



Sampling biases and reproducibility: experimental design decisions affect behavioural responses in hermit crabs

Anna Hills^{a, b}, Mike M. Webster^{b, *} 

^a School of Biological Sciences, University of Manchester, Manchester, U.K.

^b Centre for Biological Diversity, School of Biology, University of St Andrews, St Andrews, U.K.

ARTICLE INFO

Article history:

Received 29 March 2022
Initial acceptance 3 May 2022
Final acceptance 22 August 2022

MS. number: 22-00166R

Keywords:

boldness
enrichment
neophobia
replication
reproducibility
STRANGE

How important are sampling and experimental design decisions in shaping test subject behaviour under laboratory conditions? We examined the effects of circatidal rhythm, time held in captivity, sampling location (open or covered areas of habitat), acclimation period and water depth on activity and emergence latency in hermit crabs (*Pagurus bernhardus*). We found that subjects held in captivity for 1 month and those collected from the open were faster to emerge from their shells after disturbance compared, respectively, to those tested after 1 day in captivity or collected from beneath cover. We also found that subjects tested after shorter acclimation periods were more active than those tested following longer acclimation periods. Our findings reveal that sampling and study design decisions can have pronounced influences on subject behaviour measured under otherwise common conditions, with potentially important implications for interpretation and reproducibility of findings. As researchers we should take care to explicitly consider how sampling biases and effects arising from our experimental protocols might affect the behavioural responses of test subjects. Doing so can help us make more reasonable generalizations beyond our subject pool, draw better-informed comparisons between studies and achieve greater reproducibility of findings.

© 2022 The Author(s). Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

When we design experimental investigations of animal behaviour, we make decisions over factors such as which subjects to test, when to test them and how to house them prior to and during the experiments. How important are these decisions in influencing the behaviour exhibited by subjects when tested? If the subject pool is not representative of the wider population that the researchers are seeking to understand, or if the environments that test subjects experience prior to and during testing shape their behaviour in ways that are not accounted for, then this can be problematic, with implications for the interpretation, comparison and replication of findings. Webster and Rutz (2020) created the STRANGE framework to highlight this issue and to help researchers identify, mitigate and report such biases. The acronym STRANGE refers to test subjects' social background (S), trappability (T), rearing history (R), acclimation and habituation (A), natural changes in responsiveness (N), genetic make-up (G) and experience (E). Variation among test subjects with respect to these factors is not necessarily problematic: in fact, these are often the focus of well-designed experiments, or else are explicitly

controlled for. In cases where these factors are not accounted for, or are not clearly reported, they may constitute artefacts, affecting test subject behaviour in unforeseen or unknown ways.

The procurement of animals for use in experiments may be the first opportunity for sampling bias to occur. Many research projects test subjects collected from the wild, using a variety of passive and active trapping methods. Numerous studies have revealed that animals with particular behavioural or personality traits may be more likely to enter traps than others (e.g. Carter et al., 2012; Garamszegi et al., 2009; Wilson et al., 1993; Álvarez-Quintero et al., 2021). Where biased traps are used to collect subjects for behavioural studies, animals with particular attributes may end up being over-represented within the subject pool (Álvarez-Quintero et al., 2021; Kressler et al., 2021; Wilson et al., 1993), giving a skewed estimate of the behaviour of the wider population. Where subjects are allowed to freely participate in experiments, for example by interacting with experimental apparatus, and animals with particular personality traits are more likely to engage with experimental equipment, then self-selection effects, analogous to trapping biases, can apply (Morton et al., 2013).

The housing of subjects prior to testing can also influence how they behave when tested. Physical enrichment, the degree of

* Corresponding author.

E-mail address: mmw1@st-andrews.ac.uk (M. M. Webster).

stimulation an animal receives in its environment, is known to affect brain development and subsequent behaviour (van Praag et al., 2000). This has been documented in a range of species and behavioural contexts. Greater degrees of environmental enrichment are associated with more optimistic response biases in starlings (*Sturnus vulgaris*; Matheson et al., 2008), enhanced spatial memory in mice, *Mus musculus* (Frick & Fernandez, 2003), reduced latency to complete a cognitive test in rattlesnakes, *Elaphe obsoleta* (Almli & Burghardt, 2006) and increased aggression in zebrafish, *Danio rerio* (Woodward et al., 2019). Similarly, the social environment experienced by animals can shape subsequent behaviour. Canaries, *Serinus canaria*, that had been housed in pairs approached an ambiguous cue associated with food palatability sooner than those held alone (Lalot et al., 2017); cranes, *Grus americana* (Kreger et al., 2004) reared by their parents were more vigilant than those reared by humans; male guppies, *Poecilia reticulata*, raised in male-biased sex ratio groups used coercive mating tactics more frequently than those raised in female-biased and equal sex ratio group compositions (Evans & Magurran, 1999); and pheasants, *Phasianus colchicus* housed in groups of five outperformed those raised in groups of three in two different spatial discrimination tasks (Langley et al., 2018).

While test subject origins and experience prior to testing can affect their behaviour at test, these effects can be moderated further by the design of study protocols (Webster & Rutz, 2020). The timing of testing can be important, since circadian rhythms, photoperiod and circatidal rhythms can affect behaviours, from activity to aggregation (Imafuku, 1981; Saunders, 1997; Simon et al., 2012; Turra & Denadai, 2003). Habituation to the captive environment, or, conversely, chronic stress associated with captivity, and acclimation to experimental conditions can also shape behavioural responses (Adams et al., 2011; Butler et al., 2006; O'Neill et al., 2018), meaning that decisions over how long to house animals before running experiments and the duration of settling periods prior to commencing observations can also play a role in shaping experimental outcomes. Unpicking the effects of these different sources of behavioural variation, from sampling biases, to test subject experience and artefacts of experimental protocols, presents a challenge for researchers.

In this study we investigated how sampling and experimental design decisions affected the behaviour of hermit crabs (*Pagurus bernhardus*) tested under standardized conditions in the laboratory. Our aim was to explore the effects of multiple sources of potential bias in a single study system. We focused on two behaviours, a measure of activity and the time taken by hermit crabs to emerge from their shells after being disturbed. Across three factorially designed experiments, we investigated effects of time in captivity before testing, the scheduling of testing relative to tidal cycle (we tested smaller hermit crabs from the intertidal habitat), the depth of water in the testing arena, the microhabitat from which the hermit crabs were collected and the amount of time they were given to acclimate to the experimental setting before being tested. Four of these factors correspond to T, A and N components of STRANGE (Webster & Rutz, 2020): microhabitat of origin corresponds to trappability (T); self-selection, time in captivity and acclimation to the experimental setting correspond to acclimation and habituation (A); testing relative to tidal cycle corresponds to natural changes in responsiveness (N). Varying water depth falls outside the STRANGE framework but is an example of an experimental design decision that may also affect behavioural responses.

