormick M.I., Allan B.J., Mitchell M.D., Gonçalves E.J., Bryshun R., Ferrari M.C. 2016 At odds with the group: changes in lateralization and escape performance reveal conformity and conflict in fish schools. *Proc. R. Soc. B* **283**, 20161127.
32. Ferrari M.C.O., McCormick M.I., Allan B.J.M., Choi R.B., Ramasamy R.A., Chivers D.P. 2015 The effects of background risk on behavioural lateralization in a coral reef fish. *Funct. Ecol.* **29**, 1553-1559. (doi:10.1111/1365-2435.12483)
33. Bisazza A., Dadda M. 2005 Enhanced schooling performance in lateralized fishes. *Proc. R. Soc. B* **272**, 1677-1681.
34. Stoffel M.A., Nakagawa S., Schielzeth H. 2017 rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol. Evol.* **8**, 1639-1644. (doi:10.1111/2041-210X.12797)

RESPONSE LETTER
Replication alert: behavioural lateralisation in a detour test is not repeatable in fishes

We thank the editor and reviewers for their comments and suggestions. We have revised the manuscript to address them and include a detailed response (in black) to each of the editor and reviewer's points below (indicated in grey). Line numbers in this response letter refer to the revised manuscript, unless stated otherwise.

In addition to these revisions, we have added the following text, tables and figures to the manuscript and Electronic Supplementary Material (ESM):

- **Table 1:** Size, mass, sample size, agreement repeatability, adjusted repeatability and test statistics for the five species examined in our study.
- **Supplementary materials and methods:** an explanation of issues with tests of behavioural lateralisation in the literature and new, appropriate statistical methods.
- **Archived code:** for new statistical analyses and simulations explained in the ESM.
- **Table S1:** three new columns with information on sample size, fish size range and maze dimensions for the 35 rows (i.e. species and stressor or trait tested).
- **Table S2:** statistics for tests of individual- and population-level lateralisation for the five species examined in our study.
- **Table S3:** statistical analyses used by all studies listed in Table S1.
- **Figs S5-S9:** Frequency distribution of the number of right turns taken by the five species examined in our study across the different trial series. Serves as a visual illustration of individual- and population-level lateralisation (tested in Table S2).

Associate Editor

Your manuscript "Replication alert: behavioural lateralization in a detour test is non-repeatable in fishes" has now been reviewed by three experts in the field. As you will see, the reviews are somewhat mixed. All three referees found the topic to be timely and important. One referee, in particular, believes that the work in general has the potentially to be highly influential. Unfortunately, though, the referees also had several major concerns about the research, many of which related to the design of the study. Given these concerns, I unfortunately cannot recommend your manuscript for publication. I am sorry that I don't have better news for you, and I hope that you will find the detailed reviews below useful.

We are glad that all three reviewers found the topic important and timely. We hope that our study will be influential in correcting an important shortcoming in the study of behavioural lateralisation and setting new transparency and reporting standards for the field. Accordingly, Reviewer 1 recognized the considerable effort we put into ensuring that our study design and results are transparent, reproducible, generalizable, and intended to facilitate replication. Below, we comment on the five major concerns raised by Reviewers 2 and 3 and explain how these concerns 1) do not apply to our study, 2) do not affect the validity of our findings, and/or 3) are not unique to our study.

1) One reviewer indicated that the test of lateralisation was flawed due to the fact that runway size was not controlled for.

Reviewer 1's comment only applies to one of the five species used in this study, *C. rupestris*. The size-range of the four other species examined was much narrower (see Table S1). Importantly, our results are consistent across all five species, irrespective of differences in size-range and fish-to-runway ratios: all repeatability estimates are extremely low. If fish size-to-runway ratio had been an issue for *C. rupestris,* one would have expected a low R value for that species and high R values for the other species, which is not the case. Please note that body size was controlled for in the analyses to compute adjusted repeatability (see *Statistical analysis* and Table 1). Moreover, most studies of lateralisation using a detour test do not control for the width/length of the runway in relation to the size of the fish. We have added two columns to Table S1 documenting the size-range of tested fishes and runway widths from previous studies (as well as the four species and two different detour mazes used in our study). As you will see most studies (22 of the 35 rows in Table S1) do not provide information on fish size and/or maze dimensions, which is a shortcoming.

2) Two reviewers expressed concerns that the lateral tagging could affect lateralisation and thus affect your results.

This was an unfortunate misunderstanding based on the wording in our original submission. We did indeed tag fish uniformly on both sides of the body. As Reviewer 2 pointed out, tagging animals on one side of the body could introduce a side bias. The lead author (DGR; English second language) wrote "on either side of the dorsal fin a minimum of two days prior to testing" to mean that tags were placed on both sides. Unfortunately, none of the co-authors recognized that this wording could be misconstrued, but all of the authors who were present and participated in the data collection can attest to this, including PIs (Timothy Clark, Fredrik Jutfelt, Ben Speers-Roesch, Sandra Binning), postdocs (Dominique Roche, Mirjam Amcoff, Tommy Norin, Josefin Sundin) and graduate students (Rachael Morgan, Anna Andreassen, Mette Finnøen, Michael Lawrence, and Eleanor Henderson). This has now been corrected in the revised manuscript (L147-148) as follows: "on both sides of the dorsal fin a minimum of two days prior to the first test".

3) At least one referee was concerned that using a different stimulus for each trial could have affected the potential repeatability of the test.

While we understand this concern, the same could be said about re-using the same stimulus, which would conceivably result in habituation and a loss of lateralisation as the trials progressed. We thought about our choice of stimulus carefully for this reason and opted for neutral stimuli that were slightly modified between trials (a cross, two parallel black bars, a cross with a solid circle above it, a cross with a horizontal bar below it). This information has been added to the manuscript on L159-161. This choice was meant to avoid changing the nature of the stimulus, for example from a neutral one (e.g. a cross) to a potentially threatening one (two dots, which could be perceived as eyes). We also used symmetric stimuli and carefully centered them on the barrier to avoid artificially creating a side-bias. This information was present in the original manuscript.

4) One reviewer in particular was concerned that the populations tested are not lateralized as they were in past papers.

