2017-01-20

Does repeatable behaviour in the laboratory represent behaviour under natural conditions? A formal comparison in sea anemones

Osborn, A

10.1016/j.anbehav.2016.10.036
Animal Behaviour
Elsevier BV

Does repeatable behaviour in the laboratory represent behaviour under natural conditions? A formal comparison in sea anemones.

Andrew Osborn

Mark Briffa

This is a manuscript of an accepted article published by Elsevier in *Animal Behaviour* (2017), available at http://dx.doi.org/10.1016/j.anbehav.2016.10.036

Marine Biology & Ecology Research Centre,

Plymouth University,

Plymouth PL3 8AA,

UK.

mark.briffa@plymouth.ac.uk

Short title

Repeatability in the laboratory and field

Word count: 9,901

Correspondence:

Mark Briffa

Marine Biology & Ecology Research Centre,

Plymouth University,

Plymouth PL3 8AA,

UK.

mark.briffa@plymouth.ac.uk

0044 (0)1752 584632

Abstract

Animal personality studies rely on collecting repeated behavioural data either in the field or in animals under laboratory conditions. Conditions in the field should be far less stable than controlled laboratory conditions, and hence represent a potential source of variation in behaviour. Here we report on the first experiment to our knowledge that formally compares the repeatability of identical behaviours in the laboratory and the field, and across the transition from laboratory to field. Using a design that controls for observation number we compared two groups of sea anemones, observed across two experimental phases, either (a) in the field followed by the laboratory or (b) in the laboratory only. We analysed differences in behaviour across a range of levels including repeatability and its among- and within-individual variance components. Although mean startle response durations vary between the laboratory and field, there was no significant difference in repeatability across situations. Within-individual variance differed between the two periods of the experiment for animals observed only in the laboratory but this effect was not present for those that transitioned from field to lab. Furthermore, the rank order of individual responses was stable for animals observed only in the laboratory but changed for those that transitioned from field to lab. These data show that although repeatability estimates in the laboratory can yield results that are similar to data collected in the field, the underlying components of consistent variation in behaviour might be influenced by an interaction between prior experiences and the current situation in which the animals are observed.

Keywords: *Actinia equina*, animal personality, field, IIV, laboratory, Repeatability

Introduction

Consistent among individual differences in behaviour have now been described in a wide array of animal taxa (Carere & Maestripieri, 2013) and although most studies have been performed on vertebrates there are several examples across invertebrate phyla including arthropods and cnidarians (Kralj-Fiŝer & Schuett, 2014). Usually referred to as animal personality (Dall, Houston, & McNamara, 2004; Sih, Bell, & Johnson, 2004) consistent among individual variation can be quantified as repeatability, R (Bell, Hankison &, Laskowski, 2009). This metric describes the proportion of total variation attributed to among individual variation, once other obvious covariates (e.g. body size) have been accounted for. Thus $R = V_{BI} /(V_{BI} + V_{WI})$, where V_{BI} denotes among (or 'between') individual variance and V_{WI} denotes within individual variance, typically the residual variance in statistical models that also include terms for V_{BI} (Cleasby, Nakagawa, & Schielzeth, 2015; Royauté, Buddle, & Vincent, 2015; Stamps, Briffa, & Biro, 2012; Westneat, Wright, & Dingemanse, 2014). Residual within-individual variance (also called intra individual variation, IIV; Stamps et al., 2012) has itself been the focus of recent interest. Along with understanding what factors might influence overall measures of repeatability the effects of external conditions on IIV are also of interest across a range of fields including psychology (Asendorpf, 1992), cognitive neuroscience (MacDonald, Nyberg, & Bäckman, 2006) and latterly behavioural ecology (e.g. Westneat et al., 2014). This measure quantifies the predictability or consistency of behaviour. In addition to biotic covariates that could influence V_{BI} and V_{WI} abiotic aspects of the environment could also influence repeatability and its variance components (Briffa, Bridger, & Biro, 2013; Royauté et al., 2015). For example, in poikilotherms metabolic rate and hence behaviour should fluctuate with temperature. For aquatic animals we also expect temporal fluctuations in a suite of physicochemical parameters that might influence behaviour including pH, dissolved oxygen and conductivity or salinity. Furthermore, spatial variation in

these parameters can lead to among individual variation in microhabitat, especially in sessile or sedentary animals. Therefore there has been some concern that there may be a disjoint between repeatability estimates made under relatively stable (temporally and spatially) laboratory conditions and those made under variable field conditions (e.g. Fisher, James, Rodríguez-Muñoz, & Tregenza, 2015a). Temporal variation in the field might increase V_{WI}, reducing repeatability (compared to laboratory conditions) whereas stable among individual variation in microhabitat may increase V_{BI}, increasing repeatability (compared to laboratory conditions). Such concerns are especially pertinent when laboratory repeatability estimates are obtained for wild caught animals, which have been bought into a novel environment, and when studies aim to make inferences about the fitness consequences of animal personality variation (Fisher et al., 2015a). Specific experimental paradigms where these concerns are important include 'two-step' studies where animals are observed in the laboratory and assigned to a given behavioural type (e.g. 'shy' or 'bold', or given a value along a shy-bold continuum) before being released and observed in the field (Niemelä, & Dingemanse, 2014), or when captive bred populations are used in personality research (Archard, & Braithwaite, 2010).

Although a number of animal personality studies have compared laboratory and field data, few have been designed to formally test the idea that estimates of repeatability could vary between the two situations (Fisher et al., 2015a; Fisher, David, Tregenza, & Rodríguez-Muñoz, 2015b; McCowan, Mainwaring, Prior, & Griffith, 2015; Yuen, Pillay, Heinrichs, Schoepf, & Schradin, 2016). In a recent example by Fisher et al. (2015a) wild field crickets, *Gryllus campestris*, were observed both *in situ* and in the laboratory through repeated cycles of recapture and release. They found that while three traits were significantly repeatable in the laboratory only activity and exploration were repeatable in the field setting. Furthermore, while exploration and activity in the laboratory correlated with the same traits in the field,

there was no correlation between boldness in laboratory and field conditions. Similarly, McCowan *et al*. (2015) found that although laboratory and field measures of exploration were both repeatable in zebra finches, *Taeniopygia guttata*, there was no correlation between the two situations. Results such as these imply that there is a mismatch between personality traits under laboratory and natural conditions, and therefore interpreting the evolutionary and ecological significance of laboratory based personality studies may be less than straightforward. Similarly, laboratory measures related well to field measures in blue tits, *Cyanistes caeruleus* (Herborn, Macleod, Miles, Schofield, Alexander, & Arnold, 2010), red squirrels, *Tamiasciurus hudsonicus* (Boon, Réale, & Boutin, 2008) and African striped mice, *Rhabdomys pumilio* (Yuen et al. 2016) but for Siberian chipmunks, *Tamias sibiricus*, there was no correlation between behaviours observed across the different situations (Boyer, Réale, Marmet, Pisanu, & Chapuis, 2010). While these studies have compared repeatabilities or mean level effects across laboratory and field settings an aspect that has yet to be investigated is how the specific variance components, V_{BI} and V_{WI}, might be affected by the experimental setting. Studies focussing on these components could identify the causes of differences in behaviour between the laboratory and field.