Our study consisted of three experiments. In experiment 1, we investigated the effects of circatidal rhythm and time held in captivity. Subject were held in the laboratory for either 1 day or 28 days before testing, with trials performed at times corresponding to low, mid or high tide. We predicted that hermit crabs would be less

active and would take longer to emerge from their shells at times corresponding to low tide (Imafuku, 1981; Turra & Denadai, 2003), and that this effect would be diminished in those held in captivity for longer. We predicted that across all tidal stages, activity levels would be higher and emergence times lower in the hermit crabs that had been in the laboratory for longer and that had had more time to adjust to laboratory conditions. Experiment 2 also measured these behaviours at times corresponding to low, mid or high tide, with testing taking place in shallow or deeper water. In experiment 2, all subjects were tested after 24 h in captivity. In addition to the circatidal effects on behaviour predicted for experiment 1, we further predicted that activity levels would be higher and emergence times lower in hermit crabs tested in deeper water, reasoning that deeper water provides refuge from predatory birds such as carrion crows, *Corvus corone*, which have been observed to feed on this particular population, with foraging depth likely limited by bill length. Finally, in experiment 3, we investigated the effects of sampling origin and differing holding periods prior to release into the testing arena on behaviour. We compared hermit crabs that were sampled from open areas or sandy or bare rock substrate with those that were collected from beneath flat rocks or seaweed cover. We predicted that subjects collected from open areas would be more active and would emerge sooner than those collected from cover, reasoning that bolder individuals should spend more time in the open and therefore be over-represented in open-captured samples. We also predicted that activity levels would be greater and emerge times lower when holding periods prior to testing were greater, owing to subjects having longer to acclimate to testing conditions.

METHODS

Model Organism

Pagurus bernhardus is a marine hermit crab that carries gastropod shells for protection, retreating within them when threatened. They are omnivorous detritivores, actively searching the substrate for dead animals and sifting organic particles from the sand (Ramsay et al., 1997). They can readily be studied in the laboratory and have been used as model organisms for investigating resource competition (e.g. Dowds & Elwood, 1983) and personality (e.g. Briffa et al., 2008; Gorman et al., 2018), among other behaviours. We selected *P. bernhardus* as our model system both because they are an established model and because they are locally abundant. We wish to emphasize that while we are using them to explore sources of sampling bias in animal behaviour studies, this should not be taken as a criticism of existing research using this and other hermit crab species. There is a sizeable literature on hermit crab behavioural ecology containing rigorously designed and analysed experiments, and researchers already take into account factors such as acclimation and habituation effects by conducting repeated measures on individuals, by testing in the field as well as in captivity and by sampling widely from different areas of habitat (e.g. Briffa, 2013; Briffa & Bibost, 2009; Briffa et al., 2008). Thus, the approaches outlined in our study apply to animal behaviour research generally and not to research using hermit crabs in particular.

Animal Collection and Housing

Hermit crabs were collected from rockpools at East Sands, St Andrews, Fife, U.K. and transported in covered buckets of sea water for approximately 20 min by foot, to our laboratory at the University of St Andrews. Collection and testing took place between November 2020 and May 2021. During this time, multiple batches

of 50–100 hermit crabs were collected, tested and released. In total, 1388 hermit crabs were collected, of which four were excluded after testing (detailed in experiment 1, below), leaving 1384 subjects. Collection occurred at low tide or on falling tides from rockpools. The rockpools were typically 5–30 cm deep and contained exposed rock, loose boulders, sand and macroalgae. Hermit crabs were collected by hand, and we were careful to collect hermit crabs from all areas of habitat within the rockpools except when collecting them for experiment 3, where we specifically targeted hermit crabs from open and sheltered areas, as described below. Only those inhabiting common periwinkle, *Littorina littorea*, shells with a height of 1.7–3.5 cm were collected. We estimate that >90% of the hermit crabs at this location occupied common periwinkle shells. The mean (\pm SD) shell height and aperture size of hermit crabs were 2.67 ± 0.25 cm and 1.47 ± 0.14 cm, respectively. In the laboratory, hermit crabs were placed in holding tanks (50 × 50 cm and 20 cm tall). These contained artificial sea water (salinity of 1.025, prepared using Instant Ocean® brand aquarium salt) kept at a constant 9–11 °C. Each tank contained a 2.5 cm deep layer of coarse aquarium gravel, a brick with three cores placed on its side, two artificial aquarium plants and an air-powered sponge filter. Most hermit crabs were tested 1 day after arrival, then released. In experiment 1, one treatment group was held for 28 days prior to testing and release. Individuals that were tested after 1 day were fed only on the day of arrival. Those held for 28 days were fed every 2 days, and always 20–24 h prior to testing, so that subjects in both conditions had fed the day before testing. They were fed with crustacean food pellets (Tetra Crusta Menu complete crustacean food, Tetra, Melle, Germany). Individual hermit crabs were never tested twice and only took part in either an activity level trial or an emergence trial. After testing, they were returned to the sea, in a different location approximately 600 m from where new hermit crabs were collected.

Behavioural Tests

We performed three experiments, each of which involved two tests, one of activity and one of emergence time following a disturbance. The tests broadly followed the same design, with modifications specific to each experiment. Below, we first describe the tests, then the experiments. Note that this work was performed during the global COVID-19 pandemic. Midway through the study (January 2021) the U.K. entered a period of more restrictive lockdown measures. This limited the number of persons allowed in our shared laboratory facilities at any given time and, as consequence of this, we had to alter the running times of some experiments and reduce the number of trials in some treatments. Such instances are noted below.

Activity level test

Trials were performed in batches of nine, using an array of semitransparent, pale grey plastic boxes (20 × 20 cm and 17 cm tall; Backpack brand, TakeawaySupplies.co.uk), arranged in a 3 × 3 grid, with one subject tested in each. The boxes were separated by black partitions to visually isolate subjects from one another, and all nine were screened from outside disturbance by being placed within a larger opaque circular container. Beneath each box we placed a 6 × 6 square grid. This was visible from above through the semitransparent base of the box and filled its floor. Artificial sea water was added to each box to a depth of 15 cm (but see exception in experiment 2) and was continuously aerated using airstones and an external air pump, except during the experimental phase when the airstone was removed so as not to disturb the hermit crabs. During each trial, a single hermit crab was placed

in each box (shell aperture facing down). A photography tent was placed around the larger outer container holding the nine test boxes to reduce surface light reflection and disturbance. A tripod with a horizontal boom and a GoPro Hero 5 camera (GoPro, Inc., San Mateo, CA, U.S.A.) attached was inserted through a small hold cut into the photography tent, positioned to film the subjects from above. The subjects were allowed an acclimation period of 30 min before filming began (but see exception in experiment 3), then filmed continuously for 60 min (experiment 1) or 30 min (experiments 2 and 3). We reduced the duration of the experimental period for the second and third experiments, relative to experiment 1, because institutional COVID-19 restrictions were put in place during the lockdown in early 2021, which restricted the amount of time we could spend in the laboratory. Between trials, the water was changed and aerated to ensure no chemical cues remained that may have influenced the next subject's behaviour. After testing, used subjects were placed into a separate aquarium where they were held until they could be released. From the videos of these trials, we recorded the amount of time spent moving to the nearest second and the number of times the subject crossed between squares on the grid. Where subjects were emerged from their shells but not walking, they were not counted as moving. These behaviours were found to be strongly positively correlated, and we decided to use only number of squares crossed as our measure of activity in the analyses described below.