We now show that many of our fish would be considered 'lateralised' and provide new adequate statistical approaches (supported by simulations) to test for lateralisation. This is an interesting comment given that we had not explicitly tested for individual- or population-level lateralisation in the original manuscript. It also speaks to the concern we now express in the revised manuscript (L213-215 and L429-431) and Electronic Supplementary Material (ESM) that authors often eyeball the data to assess whether there is evidence of lateralisation or not. We did not test for lateralisation in the original manuscript because our aim was to test the repeatability of lateralisation scores. However, Reviewer 3 was correct in bringing this to our attention and we have now developed appropriate statistical tests to examine both individual- and population-level lateralisation. Unfortunately, testing for lateralisation is not straightforward. As a result, previous studies (including our own previous studies) have employed inadequate statistical analyses. We now explain this in detail in the ESM (see *Supplementary materials and methods*) and document the specific analyses (or absence of) used by previous studies to test for lateralisation in Table S3. Using correct statistical analyses, we show that 15 of the 22 sample populations (i.e. 68%) exhibit individual-level lateralisation, i.e. more individuals exhibit a side preference than expected by chance based on a binomial distribution with p=0.5 (sample populations displayed in Fig. 2 and Fig. S4; 4 species x 4 trial series each + 1 species x 2 tests x 3 trial series each). Additionally, 5 of the 22 sample populations (i.e. 23%) exhibit population-level lateralisation. These results are now presented in Table S2 and Fig. S5-S9.

5) Some papers suggest that lateralisation can be context dependent, which would explain the variability observed in lateralisation scores.

While we appreciate this suggestion, results suggesting that lateralisation can be context dependent [including our own work (Roche et al., 2013)] are likely to be false positives and/or misinterpretations of real (random) patterns. False positives are common occurrences resulting from low sample sizes, P-hacking and HARKing (see Parker et al., 2016; Fraser et al., 2018). Lateralisation certainly appears to be context-dependent, as our results indicate, but this variability is random rather than adaptive. An example of a similar type of misinterpretation was made explicit in a recent reanalysis of a well-cited study on honeybee magnetoreception. This paper, by a biologist and a mathematician, revealed random patterns rather than adaptive behaviour as suggested by the original authors: Baltzley, M. J., and M. W. Nabity. 2018. Reanalysis of an oft-cited paper on honeybee magnetoreception reveals random behavior. *Journal of Experimental Biology* 221: jeb185454 doi:10.1242/jeb.185454.

To our knowledge, there are only three studies that have documented intra-individual context-dependency in lateralisation and related these observed patterns to adaptive behaviour. One of these is by the lead and senior authors (DGR, SAB) on the current manuscript. We had acknowledged these studies in the original manuscript (see quotes below with line numbers from the original manuscript):

L57-60 "Several studies also report a high degree of within-individual variation in the strength of lateralisation in individuals measured repeatedly across different contexts, suggesting that this trait may be more labile than previously believed [e.g. 8, 14]."

L287-289 "Although Chivers et al. [13] did observe high LR scores in some shoals of the schooling fusilier, *Caesio teres*, lateralisation strength varied substantially both among and within the four groups tested."

L324-331: "Interestingly, Ferrari et al. [8] found considerable variability in the lateralisation scores of *P. amboinensis* tested twice on the same day using opaque barriers with no stimuli. These results were interpreted as representing adaptive plasticity in lateralisation strength due to different predation risk scenarios [8], yet these experiments lacked a control group establishing the consistency of lateralisation strength in untreated individuals. Our results suggest that such variability is likely due to the inherent randomness of turning choice displayed by tested individuals rather than any adaptive behavioural decision."

Referee: 1

This work is exceptionally thorough. The results are based on moderately large samples and experiments conducted by different teams of researchers working in different locations. The researchers have also adhered to the highest standards of transparency. Thus I have unusually strong confidence in the results.

The question addressed by the authors is of great importance to the many researchers seeking to use the detour test as a method of assessing behavioral lateralization in animals (especially fish).

This paper will be highly influential within the realms of behavioral lateralization research, and more broadly with the growing number of researchers interested in transparency and the rigorous evaluation of the validity of published results and methods.

This paper was a pleasure to review. I have no substantial concerns or major suggestions for improvement.

Thank you.

line 160 – how many T-mazes were used at each location?

This information is in the caption of Fig. S1 but we now also mention it explicitly in the main text on L156-157: "*C. rupestris* and *D. rerio* were tested in one maze and *N. azysron* and *P. amboinensis* in another (see Fig. S1)."

line 228 – I was somewhat frustrated by the incomplete reporting of details of results. Most estimates of repeatability are available only in figures, not tables. Availability in figures is better than nothing, but easy access to the actual numbers is desirable. Somewhat more of a problem is that no details are presented regarding the changes in R when accounting for trial series or fish mass, and only p-values and test statistics are presented for identity of experimenter and start-side. Why not include a few tables with this sort of information in the supplement?

Thank you for this suggestion. We have added the suggested table as Table 1 in the main text.

Referee: 2

Understanding the repeatability of studies is very important, so I read this paper with considerable interest. The implication that lateralization scores are not repeatable is important given that they appear to be commonly used, as the authors have pointed out.
I looked at many of the papers cited by the authors to get a sense of this field. I am a bit torn about this paper. While I love the idea of doing repeatability tests, I must say that this paper does appear to have a few biases that give me pause.

We hope that our revisions and response have addressed your concerns.

If you are running a fish down a runway and it has to make a quick decision to turn right or left, then don’t you have to control the width of the runway? Consider a C. rupestris that is 1.38 grams (probably about 2 cm long) and another that is 46.52 grams (length????, line 121) running down a runway that has a certain width. The runway is vastly different. If it is the appropriate size for the large fish, then it is vastly too big for the small fish. The small fish is not coming to a wall in a somewhat confined space and making a decision to turn left or right in rapid time as would be necessary for a real lateralization test. I suspect that not adjusting the tank size to match the fish is a big problem here. I can’t help think that you had a rather poor test of lateralization, hence I am not surprised you have no consistent effect.