Beadlet sea anemones, *Actinia equina*, are sedentary cnidarians, common on the intertidal zone in north-western Europe. Mature polyps attach their basal disc to rocky substrata, a condition that can be reproduced in the laboratory (Rudin, & Briffa, 2011; 2012). At the opposite end of the column (the main mass of the body) is the oral disc through which they ingest food and eject waste. The oral disc is surrounded by six rows of feeding tentacles, used to trap prey and detritus from the water column and then to guide it to the oral disc. They also possess a single row of specialized acrorhagial tentacles, which are only easily visible during agonistic encounters (Fish, & Fish, 2011). When disturbed, *A. equina*, will retract their feeding tentacles to cover the oral disc. Following retraction of the tentacles, the

anemone will slowly reopen so that the tentacles and oral disc are again visible. In previous studies of behavioural variation in sea anemones the duration of tentacle retraction, also termed the 'startle response duration', has been used as an index of boldness and this is significantly repeatable in *A. equina* (Briffa, & Greenaway, 2011; Rudin, & Briffa, 2012) and in the Caribbean giant sea anemone *Condylactis gigantean* (Hensley, Cook, Lang, Petelle, & Blumstein, 2012). In *A. equnia*, significant repeatability has been found in studies based in the field (Briffa, & Greenaway, 2011) and in the laboratory (Rudin, & Briffa, 2012) but these studies differed in key experimental details (e.g. number of within individual replicates, presence of aggression) hindering the direct comparison of laboratory and field data. Since the retraction response can be evoked in both situations, using identical methods to disturb the anemones, *A. equina* is an ideal study subject for experiments specifically designed to compare laboratory and field based repeatabilities (e.g. see Carter, Marshall, Heinsohn, & Cowlishaw, 2012; Dochtermann, & Nelson, 2014). Here we investigate personality differences between typical laboratory conditions and field conditions, using an experimental design that controls for the duration of the experiment and the number of observations conducted in each situation: One group of anemones is repeatedly observed in the field for the first period of the experiment and then in the laboratory for the second period. A second group is observed concurrently with the first group and on an equal number of occasions but only in the lab, across both periods of the experiment.

First we ask whether sample level mean behaviour (i.e. the average of all individuals in the experiment) differs between laboratory and field based observations. Second, we ask whether repeatability differs between the two situations and whether its variance components (V_{BI} and V_{WI}) differ between situations. Third we investigate the specific effects of transitioning individuals from the field to the lab. For those anemones observed in the field, we also analyse key physicochemical seawater parameters that might influence their

behaviour. If laboratory based estimates of repeatability are comparable to those made in the field we would expect no significant differences in repeatability (or its variance components) across the two situations. If the transition from field to laboratory is important, we would expect to see differences between the two periods of the experiment in the group that transitioned from field to lab, but not in the group that spent the whole experiment in the lab.

Materials and methods

Study animals

Anemones were collected from tidepools at the base of rocky outcrops at the intertidal at Portwrinkle harbour beach, Cornwall, UK (50.361°N, 4.315°W), during July 2014. We identified 78 individuals of the red colour morph of *A. equina*, separated from one another by a distance of at least 1m, for use in the experiment. Previous molecular studies have shown that individuals separated by this distance are unlikely to be clone mates (e.g. Foster & Briffa, 2014; Turner, Lynch, Paterson, León-Cortés, & Thorpe, 2003). We immediately removed 39 individuals from the rocks, by inserting the edge of a thin silicone spatula under the pedal disk allowing them to be prized from the substrate. These were placed in individual containers and transported back to the laboratory. The remaining 39 individuals were left *in situ* during the first period of the experiment. For these individuals we applied unique identifying marks on the adjacent rock surface (using a non-toxic water-based paint) and took a digital photograph of each anemone and the identifying mark. This allowed us to re-identify the majority of individuals each time we returned to the study site. Half way though the experiment these individuals were removed from the rock and transported back to the laboratory as above. In the laboratory, all anemones were housed in individual tanks containing aerated seawater under a typical laboratory regime (Foster & Briffa, 2014; Rudin

& Briffa, 2011; 2012) for marine invertebrates: the seawater temperature was maintained at a constant 15°C, there was a 12:12h light:dark cycle and they were fed *ad libitum* on marine fish flakes.

Experimental design

Data were collected for two groups of anemones across two distinct time periods, A and B. The two groups were Laboratory-Laboratory (LL) (n = 39) and Field-Laboratory (FL) (n = 39). In the FL group observations during Period A were conducted in the field and during Period B they were conducted in the laboratory. In the LL group observations were conducted in the laboratory during both periods. Observations began after an interval of 3 days from the initial collection, to allow those that had been transferred to the laboratory a reasonable amount of time to attach to small flat rocks provided in their tanks and to acclimatize to the new environment. During each period observations were conducted on alternate days for anemones in each group such that the interval between observations was ca 48 h. The order of observation was randomized within groups. During the transition from period A to B there was a second 3 day interval where no data were collected for either group so that anemones in the FL group could acclimatize to laboratory conditions and attach to the rocks provided. Thus, the experiment contained four blocks of data, defined by the combination of Group and Period (FL.A, FL.B, LL.A, LL.B). We aimed to record 6 startle responses from each anemone in each of the four blocks of the experiment. For those in the FL group we were unable to re-identify 9 anemones at varying points during period A and hence for these 9 individuals the number of observations during period A was fewer than 6 (range = 3 to 5 observations; for all other individuals we obtained the full set of 6 observations). Obviously, it was not possible to collect these anemones for use in period B of the experiment. Therefore, to allow a balanced comparison of startle responses between the FL and LL groups during period B, we collected data from only 30 individuals (selected at

random) in the LL group during period B. Thus, the number of individuals in each period of the experiment was identical for the two treatment groups. During period A there were slightly more observations in the LL group compared to the FL group (LL, n = 234; FL, n = 221) but during period B the number of observations in each group was identical (LL, n = 180; FL, n = 180). Thus the total number of observations was N = 815 (see table A1 for a summary of the distribution of observations among individual anemones). For logistical reasons we set a maximum observation time of 2 h. Only 10 observations from the total of 815 reached this limit and, since the proportion was very low (0.012), we included these observations in the analyses that follow. It was not possible to record data blind because our study involved focal animals in the field.

Behavioural tests

To stimulate the startle response, we aimed a jet of sea water at the oral disc of open anemones. This was applied through the rapid discharge of 10 ml of seawater from a syringe at a distance of 2 cm from the oral disc. The duration of the startle response was timed with a stopwatch to the nearest second from the end of the stimulus until the point at which the anemone had re-opened its feeding tentacles and the oral disc was again fully visible to the observer (Rudin, & Briffa, 2012). For observations that were conducted in situ we collected startle response data at low tide, while the tidepools containing the anemones were readily accessible. We measured the following seawater parameters from the tidepools; dissolved oxygen, temperature, pH and salinity (YSI 2030 Pro Plus portable meter). At the end of the experiment we estimated the body size of each anemone by taking two perpendicular measures of the pedal disc diameter and finding the average. We chose this method of

estimating their size over the alternative of measuring their dry weight because we wanted to return the anemones to the shore at the end of the experiment.