Emergence test

Trials were performed in batches of eight, with each subject tested within its own plastic container. In experiments 1 and 3, these measured 15 × 6.2 cm and 7.6 cm tall. These were transparent, but all sides except the 15 cm front-facing wall were painted externally with matt black acrylic paint to minimize outside disturbance. In experiment 2, where we varied water depth, we used slightly larger containers, measuring 20 × 20 cm and 17 cm tall. These containers were slightly opaque, but the crabs were visible through the walls of the container. As with the smaller containers, all sides except the front-facing wall were painted externally with matt black acrylic paint. Each was filled with 1 cm of coarse aquarium gravel and was filled with artificial sea water up to 1 cm from the top. A LogitechHD webcam C920 (Logitech International S.A., Lausanne, Switzerland) mounted on a tripod and connected to a laptop (Acer Aspire One, Acer Inc., New Taipei City, Taiwan) was used to film the trials. The webcam was positioned in front of the subject's container and filmed through the unpainted front wall. One subject was placed (shell aperture facing down) individually on the substrate into each container. Hermit crabs were then allowed 10 min to acclimate (but see exception in experiment 3). During testing, each hermit crab was lifted out of the water, turned over with their aperture facing upwards and held in the air for 5 s. This caused the hermit crab to withdraw into its shell. The hermit crab was then lowered into the water, still in an inverted position (aperture facing up), and placed on the substrate. The time taken to the nearest second for the hermit crab to emerge from its shell and completely upright itself after the researcher had stopped handling it was determined from the video. This method presents a standardized disturbance stimulus, which has been used and verified by other researchers (Briffa et al., 2008). Any subject that took >15 min to emerge from its shell after the startle response was removed from the analysis ($N = 4$, all in experiment 1). Between trials, the water was changed and aerated and the gravel was thoroughly rinsed to ensure no chemical cues remained that could influence the next subject's behaviour. After testing, used subjects were placed into a separate aquarium where they were held until

they could be released. These trials were carried out at the same time as the activity levels trials, described above.

Experiments

Experiment 1: effects of circatidal rhythm and time held in laboratory before testing

This experiment investigated whether activity and emergence times are linked to circatidal rhythm in hermit crabs, and whether any such circatidal effects are diminished by time spent in captivity. Testing took place at times that corresponded to local low, mid or high tide, with trials performed between November 2020 and February 2021. Subjects were either tested after 24 h in captivity (i.e. 1 day after collection), or tested after 28 days under laboratory conditions, as described above. The activity and emergence trials were performed as described above. A total of 614 hermit crabs were tested during this experiment. Of these, four were excluded from 24 h high tide emergence condition because they failed to re-emerge after 15 min. This left 610 crabs in the analyses, divided between treatments as follows: activity after 24 h in captivity (low, mid and high tide) = 63, 72, 54 subjects; activity after 28 days in captivity (low, mid and high tide) = 18, 18, 18 subjects; emergence after 24 h in captivity (low, mid and high tide) = 97, 106, 92 subjects; emergence after 28 days in captivity (low, mid and high tide) = 24, 24, 24 subjects. Sample sizes were uneven between low, mid and high tide treatments because access to the laboratory was limited by a rota system during the COVID-19 pandemic that capped the number of users allowed inside the laboratory and restricted access times to 0900–1700 hours. Times of high and low tide sometimes fell outside this period, preventing us from testing subjects at these times. Trial numbers were lower for the 28-day captivity treatments than for the 24 h treatments because, midway through the study period, the U.K. entered a more restrictive second lockdown and laboratory access was further limited. These restrictions also affected experiments 2 and 3.

Experiment 2: effects of circatidal rhythm and water depth

We investigated the effects of circatidal rhythm and water depth in the testing arena on activity and emergence times in hermit crabs. Trials took place at times corresponding to local low, mid or high tide, as in experiment 1, with trials performed over a 5-week period between February and April 2021. All subjects were tested the day after capture. We used water depths of 5 cm or 15 cm in both the activity and emergence trials, with subjects randomly allocated to each. We reasoned that water depth might affect the hermit crabs' perception of risk. The hermit crabs used in this study were collected from rockpools varying between 5 cm and 30 cm deep, so that our treatment conditions correspond to shallow and intermediate depths. A total of 234 hermit crabs were tested for this experiment, with no subjects excluded. Subjects were divided between treatment groups as follows: activity (5 cm: low: $N = 19$; mid: $N = 38$; high: $N = 25$; 15 cm: low: $N = 17$; mid: $N = 34$; high: $N = 29$); emergence (5 cm: low: $N = 12$; mid: $N = 14$; high: $N = 10$; 15 cm: low: $N = 12$; mid: $N = 14$; high: $N = 10$).

Experiment 3: effect of sampling location and testing arena acclimation period

Finally, we investigated whether sampling origin and duration of acclimation time in the experimental arena before testing influenced the activity and emergence response time of hermit crabs. Sampling origin refers to the type of habitat within the rockpools from which the hermit crabs were collected, which we designated as open or covered. Open areas were flat expanses of sand or bedrock without physical cover and in which hermit crabs were readily visible on the surface. Covered areas were overlain by flat rocks,

macroalgae or overhanging crevices, and the hermit crabs were collected from beneath these. Hermit crabs were collected from open areas >1 m from cover and were picked up directly from the substrate. We collected hermit crabs from cover by carefully lifting large rocks, sweeping aside seaweed or by reaching into cracks. We collected hermit crabs from both habitat types on each day of collection. Subjects were transported in separately labelled buckets to the laboratory and were not mixed in the holding aquaria. Trials took place over 4 weeks between April and May 2021. We conducted a 2×4 factorial design, testing subjects collected from open and covered habitats in the activity and emergence assays as described above, but this time using four different acclimation periods of 5, 10, 30 or 60 min. Because we found no evidence of circatidal effects on behaviour in experiments 1 and 2, we did not balance trials across the tidal cycle here. All subjects were tested the day after collection. We tested a total of 544 subjects. In the activity test, there were 36 subjects in each of the eight sampling origin and acclimation time combination groups, and in the emergence test, there were 32 subjects per treatment combination.