Please see our response to the Editor above (*point 1*). In short, this concerns only one of the five species considered in our study and therefore has very little bearing on our results. Body size was also included in our analyses and was not identified as important (see Table 1 for statistics). We now present length and mass ranges in Table 1 (for our study species) and Table S1 (for other studies, many of which lack information on body size). All the data for our study, including fish length, are publicly available on the repository figshare (see the *Data, code, and materials* section in the manuscript) but Table 1 now makes mass and length ranges easier for the reader to assess.

How long does a fish take to recover from having an elastomer tag? I have seen those tags. They are large for a fish that is a couple of cm long. I cannot imagine that you have not dramatically altered the propensity of the fish to turn left or right, especially if it is recovering from the invasive procedure as time goes on over the course of the experiment.

Visible Implant Elastomer (VIE) tagging is a common procedure in fish behaviour/physiology studies, especially when using small individuals that cannot be tagged using conventional methods. The tags on our fish were 1 to 2 mm in diameter at the maximum. Studies on small fishes (e.g. coho salmon smolts, syngnathids, small European eels, age-0 brown trout) similar in size/mass to the ones we studied indicate that VIE tags had no detectable effect on the behaviour (Bailey et al., 1998; Imbert et al., 2007; Caldwell et al., 2011) and growth/survival (Olsen and Vøllestad, 2001) of marked fish versus controls. As stated in the original manuscript, we allowed the animals a minimum of 48 h to recover prior to testing, which is ample time according to previous studies (see Woods and Martin-Smith, 2004). Our fish were feeding prior to the first lateralisation test and between subsequent tests, which is a good indication that the tags (and experimental procedures) had a minimal effect on their behaviour. Please also see our response to Reviewer 3's comment #9.

Reference 13 is incorrect as the paper cited is on frogs not fish. I found Caesio paper to which the authors referred (Chivers et al. 2016. At odds with the group: changes in lateralization and escape performance reveal conformity and conflict in fish schools). It would appear from this paper that lateralization is known to be highly variable through time. The authors link it to predation, but it could also be a result of other stressors. If fish keep getting pulled out of their tank and tested, and the others get chased, then would not the change in stress be critical. The authors seem to ignore all of this, and time of day the fish were tested. Where the fish always put back in the exact same tanks? The Caesio paper suggests that your schoolmates (from different locations or different tanks) determine your lateralization.

Please see our response to the Editor above (*point 5*). Thank you for identifying our incorrect referencing of the tadpole paper. It has been fixed in the revised manuscript. Potential stress from handling and re-testing (in the same and different conditions) is indeed a factor to consider. Yet, there are hundreds of published studies in the behavioural ecology literature that have tested the behaviour of animals repeatedly and shown that some behaviours are highly repeatable (i.e. consistent through time). For details, see a meta-analysis by Bell et al (2009) [>1000 citations] and a methodology paper by Dingemanse & Dochtermann (2013). We have added information on the time of day the fish were tested and where they were housed between test days on L152-154: "Experiments were conducted in the morning between 9:00-12:30 and fish were returned to their respective holding tank between test days.". Please note that repeat sampling of the same individuals is a pre-requisite for assessing the repeatability of a trait.

I think the authors have missed the boat when they suggest that anthropogenic effects cannot be studied using lateralization. If you have a controlled experiment, and everything is identical, except whatever anthropogenic stressor you are manipulating, then the anthropogenic stressor is responsible for any change you see. You may not know the exact mechanism, (i.e. physiological stress, change in glucose level, change in protein expression etc), but that does not mean that you cannot say anything about the stressor. The physiologists in the group of authors are proned to think small scale. Their focus is on mechanisms, however, an ecologist should see the difference between proximate and ultimate causation.

We will refrain from commenting on the last two sentences. If a trait is not repeatable, then it cannot be used to assess anthropogenic (or any other) impacts. There is an abundance of scientific papers documenting the scale of the reproducibility crisis in the life sciences and the prevalence of false positives in the literature. We will not list them all here (key references are cited in the manuscript) but wish to point the reviewer to two recent examples in ecology and evolution that exemplify the problem:

Wang, D., W. Forstmeier, M. Ihle, M. Khadraoui, S. Jerónimo, K. Martin, and B. Kempenaers. 2018. Irreproducible text-book "knowledge": The effects of color bands on zebra finch fitness. *Evolution* 72: 961-976

Baltzley, M. J., and M. W. Nabity. 2018. Reanalysis of an oft-cited paper on honeybee magnetoreception reveals random behavior. *Journal of Experimental Biology* 221: jeb185454

Other papers of interest, which are specific to E&E include:

Lemoine, N. P., A. Hoffman, A. J. Felton, L. Baur, F. Chaves, J. Gray, Q. Yu, and M. D. Smith. 2016. Underappreciated problems of low replication in ecological field studies. *Ecology* 97: 2554-2561.

Nakagawa, S., and T. H. Parker. 2015. Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. *BMC Biology* 13: 88.

Referee: 3

General comments.
This paper aims at testing the repeatability of behavioural lateralization in fish using the detour test. In last couple of decades, the detour test has become the standard way of assessing behavioural lateralization in fishes. Such a test has shown (1) that various traits are related to the lateralization levels, and (2) that various environmental stressors can affect behavioural lateralization. Given the importance of the detour test, an assessment of its repeatability is therefore a potentially important new addition to the literature. The results of this paper suggest that the commonly used detour test is not repeatable and conclude that fish simply turn left or right at random in a detour test. However, this paper has various methodological problems, including the fact that the methods used here do not match those used by most of the previous papers for which the authors are trying to replicate the work. In addition, the interpretation of the results does not appear to take into account all possible explanations, which is a fundamental point in a paper of this kind. For these reasons , the paper is not acceptable.

We respond to these general points in the specific comments below.

Comments
1) line 102: One of the main weak points is that the authors used a detour rest claimed to be similar to the one used in papers on species studied here (8,24-26, 32,33)
However, this paper used a cross or parallel bars at the end of the runway, in the detour test, while previous papers (i.e. 8,24.26,32,33) using the species in question, did not. Specifically:
Papers 8, 25,26 used an opaque barrier
Paper 24 used a mirror test

Papers 32 and 33 used a double T-chamber and no stimulus at the end of the runway.