Statistical methods

To compare mean responses and residual variance of responses between the laboratory and the field we used a double hierarchical general linear model (DHGLM), implemented within a Bayesian framework. The rationale for such models is described in detail elsewhere (Bridger et al., 2015; Westneat et al., 2012). Briefly, DHGLMs comprise a 'mean model' and a 'standard deviation model'. The mean portion models fixed and random effects on sample mean values and the standard deviation portion models the effects of fixed and random variables on the variance around these means. The parameters in the standard deviation model thus represent residual variance. We configured a DHGLM with the following fixed effects: Period, group, observation number and all of the interactions between these factors. We also included anemone size as a covariate. Observation numbers were coded 1 to 6 within each period (A and B), i.e. observation number was crossed with period such that a significant period * observation interaction would indicate different patterns of change in startle response duration during each phase of the experiment. Moreover, a three way period * observation number * group interaction would indicate that any difference in the effect of observation number between periods A and B differed between the two treatment groups (LL and FL). In the mean model we allowed random intercepts for each individual and random slopes across repeated observations. The random intercept term allows for among individual variation (V_{BI}) in startle response duration and the random slope term allows among individual variation in how startle responses change over multiple observations (i.e. individual * observation number reaction norms). For the SD model we allowed random

intercepts only (V_{WI}), since there was only 1 observation for each individual per occasion, so it is not possible to model changes in variance across observations. The DHGLM was fitted using the JAGS software (Plummer, 2003), which we controlled from within the R statistical computing environment using the package RJAGS (3.13) (Plummer, 2014). Details of the modelling setup are given in the Appendix.

To compare repeatability (and its components, V_{BI} and V_{WI}) across treatment groups and periods, we used a hierarchical generalized linear model (HGLM), following Royauté et al. (2015). In contrast to the DHGLM, the HGLM does not contain an overall random intercept, but has (a) separate random intercepts specified for each block of the experiment and (b) separate residual variances for each block of the experiment. This model contained the same fixed effects as for the DHGLM described above and details of the modelling setup are given in the Appendix. In the R package MCMCglmm (Hadfield, 2010) that we used for this analysis these components are described as the G-structure and R-structure respectively and these correspond to V_{BI} and V_{WI}. Thus, we were able to calculate posterior modes for repeatability and its 95% CIs for each block of the experiment, basing all four repeatability estimates on variance components extracted from this single model. To determine whether there were significant differences in repeatability we calculated differences in repeatability (ΔR) between periods of the experiment ($\Delta R = R_B - R_A$) and between the treatment groups ($\Delta R = R_{FL} - R_{LL}$) (Royauté et al., 2015). Delta values were considered to be significant if the 95% CIs of their posterior modes did not cross zero. We also calculated the corresponding delta values for the variance components V_{BI} and V_{WI}. Note that such delta values (e.g. see Table 1) are not expected to equate to the simple arithmetical difference between posterior mode values for repeatability (or the variance components). This is a consequence of the Bayesian framework, which in this case produces a vector of 1000 posterior estimates for repeatability (for each block of the experiment), of which we report the posterior modes. Our

ΔR values are thus based on the posterior modal difference between pairs of vectors of repeatability estimates (rather than the difference between two single repeatability estimates).

To determine whether startle responses in the first phase of the experiment could predict those in the second phase within each treatment group (LL and LF) we first tested for experiment-wide repeatability (i.e. across all 12 observations, rather than within each period-specific block of 6 observations), and then tested for between-individual correlations across each phase of the experiment. To test for experiment-wide repeatability we specified a HGLM similar to that described above (see Appendix) but in this case we specified random intercepts (V_{BI} or G-structure) and residual variance (V_{WI} or R-structure) by treatment group (LL and FL) rather than by each block of the experiment. Therefore in this analysis observations were coded 1-12 (i.e. observations 1-6 occurred in block A and observations 7-12 appeared in block B) and we did not include interactions between observation number and any of the other predictors. We then extracted V_{BI} and V_{WI} to calculate treatment group specific repeatabilities across all 12 observations. To test for between individual correlations across periods A and B, one approach would be to use a multivariate mixed model; i.e. where responses in periods A and B are both treated as response variables and the model is specified so that between individual (but not within individual) covariance is estimated (Dingemanse & Dochtermann, 2013). However, our data were not multivariate normal (MVN) and initial analyses indicated that any correlations would be likely to be of magnitudes too low to detect using this approach due to low statistical power (Dingemanse & Dochtermann, 2013). Since we were unable to reliably use multivariate mixed models, and since within-individual covariation should not be estimated for the experimental design we used here (as startle responses for each observation in period A did not coincide with observations in period B, see Dingemanse & Dochtermann, 2013) we used an alternative approach of using a univariate mixed model where the response variable was the startle responses during period B and the

predictors were startle responses during period A, treatment group and the interaction between them. Size was also included as a covariate and anemone ID was included as a random intercept. This model was calculated using MCMC (MCMCglmm). The predictors were individual mean startle responses during period A, treatment group and the interaction between them. To further describe the treatment group specific correlations we used Spearman rank correlations between the average responses of each individual in periods A and B. For all the analyses (except Spearman's correlations) data were first $\log_{10}$ transformed to improve normality.

Ethical statement

This study was carried out following the ASAB/ABS guidelines for the use of animals in teaching and research. No licenses are required to work on this species in the UK. All anemones were returned to the shore at the end of the experiment.

Results

Mean and residual variance of startle response duration in the laboratory and field

The parameter estimates from the DHGLM and their 95% CIs are given in table 1. The fixed effects component of the mean model provided no evidence that startle response durations varied the two treatment groups ($P = 0.53$) but there was a non-significant trend for startle responses to decline with body size ($P < 0.052$). Startle responses varied across observation number ($P < 0.0001$) and period ($P < 0.0001$). A significant observation * period interaction effect ($P < 0.0001$) indicates that startle responses increased during period A and, although starting at relatively high level (i.e. in observation B1), decreased during period B (Fig. 1).

There was also an interaction effect between group and period (P = 0.009). Inspection of the degree of overlap in 95% credible intervals shows that for anemones in the LL group, which remained in the laboratory across both periods, startle response durations were stable across both periods of the experiment (Fig. 2). In contrast, those in the FL group showed shorter startle responses during period A compared with period B (Fig. 2). There was no interaction between observation and group (P = 0.41) and no three-way interaction (P = 0.78). The random intercept in the mean model provides strong evidence that individuals differed in intercepts, indicating among individual variation in startle response duration. Similarly, there was evidence for a random slope effect with respect to observation number, indicating among-individual differences in how the anemones responded to repeated observations.

The fixed effects of the SD model indicate that the IIV of startle response duration was unaffected by treatment group (P = 0.89) or between periods of the experiment (P = 0.98) but there was a non-significant trend for startle responses to increase with anemone size (P = 0.053). There was no evidence for an interaction effect between group and period (P = 0.21). The random intercept portion of the SD model provides strong evidence that individuals differ in intercepts, suggesting among-individual variation in IIV across the experiment as a whole.

Comparing the repeatability of startle responses in the laboratory and field

Repeatability estimates derived from the HGLM (table A2) indicate that startle responses were significantly repeatable for each block of data in the experiment (FL.A, FL.B, LL.A, LL.B) and that there were no significant differences in repeatability, either between periods within treatment groups, or between treatment groups within periods (table 2). Similarly, comparison of the posterior modes of block specific estimates of V_{BI} (ΔV_{BI}) indicate that

while there was significant among individual variation in startle response duration within each block of the experiment, there were no significant differences in V_{BI} among data blocks (table 3). Comparisons of the posterior modes of estimates of V_{WI} (ΔV_{WI}) indicate that during period A, IIV was greater in the LL group compared to the FL group. Furthermore, IIV declined between periods A and B for the LL group but there was no such decline for the FL group (table 3).

Are startle responses in the laboratory and field correlated?