Statistical Analyses

We performed analyses using Rstudio version 1.3.959 (R Core Team, 2021). Each of the three experiments involved measures of activity and emergence time, and general linear models (GLM, specifically two-way ANOVAs) were used to analyse each measure in each of the three experiments. The two measures of activity (time spent moving, number of squares crossed) were strongly positively correlated (Pearson product moment correlations: experiment 1: $N = 244$, $R^2 = 0.94$, $P < 0.001$; experiment 2: $N = 162$, $R^2 = 0.95$, $P < 0.001$; experiment 3: $N = 288$, $R^2 = 0.94$, $P < 0.001$). For this reason, we decided to use only one of the measures, number of squares crossed, as our measure of activity. Emergence times and number of squares crossed were not normally distributed and were normalized through log transformation and $(x + 1)$ log transformation, respectively. The normalized emergence time data and numbers of squares crossed were used as dependent variables in the GLMs. For experiment 1, time spent in the laboratory before testing (24 h or 28 days) and stage of tidal cycle (low, mid or high) were included as fixed factors. In experiment 2, water depth (5 or 15 cm) and stage of tidal cycle were included as fixed factors. In experiment 3, the fixed factors were sampling origin (collected from open or covered areas) and acclimation time prior to testing (5, 10, 30 or 60 min). For each model, interactions between fixed factors were included.

Ethical Note

The study species used in this project is not covered by U.K. ASPA regulations and no ethical approval was required or sought. No animals became obviously ill during this experiment and none died. Four trials were excluded (see above, experiment 1) due to subjects failing to move during the experiment. All animals were held in captivity for the minimum time necessary to achieve the experimental objectives and all were released to the wild after the study as described above.

RESULTS

Experiment 1: Effects of Circatidal Rhythm and Time Held in Laboratory Before Testing

Hermit crabs that had been held in the laboratory for 28 days prior to testing emerged sooner from their shells after being disturbed than did those tested 1 day after capture. We saw no effect of tidal stage on emergence times, nor any interaction between tidal stage and time in the laboratory prior to testing (GLM:

Levene statistic = 0.77, $P = 0.57$; time in laboratory: $F_{1,361} = 59.01$, $P < 0.001$, $\eta^2 = 0.14$; tidal stage: $F_{2,361} = 1.67$, $P = 0.19$, $\eta^2 = 0.009$; interaction: $F_{2,361} = 2.09$, $P = 0.12$, $\eta^2 = 0.01$; Fig. 1a). We saw no differences in activity levels between crabs held for 1 day or 28 days, nor any effect of tidal cycle and no interaction between these factors (GLM: Levene statistic = 1.60, $P = 0.16$; time in laboratory: $F_{1,237} = 2.11$, $P = 0.12$; $\eta^2 = 0.02$, tidal stage: $F_{2,237} = 1.16$, $P = 0.28$; $\eta^2 = 0.005$, interaction: $F_{2,237} = 0.02$, $P = 0.97$, $\eta^2 = 0.001$; Fig. 1b).

Experiment 2: Effects of Circatidal Rhythm and Water Depth

Neither tidal stage nor water depth in the testing arena affected hermit crab emergence times or activity (GLM: emergence time: Levene statistic = 0.98, $P = 0.44$; water depth: $F_{1,66} = 0.22$, $P = 0.64$, $\eta^2 = 0.003$; tidal stage: $F_{2,66} = 0.19$, $P = 0.82$, $\eta^2 = 0.006$; interaction: $F_{2,66} = 2.04$, $P = 0.14$, $\eta^2 = 0.06$, Fig. 2a; activity: Levene statistic = 0.16, $P = 0.66$; water depth: $F_{1,156} = 0.09$, $P = 0.78$, $\eta^2 = 0.001$; tidal stage: $F_{2,156} = 0.28$, $P = 0.73$, $\eta^2 = 0.004$; interaction: $F_{2,156} = 0.79$, $P = 0.45$, $\eta^2 = 0.01$; Fig. 2b).

Experiment 3: Effect of Sampling Location and Testing Arena Acclimation Period

Hermit crabs that were collected from open areas of habitat emerged sooner following a disturbance than did those collected from covered areas. Acclimation time in the experimental arena had no effect on emergence time and did not interact with the type of habitat from which they were collected (GLM: Levene statistic = 1.84, $P = 0.06$; sampling origin: $F_{1,248} = 12.89$, $P < 0.001$, $\eta^2 = 0.06$; acclimation period: $F_{2,248} = 1.01$, $P = 0.39$, $\eta^2 = 0.01$; interaction: $F_{2,248} = 0.71$, $P = 0.54$, $\eta^2 = 0.01$; Fig. 3a). Activity levels were not affected by sampling origin but did vary with acclimation time. Activity was lower after a longer acclimation period. There was no interaction between these factors (GLM: Levene statistic = 1.42, $P = 0.19$; sampling origin: $F_{1,287} = 1.36$, $P = 0.143$, $\eta^2 = 0.006$; acclimation period: $F_{2,287} = 4.61$, $P = 0.004$, $\eta^2 = 0.05$; Tukey HSD post hoc tests: acclimation time: 5 min versus 60 min: $P = 0.04$; 10 min versus 60 min: $P = 0.048$; 30 min versus 60 min: $P = 0.025$; interaction: $F_{2,287} = 0.28$, $P = 0.84$, $\eta^2 = 0.03$; Fig. 3b).