This is a major weakness of this paper, because the authors did not replicate exactly the methods used in previous work (i.e. the work for which the authors are trying to assess the repeatability). This is a major shortcoming. The issue of replication is a hot one in the field of the behaviour and science in general, and previous commentaries and reviews have pointed out that there are many factors that affect the reproducibility of any given study, including the methodological accuracy of replicating previous work (Maxwell at al Am Psychologist 2015; Fanelli et al PNAS 2018). 4) A major point of reproducibility and repeatability issues is that one can think of it as two tracks. In one case, a paper can measure the repeatability of a measure used in that same paper. Another issue is that of a paper that measures the reproducibility of a result obtained in other papers. Measuring the repeatability of a measurement obtained from another laboratory (or another paper) involves adding the pitfalls of both approaches. In particular, it becomes essential to replicate the exact same conditions as the original measure.

Reviewer 3 is correct that reference 24 used a mirror test rather than a detour test. We apologize for this mistake and have removed this citation from the text. All other references in the text have been thoroughly checked.

We appreciate Reviewer 3's concerns and provide further information on replication below. There are several recent papers explaining what constitutes a replication study. Two main types of replication studies exist: direct replications and conceptual replications, both of which are designed to evaluate the robustness of scientific findings (Kelly, 2006; Nakagawa and Parker, 2015; Zwaan et al., 2018). A direct replication is "a study that attempts to recreate the critical elements (e.g., samples, procedures, and measures) of an original study where those elements are understood according to a theoretical commitment based on the current understanding of the phenomenon under study, reflecting current beliefs about what is needed to produce a finding" (Nosek and Errington, 2017; Zwaan et al., 2018). As such, a direct replication "does not have to duplicate all aspects of an original study. Rather it must only duplicate those elements that are believed necessary for producing the original effect" (Zwaan et al., 2018). A conceptual replication "tests an extension of the theory to a new context"; for example, "whether an effect extends to a different population given theoretical reasons to assume it will be either significantly weaker or stronger in different groups" (Zwaan et al., 2018). Some researchers believe that conceptual replications are more informative than direct replications because they give better evidence for the generalizability of an effect (see Yong, 2012).

In the case of our study, we performed both a direct replication (for the species previously tested in the same way) and a conceptual replication (for species tested using a different variant of the detour test). We note that, given the lack of a standardized methodology and appropriate methodological reporting for the detour test (i.e. different maze dimensions, types of stimuli and barriers, acclimation times and rest time between trials), it is virtually impossible to replicate all published studies in the way Reviewer 3 suggests (see Table S3 for information on maze dimensions). However, we have high confidence in the results and conclusions from our study given its strengths - i.e. multiple species and experimenters, detailed information about the methods (often missing from published studies – see Table S3), filmed trials, open data and analysis script, shared notebooks, re-analysis of independently collected data. This conviction is shared by Reviewer 1, and we hope that future studies on lateralisation will adopt similar standards.

2) The authors claim that the stimulus used is neutral. However, this has not be tested, nor has it been tested whether or not all the stimuli used here were indeed equally neutral. In addition, the stimulus used may elicit a non-lateralized function (exploration, object recognition). As a consequence, the right-left variations may be those due to pure chance. The authors (lines 301-323) discuss the possibility of the effect of different stimulus types. However, the authors seem to treat an opaque barrier in the same way as the stimuli they used, while this is not necessarily the case for the fish.

The purpose of the space behind the barrier of the detour test, the stimulus on the barrier, or the stimulus behind a vertical-bar barrier is to trigger an exploration/inspection behaviour where lateralized fish use a preferred eye to inspect the stimulus or hidden space. This is well explained in many empirical and review-style papers on lateralisation. Behavioural lateralisation is the result of lateralized fishes using a preferred eye to inspect objects or spaces – this preference is believed to vary depending on the stimulus type (e.g. foreign or threatening vs. familiar, conspecific vs. heterospecific, male vs. female). For a detailed explanation, see Bisazza et al. (1997) and Facchin et al. (1999).

Other authors having conducted most of the work on lateralisation in fishes (Bisazza, Vallortigara, Facchin) have also previously used neutral stimuli, such as a red (or a different coloured) ball (see Facchin et al., 1999; Bisazza et al., 2007), which could be considered as a ‘novel object’. Bisazza et al. (2007) – who are authorities in the field – argue that eye preference when inspecting a novel object correlates with eye use during the detour test: “Fish were sorted by their laterality in the detour test using a model predator as stimulus. Later they were singly housed and exposed in their home cage to two novel stimuli, a dummy predator (different from the one used in the detour test) or a neutral stimulus, represented by a red ball, recording eye preference while fixating the novel objects. Fish that tended to detour the barrier on the left side used the right eye to scrutinize the dummy predator and the left eye to scrutinize the neutral stimulus while the fish that tended to detour the barrier on the right side showed the reverse tendency in eye use.”

Facchin , Bisazza and Vallortigara (2009) also state: “This asymmetry [turning bias] has been shown to derive from different specialization of the left- and the right-eye system in processing different classes of stimuli: left-turning fish showed a bias toward right-eye use during fixation of biologically relevant stimuli (such as a potential predator) and a bias toward left-eye use with a new neutral stimuli, while right-turning fish manifested exactly the opposite preferences [23].”

Therefore, there exists strong evidence that our choice of stimulus is warranted. We have added information to the main text detailing the exact nature of neutral stimulus we employed (a cross, two parallel black bars, a cross with a solid circle above it, a cross with a horizontal bar below it). Much like the balls used in other studies, there is no reason to believe that these stimuli would be perceived as either threatening or familiar. We specifically avoided using two circles or solid dots which could have been perceived as eyes, and therefore a threatening stimulus by the fishes.

Note that the use of a dummy predator behind a vertical-bar barrier (Bisazza et al., 1997; Bisazza et al., 1998; Facchin et al., 1999; Bisazza et al., 2000; Bisazza et al., 2007; Dadda and Bisazza, 2016) acts as a threatening stimulus. From experience, it results in the experimenter having to force most of the fish down the runway with a paddle or a dip net since the animals often refrain from swimming towards it.

3) the authors changed the stimulus between trials, in order to prevent habituation. At the same time though, a different stimulus for each trial may have affected the potential repeatability of the detour test.

Please see our response to the Editor above (*point 3*).

4) the authors used a marker on the fish, unlike previous work. .This marker is lateral, and thus it may affect lateralization but this was not tested.