There was significant experiment-wide repeatability for both treatment groups (LL; R = 0.19 [95% CIs = 0.11, 0.29], FL; R = 0.22 [95% CIs = 0.13, 0.31]) and there was no significant difference in repeatability between the two groups (ΔR = >0.01[95% CIs = -0.11, 0.15]). There was a trend for startle responses during period B to be greater for the FL treatment group than for LL (posterior mean = 428.8 [lower and upper 95% CIs = -125, 868], P_{MCMC} = 0.1) but there was no overall correlation between startle responses in periods A and B (posterior mean = -0.07 [lower and upper 95% CIs = -0.19, 0.54], P_{MCMC} = 0.24) and there was no effect of anemone size (posterior mean = 159.2 [lower and upper 95% CIs = -102, 465], P_{MCMC} = 0.26). However, a significant interaction effect (posterior mean = -0.21 [lower and upper 95% CIs = -0.41, -0.02], P_{MCMC} = 0.022) indicated a trend for a positive correlation across both periods for the LL group (Rho = 0.35, P = 0.057), while there was no significant correlation for the FL group (Rho = -0.063, P = 0.73) (Fig. 3).

Physicochemical variables and startle responses in the field

We used an HGLM to analyze the effects of a range of environmental parameters (that can fluctuate in the field) on startle response duration during period A for anemones in the FL group; seawater temperature, dissolved oxygen, salinity and pH. None of these predictors influenced startle response durations (see table A3 for details of this model). We nevertheless calculated repeatability adjusted for these variables. The adjusted repeatability for startle responses in the FL.A block was slightly higher than the unadjusted value reported in table 2 (R_{ad} = 0. 46 [95% CIs = 0.34, 0.59]). However, the difference between R_{ad} and the unadjusted R reported in table 2 was not significant (ΔR (R_{ad} – R) = 0.08 [95% CIs = -0.07, 0.29]). Furthermore an unadjusted repeatability, calculated again only for the FL.A block, but this time using a version of the HGLM that excluded the environmental predictors, was virtually identical to the adjusted version, (R = 0.46 [95% CIs = 0.34, 0.58]).

Discussion

Although there was equivalent repeatability between laboratory and field conditions, our data show three main differences in the behaviour of individuals that transitioned from field to laboratory (FL) compared to those that remained in the laboratory for the entire experiment (LL). First, for the FL group, sample mean startle responses (i.e. the mean of all individuals across all observations within a given period of the experiment) in the field were of low duration then of high duration when transferred to the laboratory (Fig. 2). In contrast, sample mean responses in the LL group were of intermediate duration across both periods. Second, for the LL group within-individual variance was greater in period A than in period B, but for the FL group within-individual variance did not differ between periods (table 3). Third, the rank order of individual mean responses (i.e. individual-specific means within a given period

of the experiment) was relatively consistent for anemones in the LL group but not for those in the FL group (Fig. 3).

The interaction between group and period (Fig. 2) is surprising in two regards. First, we might have expected that the anemones would be more risk-averse (i.e. longer startle responses) in the field compared to the laboratory situation, due to the presence of predators and the possibility of damage from wave action. In this case startle responses would have been longer in the field than in the laboratory, since tentacle retraction is a protective behaviour (Edmunds, Potts, Swinfen, & Waters, 2009), but this is not what we found. Second, the group x period interaction indicates that anemones that have experienced repeated perturbation in the field are more risk-averse when they go on to receive the same type of disturbance again in the laboratory. A possible explanation for the increased duration of startle responses in the laboratory by anemones in the FL group is that it represents a reaction to recent handling where the anemones have been removed from the substrate and placed into a new environment. However, the individuals that were observed in the laboratory during period A (LL.A) did not show similarly long startle responses after identical recent handling procedures. Therefore, an alternative explanation is that the longer startle responses in period B by anemones in the FL group indicates a combined effect of prior experience of the stimulus in the field with recent handling and/or the change in environment. We note, however, that this possibility is difficult to test directly with the current data. This would require an experiment where both groups were removed from the substrate and allowed to resettle at the beginning of each period of the experiment, although this would be logistically difficult to achieve.

Regardless of its underlying causation the period x group interaction does suggest that anemones retain information on recent experiences of disturbance. Previous studies have

demonstrated simple learning in the colonial sea anemone *Anthopleura elegantissima*, where tentacle retraction times decrease following repeated stimulation with a water jet (Logan, 1975; Logan & Beck, 1978), i.e. they habituate. Here, we did not find a simple pattern of habituation. Rather, an observation x period interaction (Fig. 1) indicates that anemones in both groups appeared to sensitize during the first phase of the experiment and then habituate during period B. A possible explanation is that these changes in startle responses reflect a 'dual process habituation' (Thompson, Groves, Teyler, & Roemer, 1973). Here, both sensitisation and habituation occur simultaneously, but high initial levels of sensitisation eventually dissipate, to reveal an underlying pattern of habituation. Among non-human animals a study of the bull frog, *Rana catesbeiana*, (Bee 2001) provides limited support for dual process habituation but further experiments would be needed to confirm whether dual process habituation occurs in sea anemones. A second possibility is that the underlying cause is similar to that which drove the group x period interaction. That is, experience of the stimulus in period A is retained through to period B such that responses to repeated stimulation (when it is encountered again after a pause) are modified by the prior experience of repeated stimulation. Adaptive behavioural plasticity, based on prior experience, has been shown (at the individual mean level) in rainbow trout (*Onchorhyncus mykiss*) subjected to different types of social interaction (Frost, Winrow-Giffen, Ashley, & Sneddon, 2007). In the current study, where the stimulus elicited tentacle retraction but did not cause damage, continued sensitisation might be maladaptive due to loss of feeding time. Therefore it is possible that an initial period of sensitisation (during which some learning has taken place) was replaced by habituation during the second part of the experiment.

Within these effects at the level of mean responses, the initial analysis indicated that (a) there was significant consistent among individual variation in behaviour (random intercepts in the mean portion of the DHGLM) and (b) significant among individual variation

in IIV (random intercepts in the SD portion of the DHGLM), similar to a range of other animals where this has been specifically modelled (Biro & Adriaenssens, 2013; Bridger et al., 2015; Jennings, Hayden, & Gammell, 2013; Stamps et al., 2012). In other words, we found among-individual variation in the consistency of behaviour. In the follow-up HGLM analyses, where we specified random intercepts and residual variances per block of the experiment, we further investigated these patterns of among and within-individual variation. In contrast to the sample mean level effects, there were no significant differences in repeatability between laboratory and field data or between periods of the experiment. Similarly, there were no differences in among individual variance (V_{BI}) between treatments or periods. The patterns for IIV (i.e. V_{WI}), however, were more complex. Anemones in the LL treatment showed greater IIV during period A compared to those in the FL treatment. In period B both treatment groups showed similar levels of IIV. Thus, as with mean-level behaviour, the situation in which the startling stimulus is first encountered appears to influence the behavioural responses to it. While encountering responses first in the field appears to lead to longer mean level responses, encountering responses first in the laboratory leads to elevated IIV.

This result highlights the need for experiments that control for the number of stimulus exposures when comparing repeatabilities across situations, whether the change in situation occurs due to a transition between field and laboratory conditions or across some other variable such as temperature (Briffa et al., 2013) or the presence of a predator (Briffa, 2013; Briffa, Rundle & Fryer, 2008). High IIV (low predictability) might be associated with high risk situations (Briffa, 2013; Briffa et al., 2013) and thus anemones in the LL group may have initially perceived their new environment (during period A) as a high risk one. It is curious then that IIV was not similarly elevated during period B for the anemones in the FL group. The difference between the FL.B and LL.A blocks of the experiment was that those in the

FL.B block had experienced the stimulus prior to being observed in the lab, whereas those in the LL.A block had not. Again, it appears that responses to the stimulus are modified by prior exposure to it.