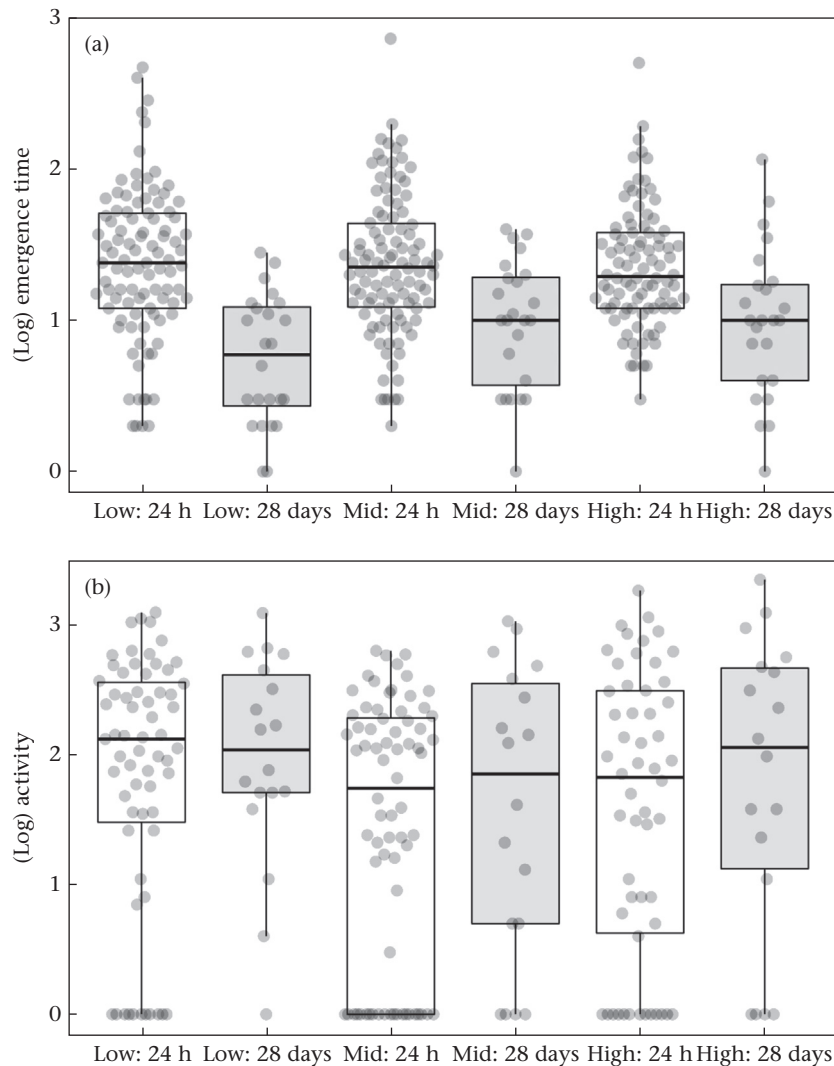


Figure 1. Results from experiment 1, investigating the effects of time in captivity (24 h or 28 days, white and grey bars, respectively) and tidal cycle at time of testing (low, mid or high tide) on (a) latency to emerge from shell after disturbance and (b) activity, the number of squares crossed on the arena floor. Box plots show the median, interquartile range and 95% confidence intervals and the points display the raw data. This figure was produced using the 'PlotsOfData' app (Postma & Goedhart, 2019).

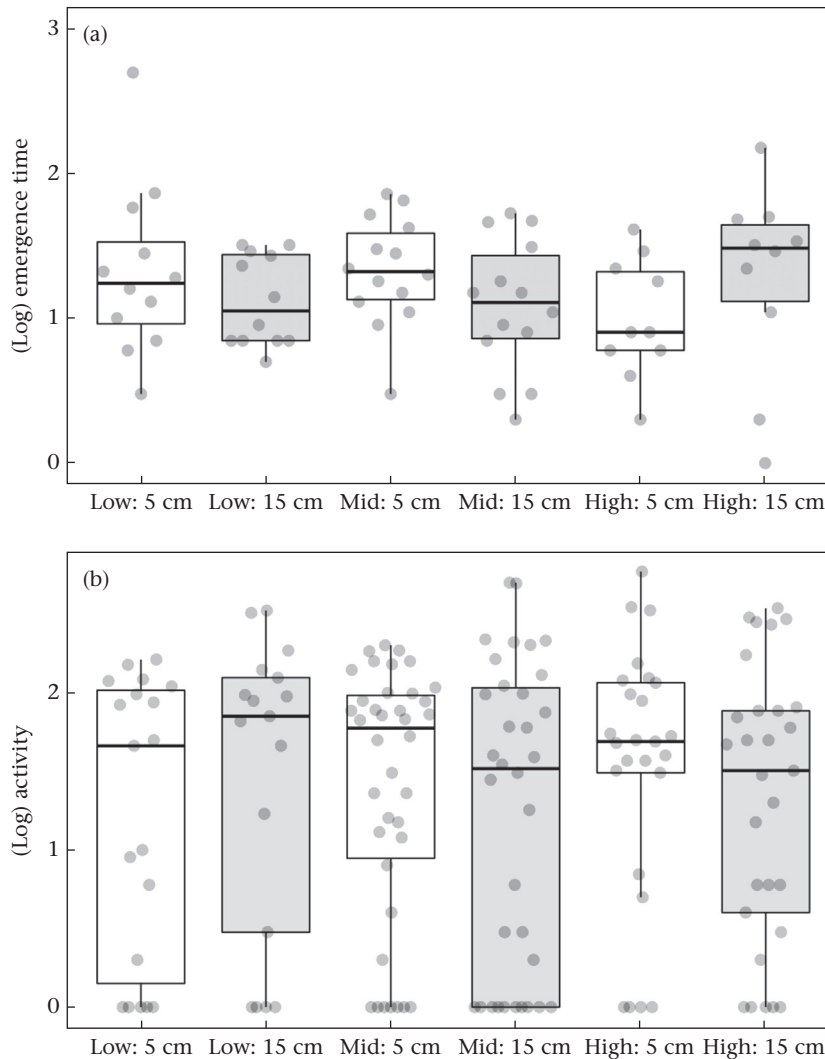


Figure 2. Results from experiment 2, investigating the effects of water depth in the testing arena (5 or 15 cm, white and grey bars, respectively) and tidal cycle at time of testing (low, mid or high tide) on (a) latency to emerge from shell after disturbance and (b) activity, the number of squares crossed on the arena floor. Box plots show the median, interquartile range and 95% confidence intervals and the points display the raw data. This figure was produced using the 'PlotsOfData' app (Postma & Goedhart, 2019).

DISCUSSION

In this project we asked how important sampling and experimental design decisions are in shaping test subject behaviour measured under controlled conditions. Our results suggest that simple sampling and experimental design decisions can have pronounced effects on the behaviours observed in the laboratory, with implications for replication, comparison of findings between studies and extrapolation from effects seen in captive settings back to natural conditions.

First, we found that the amount of time hermit crabs were held in captivity for prior to testing affected their latency to re-emerge from their shells after being disturbed, with those held in captivity for 28 days emerging sooner than those tested 1 day after capture. We saw no effect of time in captivity on activity levels, which remained broadly similar after 1 day and 28 days in captivity. Second, we found that the site of collection affected re-emergence latency in the laboratory. We compared subjects collected from adjacent areas of microhabitat, metres apart but within the same rockpools, which were either in open expanses of sand or bedrock substrate, or in covered areas, beneath large flat rocks, overhanging

ledges and macroalgae patches. We saw that subjects collected from open areas re-emerged sooner compared to those collected from covered patches. Third, when investigating sampling origin, we also varied acclimation time to the experimental areas, and this was seen to be related to activity, with activity levels decreasing after longer periods of acclimation prior to testing. Acclimation period was not related to re-emergence time.

We found that the timing of testing relative to the tidal cycle was unimportant: over two separate experiments, we saw no evidence that test subject activity or emergence times varied as a function of tidal stage. In contrast, the hermit crab *Pagurus geminus* was found to be most active in the laboratory at times that corresponded to high tide at their site of collection (Imafuku, 1981). Turra and Denadai (2003) reported possible circatidal effects on behaviour in the hermit crab *Clibanarius scolopetarius*, but not in *Pagurus criniticornis*, *Clibanarius antillensis* or *Clibanarius vittatus*, where circadian activity with pronounced diurnal and nocturnal behavioural variation was seen. We also saw no effects of water depth on behaviour in our study. We had reasoned that water depth might affect perception of risk, leading subjects to be less active in shallower water and to take longer to emerge. It is possible that the