Please see our response to the Editor above (*point 2*) and our response to Reviewer 2 with respect to our symmetrical application of VIE tags.

5) the interpretation of the results and, specifically, the potential explanation for positive results in previous studies (line 396-410) is a fundamental part in a paper of this kind. As it is , however, this part is rather simplistic and weak. The authors suggest that the reasons for positive results in the past can be due to two main issues (1) the low n used in previous work (<20) and (2) P-hacking and selective reporting of significant results .
However there can be many other reasons for the fact that previous work shows a relationship between lateralization and a given trait (or the effect of an environmental factor on lateralization), while the present work suggests that because fish behave randomly in a detour test, previous results should not be expected unless there is low statistical power or publication bias.

a) The stimulus used here is different from those used in previous work.

Please refer to our response on this point above.

b) the stimulus used here was changed in every trial

Please refer to our response on this point above and our response below.

c) After been tested in a detour test once, fish are no longer the same as they were before (they have already seen the tank, the barrier, they have become familiar with a place that was previously unknown, they have been manipulated several times etc.). Indeed, there are many tests in animal behaviour that are not immediately repeatable. For example, the open field test.. What can be done (and has been done in previous work) is to correlate the result of the detour test with other personality or laterality tests that measure the same or different cognitive functions.

As we outline in our response to Reviewer 2's comments, testing animals under the same conditions multiple times is a common and well-established means of assessing the repeatability of physiological and behavioural traits. See Bell et al (2009) for a review and examples of highly repeatable behaviours. We also wish to highlight that behaviours which are highly consistent through time have a high repeatability despite animals becoming familiar with the test arena and being handled on multiple occasions. Clearly, this is not the case for lateralisation assessed with a detour test (i.e. very low repeatability), as our study demonstrates.

We are fully aware of the need to control for habituation (see Roche et al., 2016 for a review paper on repeatability by DGR and SAB), which we have done in our experiment, both statistically – by including trial number as a fixed factor and computing an adjusted repeatability (see Biro and Stamps, 2015) – and by changing the stimulus between series of trials. Adjusted repeatability estimates are now explicitly presented in Table 1. Correlating the results of the detour test with the results of other personality or laterality tests is a useful approach (as we emphasized in the conclusion of our original manuscript) but this method also has shortcomings and would be more informative if lateralisation in a detour tests was repeatable in the first place. Note that there are only two published studies (by the same group) suggesting that lateralisation scores obtained with different methods are sometimes correlated (Sovrano et al., 2001; Dadda et al., 2012). Only one of these two employed a detour test. Other studies have found no correlation between lateralisation measured using different methods: for example, during rotational swimming and escape response (Izvekov et al., 2014), and using a social and non-social stimulus in different apparatuses (Moscicki et al., 2011). Therefore, further evidence in support of this assertion is still needed.

d) In addition, the argument for the possibility that a low n in previous work may have caused biased results is weak. The authors mention 8 papers with n<20 (line 401). However, there are many other papers on lateralization (relationship with other traits, effect of environmental factors) in which the n was much higher than 20, even higher than 50, even among those in the list (table S1). This needs to be acknowledged clearly.

We have added a column with sample sizes for all studies listed in Table S1. We are unaware of additional papers to those listed in Table S1 that test relationships between lateralisation and other traits or the effects of environmental stressors. Unfortunately, Reviewer 3 did not provide references to additional papers. Please see our response to point #8 below, where some papers were referenced. Please also note that we have added the following papers to Table S1 (Bisazza and Dadda, 2005; Sovrano et al., 2005; Dadda and Bisazza, 2006a; Dadda and Bisazza, 2006b), which used individuals from selection lines for lateralisation. These were initially excluded because the fish themselves were not tested in a detour test (although their parents had been).

e) The starting (control) population in many previous studies was lateralized, either at the population or at the individual level. Here, all the populations analysed are not lateralized at either level. This may be a consequence of the fact that the fish behave randomly in the detour test (as the authors suggest). Following this argument, repeatability is not even much an issue, i.e. fish simply behave randomly in a detour test, and therefore all previous work that shows lateralization at the individual or population level was biased or not well executed (e.g. see examples in line 333-353). However, a stronger argument would have been to test the repeatability of a population that was lateralized (at the population or individual level), to see if this lateralization was simply the result of an experimental bias. A counter argument to this idea is that there will never be such a population because fish in a detour test simply turn at random. This would have to been more strongly supported by analysing more than just 5 populations, since there are a lot more than 5 studies that show populations with an individual or a population side bias. In addition, many of these studies show such bias in more than one single paper. (e.g. left bias at population level in Bisazza et al 1997, Bisazza et al 1999, Bisazza et al 1997, Facchin et al 1999, Bisazza et al 2000; Lateralized control population at the individual level, in Nilsson et al 2012, Domenici et al 2012).

Please see our response to the Editor above (*point 4*). In summary, populations do exhibit individual- and population-level lateralisation but these patterns *are not consistent through time*. We note that previous studies mentioned by Reviewer 3 either did not use statistics to explicitly test for lateralisation or used inadequate tests (see Table S3), which is another possible explanation for the positive results in published studies (notwithstanding other issues such as publication bias and those listed in the manuscript). It is also noteworthy that, while several species have previously been tested as pointed out by Reviewer 3, most experiments have been conducted by the same research group, weakening the cumulative evidence they provide (see Voelkl et al., 2018). We do not wish to single out or unduly criticize this group's work but simply highlight that strong evidence for biological patterns requires (multiple) independent replications (Lemoine et al., 2016; Munafò et al., 2017; Royal Netherlands Academy of Arts and Sciences, 2018; Zwaan et al., 2018). In the revised manuscript, we provide extensive information on correct statistical tests for assessing individual- and population-level lateralisation. This is a considerable addition to the original paper and we hope that it will encourage a better use of statistics in this field in the future.

6) the author point out that there does not seem to be a relationship between schooling and lateralization. According to their other view (i.e. lateralization in the detour test is random) , this is not surprising . However, the other point is that indeed some previous work (cited here) has also shown that schooling species do not necessarily show a population level lateralization, in line with the work presented here. Clearly , there is a need for a more comprehensive test (taxonomically correct) that compares many schooling and solitary species in their lateralization tendencies.