Up to this point, it has been difficult to directly compare repeatabilities of startle responses in beadlet sea anemones under laboratory and field conditions. In two previous studies (Briffa, & Greenaway, 2011; Rudin, & Briffa, 2012) there were similarly high levels of consistent among individuals variation in the laboratory and the field. However, these two studies were not conducted concurrently and there were other differences between the experimental conditions. In both of these studies the estimates of repeatability were greater than those in the current study. A number of factors differed between the previous studies and this one, including the year of study and different study sites, again highlighting the need for experiments designed specifically to compare repeatabilities simultaneously under the two situations. The most obvious explanation for the lower estimates of R in the current study is that here we estimated repeatability over a greater number of observations per individual than in the previous studies where it was estimated over only two to three observations compared to a maximum of twelve observations per individual here. Indeed, although we found significant repeatability across all twelve observations in our study, these estimates were lower than those made across the six observations in each period of the experiment. A negative correlation between repeatability and number of observations has been noted across a wide range of study systems (Bell et al., 2009). In the previous study of *in situ* repeatability in this species (Briffa & Greenaway, 2011) R was adjusted on seawater temperature variation, since individual anemones in the field are likely to experience spatially consistent differences in microchabitat (although they may also experience significant levels of temporal variation in microhabitat). In that study, adjusting for temperature still yielded high estimates of repeatability. Similarly in the current study, for those individuals observed in the field we

found that none of the physicochemical variables that we measured influenced startle response durations, or had any significant effects on repeatability, since there was no difference in adjusted and unadjusted estimates. Thus for *A. equnia* it appears unlikely that spatial variation in microhabitat is sufficient to drive consistent among individual differences in behaviour.

Overall, among-individual variation in behaviour was maintained across both periods, which for one group represented a change in environment. Nevertheless, we only found a trend for positive among-individual correlation between average startle responses across both periods of the experiment for the anemones that were in the laboratory for both periods. In this group, those that gave the longest startle responses in period A also tended to give long startle responses in period B. For those in the field to laboratory treatment, those that gave the longest startle responses in period A were not significantly likely to give the longest responses later in the lab. These results for boldness in *A. equina*, that it is repeatable in both the laboratory and the field (and repeatable across both periods of the experiment for both treatment groups) but there is a lack of correlation between the two situations, are similar to those for crickets (Fisher et al., 2015a) and zebra finches (McCowan et al., 2015). However, compared to these studies, our data were collected in a different way. Here we observed anemones in two discrete observation periods; in contrast, the studies on crickets and zebra finches involved repeatedly releasing and recapturing the animals between the laboratory trials. Thus, while boldness in the laboratory may be influenced by stress responses to repeated recapture (Fisher et al., 2015a) in some studies, this is not the case for the present data. The lack of correlation between data collected in periods A and B for anemones in the FL treatment is likely to contribute to the random slope effect seen in the initial double hierarchical model. For this group, individual mean responses tended to change across the two periods of the experiment. In contrast, individual mean responses were more similar

across the two periods for those anemones in the LL group. Thus, individuals differed in their reaction norms (Dingemanse, Kazem, Réale, & Wright, 2010) across the twelve observations.

Change among individuals in their rank order of responses, as seen here in the FL group, is not unusual in experiments where a change in situation is experienced (e.g. Briffa, 2013; Briffa et al., 2013; Fisher et al., 2015a; although see Fürtbauer, Pond, Heistermann, & King, 2015). Such patterns indicate that individuals differ in the amount of behavioural plasticity that they show, i.e. they have different behavioural reaction norms (Dingemanse et al., 2010). The reasons why the transition from field to laboratory should lead to such changes in the current data are not clear. It is not an effect of repeated stimulation, since there was no similar change in the LL group. Indeed, even though there were some changes in the rank order of individual mean responses, responses in the FL group were repeatable across all twelve trials. It is also difficult to ascribe the effect to a change in the abiotic habitat since none of the physicochemical variables measured in the field had any noticeable effect on startle response durations. Further, the changes were not due to handling stress resulting from repeated recapture and release, which was absent in our experiment. However, they may represent stress responses to the single collection procedure and transport to the lab, or to unmeasured aspects of the environment that differ between the field and laboratory situation. If so, it appears that for such effects to impact on startle response duration, they must be present in combination with prior experience of the startling stimulus, since they were absent in the LL group. Added to the results for sample mean level behaviour and for within-individual variation in behaviour, this result provides further evidence that the situation in which a novel stimulus is first encountered can influence subsequent behavioural responses to it.

Although we found equivalent repeatabilities in the field and laboratory, a component of behavioural variation that contributes to repeatability (V_{WI}) differed between groups that

were observed under each condition. In addition, the results of the mean level part of the analysis indicate that the startle responses during period B, at the levels of (a) overall average startle responses during period B and (b) changes in startle response across repeated stimulation during period B, might be influenced by experiences during period A. These differences among the four blocks of data indicate that the combination of prior experience and the situation where the prior experiences occurred could influence behaviour. We therefore suggest the following points are worth considering when designing experiments using wild caught animals: (1) the likely amount of handling stress during capture, (2) the time allowed for acclimation to the laboratory situation and (3) whether or not the animals are naïve to the test procedures used. The importance of comparing laboratory and field collected data in longitudinal behavioural studies has been highlighted by a range of authors (Carter et al., 2012; Dochtermann & Nelson, 2014; Niemelä & Dingemanse, 2014; Stamps et al., 2012). We suggest that studies comparing behaviour in the laboratory and field should also be relevant to a much wider range of studies that do not necessarily involve testing for repeatability, but where behaviour in wild caught animals is nevertheless observed in the laboratory. Such studies are often the only feasible way of performing manipulative experiments in animal behaviour but insights into how the behaviour of interest (and across which levels of analysis) might be influenced by bringing animals into the laboratory and by prior experiences can only help in the interpretation of the results that they yield.

Acknowledgements

We are grateful to two reviewers for their constructive comments that have improved this study.

References

Archard, G.A., & Braithwaite, V.A. (2010). The importance of wild populations in studies of animal temperament. *Journal of Zoology*, 28, 149-160.

Asendorpf, J.B. (1992). Beyond stability: Predicting inter-individual differences in intra-individual change. *European Journal of Personality*, 6, 103–117.

Bee, M.A. (2001). Habituation and sensitization of aggression in bullfrogs (*Rana catesbeiana*): Testing the dual-process theory of habituation. *Journal of Comparative Psychology*, 115, 307-316.

Bell, A.M., Hankison S.J., & Laskowski, K.L. (2009). The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, 77, 771–783.

Biro, P.A., & Adriaenssens, B. (2013). Predictability as a personality trait: consistent differences in intraindividual behavioral variation. *American Naturalist*, 182, 621-629.

Boon, A.K., Réale, D., & Boutin, S. (2008). Personality, habitat use, and their consequences for survival in North American red squirrels *Tamiasciurus hudsonicus*. *Oikos*, 117, 1321–1328.

Boyer., N., Réale, D., Marmet, J., Pisanu, B., & Chapuis, J.L. (2010). Personality, space use and tick load in an introduced population of Siberian chipmunks *Tamias sibiricus*. *Journal of Animal Ecology*, 79, 538–547.

Bridger, D., Bonner, S.J., & Briffa, M. (2015). Individual quality and personality: bolder males are less fecund in the hermit crab *Pagurus bernhardus*. *Proceedings of the Royal Society B: Biological Sciences*, 282, 20142492.