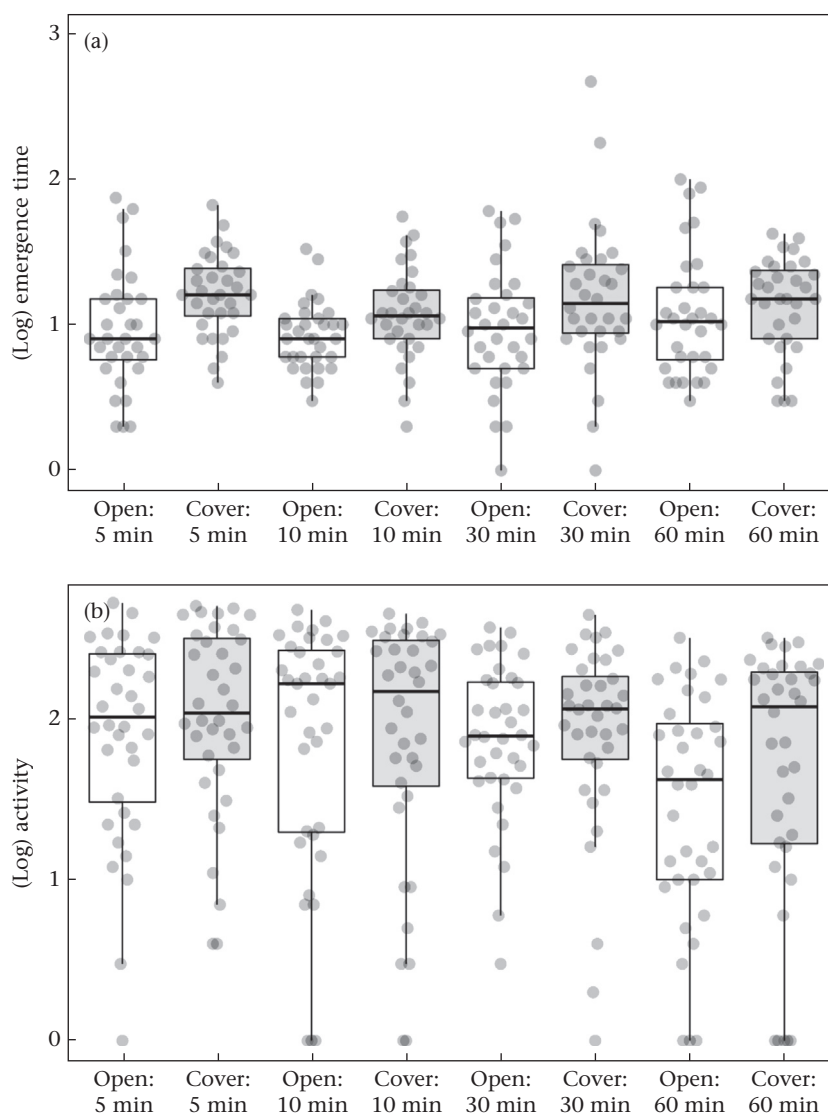


Figure 3. Results from experiment 3, investigating the effects of collection site habitat (open or covered, white and grey bars, respectively) and acclimation time to the experimental arena (5, 10, 30 or 60 min) on (a) latency to emerge from shell after disturbance and (b) activity, the number of squares crossed on the arena floor. Box plots show the median, interquartile range and 95% confidence intervals and the points display the raw data. This figure was produced using the 'PlotsOfData' app (Postma & Goedhart, 2019).

water depths that we used were not great enough to observe effects on behaviour, although they were representative of the range of depths in which the hermit crabs occurred at the site of collection at low tide.

Hermit crabs held in captivity for 28 days re-emerged from their shells sooner after being startled than those tested 1 day after capture. We suggest that this may reflect habituation to the laboratory environment. Note that many studies examining hermit crab behaviour account for time-in-captivity effects, for example by conducting repeated tests over the captive period that allow for changes in behaviour to be quantified (e.g. Briffa, 2013; Briffa et al., 2008). Other studies have also reported time-in-captivity effects on behavioural responses in different species. Butler et al. (2006) found that the likelihood of chaffinches, *Fringilla coelebs*, performing foraging behaviour in captive trials decreased with time held in captivity. In European blackbirds, *Turdus merula*, locomotion, maintenance behaviours (preening and beak rubbing) and alarm calls increased, while alert behaviours decreased over the duration of a 20-day period in captivity (Adams et al., 2011). Baseline levels of corticosterone were significantly greater at the

end of the period of captivity than at the beginning, suggesting that confinement may be a chronic stressor in this species. Researchers use a range of holding periods when designing captive experiments for wild animals, and studies such as Butler et al.'s (2006) study demonstrate that short periods of confinement might be better than longer habituation times in some cases, if the animals are more likely to engage in the behaviours of interest. Some researchers have compared behaviours measured in the laboratory to those recorded under natural conditions in the wild (e.g. Fisher et al., 2015; Gilby et al., 2011; Herborn et al., 2010). Validation studies such as these are important, allowing us to check that the behaviours we quantify under controlled conditions actually capture the responses that we seek to understand in wild animals. Useful further work might extend this approach to investigate how the range of time in captivity affects the differences in captive and wild behaviour, to determine whether shorter or longer holding periods are more appropriate for answering particular questions about given behaviours in particular study species.

Our third experiment revealed that hermit crabs that had been allowed to acclimate for longer tended to be less active than those

allowed to acclimate for shorter periods. Heightened activity may reflect a neophobic response to the test environment, an effect of stress that declines with acclimation time. O'Neill et al. (2018) measured activity levels in domestic and feral guppies that were allowed to acclimate for different periods, ranging from minutes to days. They found that for both populations, activity levels were highest in the groups given the shortest acclimation times and lowest for those given intermediate acclimation times, and that activity levels rose again for the longest acclimation times, although without getting as high as those seen for the shortest acclimation times. Ward (2012) reported that activity rates decreased over time in mosquitofish, *Gambusia holbrooki*, exploring a novel arena. This was true for individual fish and for groups of various sizes. In Ward's (2012) study, acclimation time was not varied, but changes in activity were measured across the observation period. Ward (2012) suggested that initial greater activity may reflect a tendency to explore unfamiliar surroundings. A similar tendency may underlie the initially higher levels of activity seen in our test subjects, with the stress of being placed in a novel environment perhaps causing the hermit crabs to seek shelter or escape. Further work is necessary to test this suggestion.

Our third experiment also revealed that hermit crabs collected from open areas tended to re-emerge sooner than those collected from covered areas. Re-emergence times are known to be repeatable within individuals in this species (e.g. Briffa et al., 2008; Mowles et al., 2012), and it is possible that bolder, faster re-emergers are more risk-prone, spending more time in the open where they are visible to predators compared to risk-averse shy hermit crabs. Another (not mutually exclusive) possibility is that state-dependent differences, such as available energy reserves, drive both space use in the wild and emergence latency in the laboratory. These ideas are speculative, and any relationship between microhabitat use and personality or state need to be explored in a dedicated follow-up study using an experimental design that specifically captures individual repeatability of behaviour. We did not see that subjects collected from the open differed in their activity levels compared to those collected from cover. Garcia et al. (2020) documented a relationship between exploration activity, re-emergence latency and microhabitat of origin in another hermit crab, *Clibanarius symmetricus*. They considered four habitat types (sandy substrate, muddy substrate, oyster bank, muddy substrate with roots) and ranked these from highest to lowest predation risk, based on the opportunity to burrow and use physical structures as cover to avoid predators. Garcia et al. (2020) found that hermit crabs from the microhabitat that had muddy substrate with roots were more active than those from the other habitat types, while those from the sandy substrate took longer to emerge than those collected from the muddy substrate with roots.