To our knowledge, there is only one study that specifically examined and documented a relationship between lateralisation and schooling performance. Bisazza & Dadda (2005) tested a single species (*Girardinus falcatus*): schools were composed of only two individuals (i.e. pairs of fish); fish were from a multi-generation laboratory strain (held since 1992); and only females were tested. Chivers *et al* (2016) also tested one species (wild caught) of schooling fish but examined the effect of lateralisation on individual escape performance rather than any performance measure related to schooling as in Bisazza & Dadda (2005). Similarly, Bibost and Brown (2013) and Middlemiss (2018) examined the effect of lateralisation on an individual's position in a group but not on schooling 'performance'. Finally, Lopes et al. (2016) tested the effect of aquatic acidification on lateralisation and group cohesion (i.e. nearest neighbour distances) but did not include a control to test for the effect of lateralisation on schooling.

Our main objective in the present study was not to test for a relationship between lateralisation and schooling. However, we formulated our prediction given the results and hypotheses put forth in Bisazza & Dadda (2005) (see L34-36): “We expected the three shoaling species to exhibit greater within-individual consistency in lateralisation than their non-shoaling counterparts given previous reports of stronger lateralisation in group-living fishes.” We agree that a broad-scale comparative analysis would be highly informative. Note that Bisazza et al. (2000) compared 16 species, including shoaling and non-shoaling ones, but their sample sizes were low (n per species = 7 to 18) and a formal comparative analysis controlling for phylogeny was not conducted. While our study compares fewer species than Bisazza et al. (2000) (i.e. 5 rather than 16), our sample sizes are larger (n per species = 40 to 60), which increase the robustness of our lateralisation tests and repeatability estimates for the species we tested. As such, our results should not be disregarded. Please note that we were careful in our wording: we suggest that the benefits of lateralisation for schooling performance *might* not be as important as previously thought. In our view, this is a reasonable interpretation of our results and those of previously published studies, as summarised in the main text.

The author however, do not present their point very clearly. They state that the benefits of lateralization for schooling species might be overstated (which would be a correct conclusion if one believed that validity of the detour test in assessing lateralization), and/or that the detour test is not adequate for assessing this trait in group-living species. The authors are not very clear here. It seems that they previously state (line 267) that the detour test is not adequate in fishes in general, not just in group-living species, therefore it is unclear how this paper can make any comment about whether or not schooling species are more lateralized than solitary species.

Our results indicate that the detour test is not an adequate method of assessing for lateralisation in solitary *and* group living fishes. This sentence was originally written as such because Bisazza & Dadda (2005) used a detour test [see p. 1678 in the paper] to assess lateralisation in their study species (*Girardinus falcatus*). Given the issues we highlight with the detour test and the fact that Bisazza & Dadda (2005) is the only study having looked at the relationship between lateralisation and schooling, we feel it is sensible and warranted to question whether or not schooling species are in fact more lateralized than solitary species. For clarity, we have re-phrased the sentence as such (L313-315): “Taken together, these results suggest that that the detour test is not adequate for assessing lateralisation in fishes and that the benefits of lateralisation for shoaling species might be overstated.”

7) The authors reanalysed previous work that showed repeatability and, using a new statistical test, conclude that this work does not support repeatability of lateralization. However, there is also previous work that show heritability of lateralization using the detour test. This heritability cannot be explained when using a detour test in which fish turn at random, as the authors suggest. The authors cite this work (line 77), however they "play it down" by stating that lateralization strength decreases across generations.

The findings that lateralisation is heritable but that its strength decreases across generations were produced by the same research group (please see cited papers in the text). In addition, details about fish size and maze dimensions in studies on these selection lines are incomplete or missing – please see Table S1.

The probability that selection using a detour test in which fish turn at random, would produce a line with same results in successive generations and opposed to those of the line selected in the opposite direction and consisting of two distinct replicate lines of the experiment is infinitely low. In each generation, 100 animals were tested for each replica. Regarding the fact that the scores are reduced after a few generations, articles 23 and 24 give some potential explanations. One is that there can be other selective factors that contrast artificial selection. The fact remains that the detour test can be used to select for lateralization bias, and this cannot be explained by a test in which fish turn at random.

As we explain in the main text, false positives are common in ecology and evolution (and more broadly, in the life sciences). For a recent example, please see the following paper in which a re-analysis of data from a highly cited study showed that a behaviour was random rather than adaptive:

Baltzley, M. J., and M. W. Nabity. 2018. Reanalysis of an oft-cited paper on honeybee magnetoreception reveals random behavior. *Journal of Experimental Biology* 221: jeb185454

Multiple (non-exclusive) factors can explain false positives in the literature: poor statistical practices, selective reporting to support preferred hypotheses, p-hacking, HARKing, confirmation bias and publication bias, to name a few. For a discussion of these issues and their prevalence, see: Ioannidis (2005), Colquhoun (2014), Halsey et al. (2015), Nuzzo (2015), Parker et al. (2016), Forstmeier et al. (2017), and Fraser et al. (2018). For suggestions on approaches to improve the reliability and reproducibility/replication of published studies, please see: Munafò et al. (2017). As mentioned above, it is important to consider that that all studies on the heritability of lateralisation were carried out by the same research group and that information about fish sizes and maze dimensions is often missing, making it difficult to appropriately evaluate these studies. While we do not question the scientific integrity of researchers having conducted this work, we note that these studies are non-independent and do not provide the same strength of evidence as if they had been carried out by different, independent research groups. Confidence in the assertion that lateralisation is heritable would benefit from further replication efforts by other research groups using methods that are as reproducible as possible.

8) There are many other papers that correlate lateralization in the detour test, with other lateral function. (e.g. Dadda et al 2012 Behav Neuroscience, Facchin et al 1999, Behav Brain Res) or with other functions or factors (Chivers et al 2016, Proc Roy B; Broder and Angeloni, 2014 Animal Behaviour; Bisazza et al 2001 Neuropsychologia; Bisazza et al 2005, Behav Brain Research) . These papers found a correlation that would not have been possible if individuals turn at random in the detour test. All this evidence needs to be taken into serious account in the discussion and interpretation of the results of this study.