Briffa, M. (2013). Plastic proteans: reduced predictability in the face of predation risk in

hermit crabs. *Biology Letters* 9, 20130592.

Briffa M., Bridger D., & Biro, P.A. (2013). How does temperature affect behaviour? Multilevel analysis of plasticity, personality and predictability in hermit crabs. *Animal Behaviour*, 86, 47–54.

Briffa, M., & Greenaway, J. (2011). High *In Situ* Repeatability of Behaviour Indicates Animal Personality in the Beadlet Anemone *Actinia equina* (Cnidaria). *PLoS One*, 6, e21963.

Briffa, M., Rundle, S.D., & Fryer, A. (2008). Comparing the strength of behavioural plasticity and consistency across situations: animal personalities in the hermit crab *Pagurus bernhardus*. *Proceedings of the Royal Society B: Biological Sciences*, 275, 1305–11.

Carter, A.J., Marshall, H.H., Heinsohn, R., & Cowlishaw, G. (2012). How not to measure boldness: novel object and antipredator responses are not the same in wild baboons. *Animal Behaviour*, 84, 603–609.

Carere, C., & Maestripieri, D. (editors). 2013. *Animal Personalities: Behavior, Physiology and Evolution*. Chicago: Chicago University Press.

Cleasby, I.R., Nakagawa, S., & Schielzeth, H. (2015). Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology & Evolution*, 6, 27–37.

Dall, S.R.X., Houston, A.I., & McNamara, J.M. (2004). The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters*, 7, 734–739.

Dingemanse, N.J., Kazem, A.J., Réale, D., & Wright, J. (2010). Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, 25, 81-9.

Dingemanse, N.J., & Dochtermann, N.A. (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology*, 82, 39–54.

Dochtermann, N.A. & Nelson, A.B. (2014). Multiple Facets of Exploratory Behavior in House Crickets (*Acheta domesticus*): Split Personalities or Simply Different Behaviors? *Ethology*, 120, 1110–1117.

Edmunds, M., Potts, G.W., Swinfen, R.C., & Waters, V.L. (2009). Defensive behaviour of sea anemones in response to predation by the opisthobranch mollusc *Aeolidia papillosa* (L.). *Journal of the Marine Biological Association of the UK*, 56, 65.

Fish, J.D., & Fish, S. (2011). *A Student's Guide to the Seashore*. Cambridge University Press, Cambridge.

Fisher, D.N., James, A., Rodríguez-Muñoz, R., & Tregenza, T. (2015a). Behaviour in captivity predicts some aspects of natural behaviour, but not others, in a wild cricket population. *Proceedings of the Royal Society B: Biological Sciences*, 282, 20150708.

Fisher, D.N., David, M., Tregenza, T., & Rodríguez-Muñoz, R. (2015b). Dynamics of among-individual behavioural variation over adult lifespan in a wild insect. *Behavioral Ecology*, 26, 975–985.

Foster, N.L., & Briffa, M. (2014). Familial strife on the seashore: aggression increases with relatedness in the sea anemone *Actinia equina*. *Behavioural Processes*, 103, 243–245.

Frost, A.J., Winrow-Giffen, A., Ashley, P.J., & Sneddon LU. (2007). Plasticity in animal personality traits: does prior experience alter the degree of boldness? *Proceedings of the Royal Society B: Biological Sciences*, 274, 333-339.

Fürtbauer, I., Pond, A., Heistermann, M., & King, A.J. (2015). Personality, plasticity and

predation: linking endocrine and behavioural reaction norms in stickleback fish. *Functional Ecology*, 29, 931-940.

Hadfield, J.D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The {MCMCglmm} {R} Package. *Journal of Statistcal Software*, 33, 1–22.

Hensley, N.M., Cook, T.C., Lang, M., Petelle, M.B., & Blumstein, D.T. (2012). Personality and habitat segregation in giant sea anemones (*Condylactis gigantea*). *Journal of Experimental Marine Biology & Ecology*, 426-427, 1–4.

Herborn, K.A., Macleod, R,, Miles, W.T.S., Schofield, A.N.B., Alexander, L., & Arnold K.E. (2010). Personality in captivity reflects personality in the wild. *Animal Behaviour*, 79, 835–843.

Jennings, D.J., Hayden, T.J., & Gammell, M.P. (2013). Personality and predictability in fallow deer fighting behaviour: the relationship with mating success. *Animal Behaviour*, 86, 1041–1047.

Kralj-Fiŝer, S., & Schuett, W. (2014). Studying personality variation in invertebrates: Why bother? *Animal Behaviour*, 91, 41–52.

Logan, C.A. (1975). Topographic changes in responding during habituation to water-stream simulation in sea anemones (*Anthopleura elegantissima*). *Journal of Comparative Physiology and Psychology*, 89, 105-117.

Logan, C.A., & Beck, H.P. (1978). Long-term retention of habituation in the sea anemone (*Anthopleura elegantissima*). *Journal of Comparative Physiology and Psychology*, 92, 928-936.

MacDonald, S.W.S., Nyberg, L., & Bäckman, L. (2006). Intra-individual variability in

behavior: links to brain structure, neurotransmission and neuronal activity. *Trends in Neuroscience*, 29, 474-480.

McCowan, L.S.C., Mainwaring, M.C., Prior, N.H., 7 Griffith, S.C. (2015). Personality in the wild zebra finch: exploration, sociality, and reproduction. *Behavioral Ecology*, 26, 735–746.

Niemelä, P.T., & Dingemanse, N.J. (2014). Artificial environments and the study of 'adaptive' personalities. *Trends in Ecology & Evolution*, 29: 245–7.

Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March*, pp. 20–22.(doi:10.1.1.13.3406)

Plummer, M. (2014). rjags: Bayesian graphical models using MCMC. pp. http://CRAN.R–project.org/package=rjags.

Royauté, R., Buddle, C.M., & Vincent, C. (2015). Under the influence: sublethal exposure to an insecticide affects personality expression in a jumping spider. *Functional Ecology*, 29, 962–970.

Rudin, F.S., & Briffa, M. (2011). The logical polyp: assessments and decisions during contests in the beadlet anemone *Actinia equina*. *Behavioral Ecology*, 22, 1278-1285.

Rudin, F.S., & Briffa, M. (2012). Is boldness a resource-holding potential trait? Fighting prowess and changes in startle response in the sea anemone, *Actinia equina*. *Proceedings of the Royal Society B: Biological Sciences*, 279, 1904–10.

Sih, A., Bel,l A., & Johnson, J.C. (2004). Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution*, 19, 372–8.

Stamps, J.A., Briffa, M., & Biro, P.A. (2012). Unpredictable animals: individual differences

in intraindividual variability (IIV). *Animal Behaviour*, 83, 1325–1334.

Thompson, R.F., Groves, P.M., Teyler, T.J., & Roemer, R.A. (1973). A dual process theory of habituation: Theory and behavour. In Peeke, H.V.S., & Herz, M.J. (Eds) Behavioral studies volume 1: *Habituation* (pp 239-272). Academic Press, London.

Turner, V.L.G., Lynch, S.M., Paterson, L., León-Cortés, J., & Thorpe, J. (2003). Aggression as a function of genetic relatedness in the sea anemone *Actinia equina* (Anthozoa : Actiniaria). *Marine Ecology Progress Series*, 247, 85–92.