The most important finding from our third experiment was that biased sampling, only collecting easily seen and easily captured hermit crabs from open areas, has the potential to skew samples, in this case by over-representing the fastest-emerging members of the local population, at least when tested under laboratory conditions. In this study, we collected equal subjects from both locations in all three experiments, only accounting for this as a variable of interest in the third experiment. Even this approach is not ideal, since we did not quantify the proportions of the population typically found in cover versus in the open; a more representative sampling regime would collect subjects in proportion to this distribution. Such sampling biases are well known, potentially affecting which animals are captured for use in studies and which individuals 'self-select' in free-participation experimental designs. For example, in male Namibian rock agamas, *Agama planiceps*, those with shorter flight initiation distances (FID) had a lower latency to enter baited traps and were captured sooner than those

with longer FIDs (Carter et al., 2012). More exploratory male collared flycatchers, *Ficedula albicollis*, and those with shorter FIDs were more likely to enter nestbox traps compared to less exploratory individuals and those with greater FIDs (Garamszegi et al., 2009). More active sticklebacks, *Gasterosteus aculeatus*, were more likely to swim into passive traps in a laboratory study than were less active sticklebacks (Kressler et al., 2021). A field study by Álvarez-Quintero et al. (2021) compared capture of the same species using passive traps and nets, and they found that the fish caught in the passive traps were more risk-prone, more sociable and smaller than those captured using nets. Similarly, pumpkin-seed sunfish, *Lepomis gibbosus*, collected using passive traps behaved differently in captivity than those captured in seine nets, being more likely to accept food in feeding trials (Wilson et al., 1993). An example of self-selection bias comes from capuchin monkeys, *Sapajus apella*. All individuals in a captive population were scored for personality measures and were then allowed to participate in an experimental task. More open and less assertive individuals were more likely to engage with the experiments, demonstrating that personality-related self-selection can influence which individuals opt into free-participation experimental designs (Morton et al., 2013).

Sampling issues such as these are problematic because they hamper our ability to predict natural behavioural responses in the wider population of our study species, because they can impede comparisons between studies that use differing approaches to address similar questions, and because they have the potential to complicate replication efforts. For studies using wild animals or that involve experiments conducted in the field, high-fidelity replication of studies may not always be possible (Nakagawa & Parker, 2015). If factors such as the duration of captivity prior to testing or the habitat from which subjects are collected influences behavioural responses, and if the impact of these factors is unknown to researchers, then the task of understanding why a finding fails to replicate, or separating false positive or negative findings from artefacts of differences in experimental design, becomes even more difficult.

What can we do about this? In principle, it is possible to sample pools of subjects that are as representative of a given population as possible. This might mean adjusting sampling protocols to capture behaviourally diverse individuals or adapting the testing environment to encourage more individuals to participate (Webster & Rutz, 2020). In practice, this may not always be easy – sample composition may be constrained by practical, financial or ethical considerations, while forestalling experimental design factors that influence behaviour may not be possible when these factors are unknown in advance. Webster and Rutz (2020) suggested that researchers declare factors that may lead to sampling biases in their studies and discuss the potential impacts of these when communicating research findings. This applies to the work we present here, too. While we specifically investigated STRANGE (Webster & Rutz, 2020) and design-related sampling biases in behavioural outcomes, we made a number of sampling decisions that may have implications for the behaviours we observed and that need to be declared and discussed. We limited our sample to a particular size range of hermit crabs, and to those occupying only one type of shell (albeit the majority type). Furthermore, we did not consider the size of fit of the shells relative to the hermit crabs. Hermit crabs have a preferred shell size, and those occupying shells of the preferred size are known to emerge sooner compared to those occupying either small-than-preferred shells or naturally chosen shells, which are likely to be nonpreferred due to scarcity or competition (Briffa & Bibost, 2009). We made no attempt to sex our test subjects, and so could not quantify potential sex differences in behaviour. We conducted our experiments only in the Northern

Hemisphere winter–spring months, and we only sampled from one location. Subjects were housed in groups but tested alone, in the absence of any social influences on behaviour (Webster & Ward, 2011). It is possible that larger or smaller hermit crabs may behave differently when tested under the same conditions as our sample, and the behaviours we investigated may vary over the course of the year, or with social context. It is also possible that hermit crabs from other locations or populations may behave differently too, as a result of genetic variation, experience arising from local differences in predation pressure, food or shell resource distribution or selective mortality of particular behavioural or personality types arising from these or other factors. Readers are encouraged to bear these potential limitations in mind when thinking about how far the findings presented here can be generalized to the wider population of hermit crabs and to other populations and species. Clear reporting of potential biases and their impacts can help readers understand the factors shaping behaviour observed under experimental conditions and why findings sometimes differ between otherwise similar studies.

Beyond declaration and discussion, is there anything else can be done to address biases arising from the particulars of experimental design? Some researchers test a range of design variables (e.g. housing duration, acclimation time) upon dependent variables, but performing this on a large scale and accounting for interactions between variables likely exceeds the capacity of most research groups. One solution might be a collective, community level approach to understanding how variation in aspects of experimental design affect response outcomes. Projects such as ManyLabs (Klein et al., 2014), ManyBabies (Byers-Heinlein et al., 2020), ManyDogs (ManyDogs Project et al., 2021) and ManyPrimates (ManyPrimates et al., 2019) exist to allow separate research teams to pool effort and resources into addressing research problems collectively, with overarching aims including enhancing reproducibility, assembling large and diverse subject pools and understanding the importance of variation between laboratories. In principle, similar approaches could be applied to sampling methods and experimental designs too, to drive collaborative investigations of widely used behavioural tests, charting the relationship between variation in task design and behavioural responses and backed by detailed protocols and well-annotated, open data sets. Clearly, this cannot be done for all aspects of experimental design, behavioural responses or test subject species, but for widely deployed tests (such as boldness, neophobia or open field activity assays) and commonly used model organisms, it could be valuable and informative. As primary research outputs proliferate in the field of animal behaviour, such a resource could be a valuable aid for synthesizing and making sense of this body of work.

Author Contributions

A.H. and M.M.W. conceived and designed the experiment and collected the animals in the field. A.H. built the apparatus, conducted the experiments and extracted the data from the videos. A.H. and M.M.W. analysed the data. A.H. produced a report of the project and M.M.W. edited the report to produce the manuscript, in consultation with A.H. M.M.W. acted as project supervisor and provided training and support.

Data Availability

The data set associated with this project is available from Figshare: <https://doi.org/10.6084/m9.figshare.20496930.v1>]

Declaration of Interest

None.

Acknowledgments

This work was completed as part of A.H.'s placement year for her B.Sc. Zoology with Industrial/Professional Experience degree at the University of Manchester. We thank the following people for assistance in the field and laboratory: Heidi Bough, Renee Fair, Jamie Hogarth, Nick Jones and Helen Spence-Jones. We thank Mark Briffa and an anonymous referee for their helpful comments that greatly improved this manuscript. We received no funding for this project.