We have examined these papers and considered them for inclusion in Table S1.

Dadda et al 2012 Behav Neuroscience – This paper compares different methods of measuring lateralisation and does not fit with the general topic of Table S1: how stressors affect lateralisation or how lateralisation is related to other traits.

Facchin et al 1999, Behav Brain Res – This paper is cited in the main text but also does not appear relevant for Table S1. The study examined eye use preference and asymmetry with regards to different stimuli but not in relation to other traits or stressors.

Chivers et al 2016, Proc Roy B – This paper was mis-cited as Lucon-Xiccato et al. 2016. This has now been fixed in the main text and the ESM.

Broder and Angeloni, 2014 Animal Behaviour – This study was not included because the detour test employed in this study is unlike the ones used by other studies listed in Table S1. The arena was circular rather than rectangular and did not include a barrier.

Bisazza et al 2001 Neuropsychologia – This paper also does not seem to fit Table S1. It examined populations selected for laterality and used a detour test to compare between right selected and left selected lines. A detour test (i.e. T-maze) was used to examine whether individuals were still lateralised the same way they had been selected. However, a variant of the 'detour test' (i.e. not a T-maze; see Fig. 2 in the paper), was used to test male eye preference in scrutinising females and eye preference during intrasexual attacks.

Bisazza et al 2005, Behav Brain Research – We were unable to find this paper. Is Reviewer 3 referring to Bisazza and Dadda 2005 Proc Roy Soc B (which is now included in the table – see below)?

Please note that we have added the following papers to Table S1 (Bisazza and Dadda, 2005; Sovrano et al., 2005; Dadda and Bisazza, 2006a; Dadda and Bisazza, 2006b), which used individuals from selection lines for lateralisation. These were initially excluded from the Table because the fish themselves were not tested in a detour test (but their parents had been).

9) Looking online, I found some reports on the repeatability of lateralization in fish:
https://lup.lub.lu.se/student-papers/search/publication/3809909
https://digital.lib.washington.edu/researchworks/handle/1773/26648
These reports can be considered grey literature, and they were not peer reviewed so they need to be taken with caution. Yet, these two reports appear to support repeatability using a detour test.

Irving and Brown (2013) also found that lateralisation was repeatable. However, a re-analysis using proper statistical methods changed this result. The other two studies pointed out by Reviewer 3 have unfortunately not been peer-reviewed and used the same statistical tests as in Irving and Brown (2013) (i.e. a series of Wilcoxon matched paired tests). As we illustrate in the case of Irving and Brown (2013), these tests are likely to result in type II errors (i.e. false positives) and an erroneous conclusion that lateralisation is repeatable. Hence, the conclusion of these two studies should be taken with caution. Importantly, *in both studies fish were only tested twice*, which is considered insufficient to accurately estimate repeatability (see Wolak et al., 2012).

In relation to Reviewer 2's comment about tagging fish, we note that fish in the second study highlighted by Reviewer 3 above were also tagged with Visual Implant Elastomers to allow individual identification. Fish in the first study highlighted by Reviewer 3 were injected with metal PIT-tags, which are larger and considered more invasive than VIE tags. This illustrates that VIE and other tags are commonly used by behavioural ecologists and ecophysiologists to individually mark small fishes.

\---

References

Bailey, R., Irvine, J., Dalziel, F. and Nelson, T. (1998). Evaluations of visible implant fluorescent tags for marking coho salmon smolts. *N. Am. J. Fish. Manage.* **18**, 191-196.

Bell, A.M., Hankison, S.J. and Laskowski, K.L. (2009). The repeatability of behaviour: a meta-analysis. *Anim. Behav.* **77**, 771-783.

Bibost, A.-L. and Brown, C. (2013). Laterality Influences Schooling Position in Rainbowfish, Melanotaenia spp. *PLOS ONE* **8**, e80907.

Biro, P.A. and Stamps, J.A. (2015). Using repeatability to study physiological and behavioural traits: ignore time-related change at your peril. *Anim. Behav.* **105**, 223-230.

Bisazza, A., Cantalupo, C., Capocchiano, M. and Vallortigara, G. (2000). Population lateralisation and social behaviour: a study with 16 species of fish. *Laterality* **5**, 269-284.

Bisazza, A. and Dadda, M. (2005). Enhanced schooling performance in lateralized fishes. *Proc. R. Soc. B* **272**, 1677-1681.

Bisazza, A., Dadda, M., Facchin, L. and Vigo, F. (2007). Artificial selection on laterality in the teleost fish *Girardinus falcatus*. *Behav. Brain Res.* **178**, 29-38.

Bisazza, A., Facchin, L., Pignatti, R. and Vallortigara, G. (1998). Lateralization of detour behaviour in poeciliid fish: the effect of species, gender and sexual motivation. *Behav. Brain Res.* **91**, 157-164.

Bisazza, A., Pignatti, R. and Vallortigara, G. (1997). Detour tests reveal task- and stimulus-specific behavioural lateralization in mosquitofish (*Gambusia holbrooki*). *Behav. Brain Res.* **89**, 237-242.

Caldwell, I., Correia, M., Palma, J. and Vincent, A. (2011). Advances in tagging syngnathids, with the effects of dummy tags on behaviour of Hippocampus guttulatus. *J. Fish Biol.* **78**, 1769-1785.

Chivers, D.P., McCormick, M.I., Allan, B.J., Mitchell, M.D., Gonçalves, E.J., Bryshun, R. and Ferrari, M.C. (2016). At odds with the group: changes in lateralization and escape performance reveal conformity and conflict in fish schools. *Proc. R. Soc. B* **283**, 20161127.

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *R. Soc. Open Sci.* **1**, 140216.

Dadda, M. and Bisazza, A. (2006a). Does brain asymmetry allow efficient performance of simultaneous tasks? *Anim. Behav.* **72**, 523-529.

Dadda, M. and Bisazza, A. (2006b). Lateralized female topminnows can forage and attend to a harassing male simultaneously. *Behav. Ecol.* **17**, 358-363.

Dadda, M. and Bisazza, A. (2016). Early visual experience influences behavioral lateralization in the guppy. *Anim. Cogn.* **19**, 949-958.