Westneat, D.F., Schofield, M., & Wright, J. (2012). Parental behavior exhibits among-individual variance, plasticity, and heterogeneous residual variance. *Behavioral Ecology*, 24, 598–604.

Westneat, D.F., Wright, J., & Dingemanse, N.J. (2014). The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, 90, 729-743.

Yuen, C.H., Pillay, N., Heinrichs, M., Schoepf, I., & Schradin, C. (2016). Personality traits are consistent when measured in the field and in the laboratory in African striped mice (*Rhabdomys pumilio*). *Behavioral Ecology & Sociobiology*, 70, 1235.

Table 1. Posterior summary statistics for the mean model and the SD model.

Parameter	Mean	SD	95% CI lower	95% CI upper	*P*
Mean model					
Intercept	0.11	0.09	-0.06	0.28	0.19
Size	-0.1	0.05	-0.02	0.001	0.052
Period	-0.36	0.09	-0.54	-0.20	<0.0001
Group	-0.03	0.08	-0.26	0.21	0.82
Observation	-0.30	0.06	-0.42	-0.16	<0.0001
Period * Group	0.32	0.13	0.07	0.56	0.009
Observation * Group	0.07	0.09	-0.11	0.26	0.41
Observation * Period	0.67	0.07	0.50	0.85	<0.0001
Observation * Period * Group	-0.04	0.12	-0.27	0.20	0.78
Anemone ID (intercept)	0.33	0.05	0.24	0.43	
Anemone ID (observation)	0.15	0.06	0.02	0.25	
SD model					
Intercept	-0.22	0.07	-0.36	-0.08	0.002
Size	0.06	0.03	-0.001	0.12	0.053
Period	0.004	0.08	-0.16	0.18	0.98
Group	-0.01	0.10	-0.19	0.18	0.89
Period * Group	0.15	0.12	-0.09	0.38	0.21
Anemone ID (intercept)	0.12	0.06	0.01	0.24	

Posterior estimates for factors and covariates with their standard deviations and lower and upper 95% CIs are shown. P-values are given for fixed effects only.

Table 2. Posterior mode MCMC repeatability estimates for each block of data and ΔR values.

	A	B	ΔR (B-A)
LL	0.22 [0.14, 0.35]	0.33 [0.26, 0.54]	0.15 [-0.02, 0.34]
FL	0.31 [0.19, 0.43]	0.44 [0.28, 0.57]	0.13 [-0.09, 0.29]
ΔR (FL-LL)	0.04[-0.08, 0.24]	0.04 [-0.17, 0.25]	

Posterior modes of differences between periods of the experiment (ΔR = B-A) and for the difference between treatment groups (ΔR = FL-LL) are shown. Upper and lower 95% CIs for R and ΔR values are given in square brackets.

Table 3. Posterior modes for among- and within-individual variation in startle response duration with ΔV values.

Among individual variation, V_{BI}			
	A	B	ΔV_{BI} (B-A)
LL	0.23 [0.14, 0.45]	0.35 [0.19, 0.65]	0.06 [-0.16, 0.45]
FL	0.26 [0.13, 0.43]	0.38 [0.22, 0.66]	0.13 [-0.14, 0.41]
ΔV_{BI} (FL-LL)	0.002 [-0.23, 0.22]	0.03 [-0.36, 0.34]	
Within individual variation, V_{WI}			
	A	B	ΔV_{WI} (B-A)
LL	0.84 [0.72, 1.08]	0.59 [0.48, 0.75]	**-0.27 [-0.54, -0.08]**
FL	0.57 [0.47, 0.71]	0.53 [0.42, 0.66]	-0.08[-0.22, 0.10]
ΔV_{WI} (FL-LL)	**-0.32 [-0.652, -0.10]**	-0.05 [-0.25, 0.11]	

Posterior modes for the difference between periods of the experiment (B-A) and for the difference between treatment groups (FL-LL) are shown. Upper and lower 95% CIs for V and ΔV values are given in square brackets and significant delta values are shown in bold.

Table A1. The number of observations per individual, in each treatment group and period.

	Field to Lab (FL)			Lab to Lab (LL)	
	Period			Period	
ID	A	B	ID	A	B
F00	6	6	L00	6	6
F01	6	6	L01	6	6
F02	6	6	L02	6	6
F03	6	6	L03	6	6
F04	6	6	L04	6	6
F05	6	6	L05	6	6
F06	6	6	L06	6	6
F07	6	6	L07	6	6
F08	6	6	L08	6	6
F09	6	6	L09	6	6
F10	6	6	L10	6	6
F11	6	6	L11	6	6
F12	6	6	L12	6	6
F13	4		L13	6	
F14	5		L14	6	
F15	5		L15	6	
F16	5		L16	6	
F17	6	6	L17	6	6
F18	6	6	L18	6	6
F19	6	6	L19	6	6
F20	5		L20	6	
F21	5		L21	6	
F22	6	6	L22	6	6
F23	6	6	L23	6	6
F24	6	6	L24	6	6
F25	5		L25	6	
F26	4		L26	6	
F28	6	6	L28	6	6
F29	6	6	L29	6	6
F30	6	6	L30	6	6
F31	6	6	L31	6	6
F32	6	6	L32	6	6
F33	3		L33	6	
F34	6	6	L34	6	6
F35	6	6	L35	6	6
F36	6	6	L36	6	6
F39	6	6	L39	6	6
F40	6	6	L40	6	6
F41	6	6	L41	6	6
n	221	180		234	180
N					815

Table A2. Summary of the HGLM model from which variance components were extracted for the calculation of delta-R values.

Parameter	Mean	95% CI lower	95% CI upper	P_{MCMC}
Intercept	-0.66	-1.21	-0.20	0.01
Size	-0.13	-0.31	0.05	0.15
Period B	2.87	2.17	3.57	0.001
Group LF	0.17	-0.20	0.65	0.41
Observation	0.21	-1.16	0.46	0.21
Period B * Group LF	-0.65	-0.88	0.06	0.11
Observation * Group LF	0.03	-0.06	0.11	0.52
Observation * Period B	-0.39	-0.48	-0.31	0.001
Observation * Period B * Group LF	0.01	-0.13	.14	0.88
G-structure (V_{BI})				
FL.A	0.27	0.13	0.42	
FL.B	0.40	0.22	0.66	
LL.A	0.28	0.14	0.45	
LL.B	0.40	0.19	0.65	
R-structure (V_{WI})				
FL.A	0.59	0.46	0.70	
FL.B	0.53	0.42	0.66	
LL.A	0.90	0.71	1.08	
LL.B	0.60	0.48	0.75	

Note that this summary shows posterior mean values.

Table A3. Summary of the HGLM used to analyse the effects of environmental variables on startle response duration.

Parameter	Mean	95% CI lower	95% CI upper	P_{MCMC}
Intercept	1.95	-4.54	8.99	0.56
Size	0.14	-0.07	0.37	0.18
Observation	0.23	1.16	0.29	<0.001
Temperature	-0.02	-0.09	0.04	0.50
DO_2	0.03	-0.02	0.06	0.18
pH	-0.11	-0.40	0.15	0.45
Salinity	-0.06	-0.23	0.13	0.56
G-structure (V_{BI})	0.67	0.34	0.94	
R-structure (V_{WI})	0.69	0.56	0.86	

Data are for anemones in the FL group during period A of the experiment. Note that this summary shows posterior mean values.