Dadda, M., Nepomnyashchikh, V.A., Izvekov, E.I. and Bisazza, A. (2012). Individual-level consistency of different laterality measures in the goldbelly topminnow. *Behav. Neurosci.* **126**, 845.

Dingemanse, N.J. and Dochtermann, N.A. (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *J. Anim. Ecol.* **82**, 39-54.

Facchin, L., Argenton, F. and Bisazza, A. (2009). Lines of *Danio rerio* selected for opposite behavioural lateralization show differences in anatomical left-right asymmetries. *Behav. Brain Res.* **197**, 157-165.

Facchin, L., Bisazza, A. and Vallortigara, G. (1999). What causes lateralization of detour behavior in fish? Evidence for asymmetries in eye use. *Behav. Brain Res.* **103**, 229-234.

Forstmeier, W., Wagenmakers, E.J. and Parker, T.H. (2017). Detecting and avoiding likely false-positive findings–a practical guide. *Biol. Rev.* **92**, 1941-1968.

Fraser, H., Parker, T., Nakagawa, S., Barnett, A. and Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLOS ONE* **13**, e0200303.

Halsey, L.G., Curran-Everett, D., Vowler, S.L. and Drummond, G.B. (2015). The fickle P value generates irreproducible results. *Nat. Meth.* **12**, 179-185.

Imbert, H., Beaulaton, L., Rigaud, C. and Elie, P. (2007). Evaluation of visible implant elastomer as a method for tagging small European eels. *J. Fish Biol.* **71**, 1546-1554.

Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLOS Medicine* **2**, e124.

Irving, E. and Brown, C. (2013). Examining the link between personality and laterality in a feral guppy *Poecilia reticulata* population. *J. Fish Biol.* **83**, 311-325.

Izvekov, E.I., Kuternitskaya, E.A., Pankova, N.A., Malashichev, Y.B. and Nepomnyashchikh, V.A. (2014). Lateralisation of rotational swimming but not fast escape response in the juvenile sterlet sturgeon, Acipenser ruthenus (Chondrostei: Acipenseridae). *Laterality: Asymmetries of Body, Brain and Cognition* **19**, 302-324.

Kelly, C.D. (2006). Replicating empirical research in behavioral ecology: how and why it should be done but rarely ever is. *Q. Rev. Biol.* **81**, 221-236.

Lemoine, N.P., Hoffman, A., Felton, A.J., Baur, L., Chaves, F., Gray, J., Yu, Q. and Smith, M.D. (2016). Underappreciated problems of low replication in ecological field studies. *Ecology* **97**, 2554-2561.

Lopes, A.F., Morais, P., Pimentel, M., Rosa, R., Munday, P.L., Gonçalves, E.J. and Faria, A.M. (2016). Behavioural lateralization and shoaling cohesion of fish larvae altered under ocean acidification. *Mar. Biol.* **163**, 243.

Middlemiss, K.L., Cook, D.G., Jaksons, P., Jerrett, A.R. and Davison, W. (2018). Lateralisation of visual function in yellow-eyed mullet (Aldrichetta forsteri) and its role in schooling behaviour. *Mar. Freshwat. Behav. Physiol.* **51**, 15-29.

Moscicki, M.K., Reddon, A.R. and Hurd, P.L. (2011). Lateralized behaviour of a non-social cichlid fish (Amatitlania nigrofasciata) in a social and a non-social environment. *Behav. Processes* **88**, 27-32.

Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., du Sert, N.P., Simonsohn, U., Wagenmakers, E.-J., Ware, J.J. and Ioannidis, J.P. (2017). A manifesto for reproducible science. *Nature Human Behaviour* **1**, 0021.

Nakagawa, S. and Parker, T.H. (2015). Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. *BMC biology* **13**, 88.

Nosek, B.A. and Errington, T.M. (2017). Making sense of replications. *eLife* **6**, e23383.

Nuzzo, R. (2015). How scientists fool themselves-and how they can stop. *Nature* **526**, 182-185.

Olsen, E.M. and Vøllestad, L.A. (2001). An evaluation of visible implant elastomer for marking age-0 brown trout. *N. Am. J. Fish. Manage.* **21**, 967-970.

Parker, T.H., Forstmeier, W., Koricheva, J., Fidler, F., Hadfield, J.D., Chee, Y.E., Kelly, C.D., Gurevitch, J. and Nakagawa, S. (2016). Transparency in ecology and evolution: real problems, real solutions. *Trends Ecol. Evol.* **31**, 711-719.

Roche, D.G., Binning, S.A., Strong, L.E., Davies, J.N. and Jennions, M.D. (2013). Increased behavioural lateralization in parasitized coral reef fish. *Behav. Ecol. Sociobiol.* **67**, 1339-1344.

Roche, D.G., Careau, V. and Binning, S.A. (2016). Demystifying animal 'personality'(or not): why individual variation matters to experimental biologists. *J. Exp. Biol.* **219**, 3832-3843.

Royal Netherlands Academy of Arts and Sciences. (2018). Replication studies: improving reproducibility in the empirical sciences. https://www.knaw.nl/en/news/news/make-replication-studies-a-normal-part-of-science.

Sovrano, V.A., Bisazza, A. and Vallortigara, G. (2001). Lateralization of response to social stimuli in fishes: a comparison between different methods and species. *Physiol. Behav.* **74**, 237-244.

Sovrano, V.A., Dadda, M. and Bisazza, A. (2005). Lateralized fish perform better than nonlateralized fish in spatial reorientation tasks. *Behav. Brain Res.* **163**, 122-127.

Voelkl, B., Vogt, L., Sena, E.S. and Würbel, H. (2018). Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biol.* **16**, e2003693.

Wolak, M.E., Fairbairn, D.J. and Paulsen, Y.R. (2012). Guidelines for estimating repeatability. *Methods Ecol. Evol.* **3**, 129-137.

Woods, C.M. and Martin-Smith, K.M. (2004). Visible implant fluorescent elastomer tagging of the big-bellied seahorse, Hippocampus abdominalis. *Fisheries research* **66**, 363-371.

Yong, E. (2012). Bad copy. *Nature* **485**, 298-300.

Zwaan, R.A., Etz, A., Lucas, R.E. and Donnellan, M.B. (2018). Making replication mainstream. *Behav. Brain Sci.*