APPENDIX

Details of the modelling conditions for the DHGLM

Following the usual MCMC setup, the parameters in each model were updated conditional on the remaining parameters to generate random draws from their posterior distribution. The standard deviations of the random effects and error terms in both the mean and SD models were assigned weakly informative scaled half-t prior distribution with 3 df [59] while the fixed effects parameters were assigned non-informative normal prior distributions. To aid convergence the response variable was scaled as follows. If Y_{ij} denotes the startle response of the i^{th} crab on the j^{th} occasion then the standardised startle response = $Y_{ij} - (\sum Y_{ij}/N_{ij}) / \sigma_{ij}$, i.e. each startle response is subtracted from the mean of startle responses and divided by the standard deviation of startle responses. Three chains were run in parallel so that convergence could be assessed and each chain was run with an adaptive phase of 5000 iterations and a sampling phase of 15000 iterations. We made inferences about the parameters in each model based on their posterior means and 95% credible intervals. As in previous studies, we based the primary assessment of the significance of each predictor on whether or not the 95% credible intervals for the corresponding parameter covered zero. In the case of fixed effects we are also able to judge significance by generating values analogous to classical P-values. These pseudo P-values (hereafter 'P') are obtained by calculating the tail probability for each fixed parameter. They express, as a value between 0 and 1, the probability over the set of all equal tailed credible intervals that cover zero. These values cannot be calculated for random effects as they are constrained to values > 0. Convergence across chains was assessed using the Gelman-Rubin diagnostic, which was <1.1 for each model parameter, indicating that the adaptive phase was adequate.

Details of the modelling conditions for HGLMs

The random effects and residual variances were assigned weakly informative inverse Wishart prior distributions. We also re-ran the analyses with flat priors, which yielded qualitatively identical results in each case. MCMCglmm allows a single chain, so to aid model convergence, the chain was run with an initial adaptive phase of 300000 iterations and a sampling phase of 1000000 iterations, with a thinning interval of 1000. Response data were scaled as above. We made inferences about the parameters in each model based on their posterior means and 95% credible intervals. As in previous studies, we based the primary assessment of the significance of each predictor on whether or not the 95% credible intervals for the corresponding parameter covered zero, but in this case MCMC based P values are also available. To assess convergence of each HGLM model we calculated the autocorrelation factors (ACF) for each effect and each specific variance component, which in each case were <0.1 (range = 0.0009 to 0.054), indicating that the models had converged adequately.

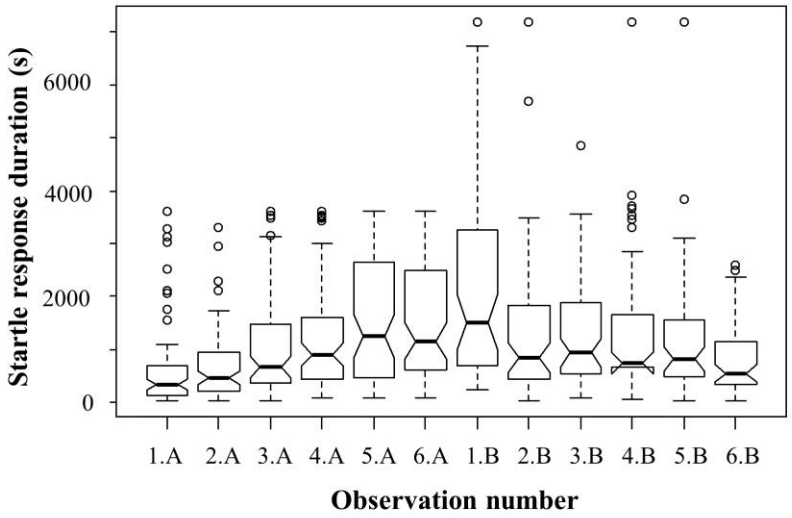

Figure 1. The significant interaction between observation number and period on the duration of startle responses. Thick horizontal bars show the median, boxes show the interquartile range (IQR) from first to third quartiles, whiskers show the nominal range of the data (maximum and minimum values that are within 1.5 x IQR) and data falling outside the nominal range are shown as dots. Notches indicate the 95% CIs of the median. Analyses were performed on standardised Log10 transformed data, but raw data are illustrated here.

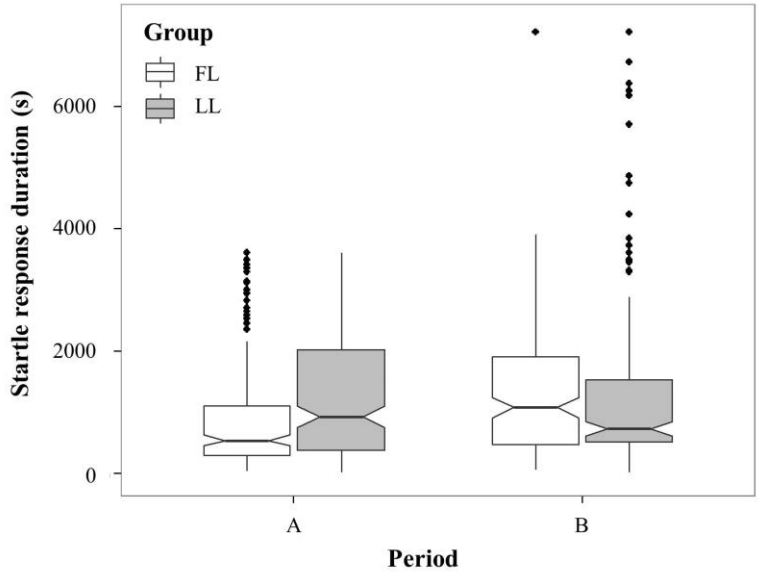

Figure 2. The significant interaction effect between treatment group and period on the duration of startle responses. Thick horizontal bars show the median, boxes show the interquartile range (IQR) from first to third quartiles, whiskers show the nominal range of the data (maximum and minimum values that are within 1.5 x IQR) and data falling outside the nominal range are shown as dots. Notches indicate the 95% CIs of the median. Analyses were performed on standardised Log_{10} transformed data, but raw data are illustrated here.

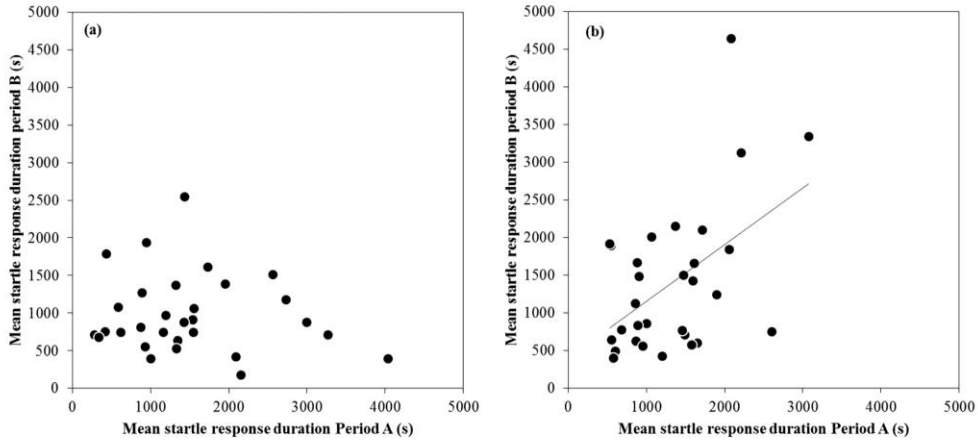

Figure 3. Among individual correlations between mean startle response durations in periods A and B of the experiment, for (a) the FL group and (b) the LL group. The correlation was positive for the LL group only (regression line added for illustration).