References

- Adams, N. J., Farnworth, M. J., Rickett, J., Parker, K. A., & Cockrem, J. (2011). Behavioural and corticosterone responses to capture and confinement of wild blackbirds (*Turdus merula*). *Applied Animal Behaviour Science*, *134*, 246–255.
- Almli, L. M., & Burghardt, G. M. (2006). Environmental enrichment alters the behavioral profile of ratsnakes (*Elaphe*). *Journal of Applied Animal Welfare Science*, *9*, 85–109.
- Álvarez-Quintero, N., Chiara, V., & Kim, S. Y. (2021). Trap versus net: Behavioural sampling bias caused by capture method in three-spined sticklebacks. *Behavioural Processes*, *193*, Article 104504.
- Briffa, M. (2013). Plastic proteans: Reduced predictability in the face of predation risk in hermit crabs. *Biology Letters*, *9*, Article 20130592.
- Briffa, M., & Bibost, A. L. (2009). Effects of shell size on behavioural consistency and flexibility in hermit crabs. *Canadian Journal of Zoology*, *87*, 597–603.
- 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*, 1305–1311.
- Butler, S. J., Whittingham, M. J., Quinn, J. L., & Cresswell, W. (2006). Time in captivity, individual differences and foraging behaviour in wild-caught chaffinches. *Behaviour*, *143*, 535–548.
- Byers-Heinlein, K., Bergmann, C., Davies, C., Frank, M. C., Hamlin, J. K., Kline, M., Kominsky, J. F., Kosie, J. E., Lew-Williams, C., Liu, L., & Mastroberardino, M. (2020). Building a collaborative psychological science: Lessons learned from ManyBabies 1. *Canadian Psychology/Psychologie Canadienne*, *61*, 349–363.
- Carter, A., Goldizen, A., & Heinsohn, R. (2012). Personality and plasticity: Temporal behavioural reaction norms in a lizard, the Namibian rock agama. *Animal Behaviour*, *84*, 471–477.
- Dowds, B. M., & Elwood, R. W. (1983). Shell wars: Assessment strategies and the timing of decisions in hermit crab shell fights. *Behaviour*, *85*, 1–24.
- Evans, J. P., & Magurran, A. E. (1999). Male mating behaviour and sperm production characteristics under varying sperm competition risk in guppies. *Animal Behaviour*, *58*, 1001–1006.
- Fisher, D. N., James, A., Rodríguez-Muñoz, R., & Tregenza, T. (2015). 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*, Article 20150708.
- Frick, K. M., & Fernandez, S. M. (2003). Enrichment enhances spatial memory and increases synaptophysin levels in aged female mice. *Neurobiology of Aging*, *24*, 615–626.
- Garamszegi, L. Z., Eens, M., & Török, J. (2009). Behavioural syndromes and trappability in free-living collared flycatchers, *Ficedula albicollis*. *Animal Behaviour*, *77*, 803–812.
- García, F. A. C., Moura, R. R., Ogawa, C. Y., Zanette, L. R. S., Silva, J. R. F., & Rezende, C. F. (2020). Never forget where you came from: Microhabitat of origin influences boldness and exploration in the hermit crab *Clibanarius symmetricus* (Diogenidae). *Journal of Experimental Marine Biology and Ecology*, *527*, Article 151365.
- Gilby, A. J., Mainwaring, M. C., Rollins, L. A., & Griffith, S. C. (2011). Parental care in wild and captive zebra finches: Measuring food delivery to quantify parental effort. *Animal Behaviour*, *81*, 289–295.
- Gorman, D., Ragagnin, M., McCarthy, I. D., & Turra, A. (2018). Risk-taking and risk-avoiding behaviors by hermit crabs across multiple environmental contexts. *Journal of Experimental Marine Biology and Ecology*, *506*, 25–29.
- Herborn, K. A., Macleod, R., Miles, W. T., Schofield, A. N., Alexander, L., & Arnold, K. E. (2010). Personality in captivity reflects personality in the wild. *Animal Behaviour*, *79*, 835–843.
- Imafuku, M. (1981). Tidal activity rhythm of the hermit crab, *Pagurus geminus* McLaughlin. *Publications of the Seto Marine Biological Laboratory*, *26*, 327–336.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahnik, S., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., & Cemalcilar, Z. (2014). Investigating variation in replicability. *Social Psychology*, *45*, 142–152.
- Kreger, M. D., Estevez, I., Hatfield, J. S., & Gee, G. F. (2004). Effects of rearing treatment on the behavior of captive whooping cranes (*Grus americana*). *Applied Animal Behaviour Science*, *89*, 243–261.

- Kressler, M. M., Gerlam, A., Spence-Jones, H., & Webster, M. M. (2021). Passive traps and sampling bias: Social effects and personality affect trap entry by sticklebacks. *Ethology*, *127*, 446–452.
- Lalot, M., Ung, D., Péron, F., d'Ettorre, P., & Bovet, D. (2017). You know what? I'm happy. Cognitive bias is not related to personality but is induced by pair-housing in canaries (*Serinus canaria*). *Behavioural Processes*, *134*, 70–77.
- Langley, E. J., van Horik, J. O., Whiteside, M. A., & Madden, J. R. (2018). Individuals in larger groups are more successful on spatial discrimination tasks. *Animal Behaviour*, *142*, 87–93.
- ManyDogs Project, Espinosa, J., Bray, E., Buchsbaum, D., Byosiére, S.-E., Byrne, M., Freeman, M. S., Gnanadesikan, G. E., Guran, C.-N. A., Horschler, D. J., Huber, L., Johnston, A. M., MacLean, E. L., Pelgrim, M. H., Santos, L., Silver, Z. A., Stevens, J. R., Völter, C. J., & Zipperling, L. (2021). ManyDogs 1: A multi-lab replication study of dogs' pointing comprehension. *PsyArXiv*. <https://doi.org/10.31234/osf.io/f86jq>
- ManyPrimates, Altschul, D. M., Beran, M. J., Bohn, M., Call, J., DeTroy, S., Duguid, S. J., Egelkamp, C. L., Fichtel, C., Fischer, J., Flessert, M., Hanus, D., Haun, D. B. M., Haux, L. M., Hernandez-Aguilar, R. A., Herrmann, E., Hopper, L. M., Joly, M., Kano, F., ... Watzek, J. (2019). Establishing an infrastructure for collaboration in primate cognition research. *PLoS One*, *14*(11), Article e0223675.
- Matheson, S. M., Asher, L., & Bateson, M. (2008). Larger, enriched cages are associated with 'optimistic' response biases in captive European starlings (*Sturnus vulgaris*). *Applied Animal Behaviour Science*, *109*, 374–383.
- Morton, F. B., Lee, P. C., & Buchanan-Smith, H. M. (2013). Taking personality selection bias seriously in animal cognition research: A case study in capuchin monkeys (*Sapajus apella*). *Animal Cognition*, *16*, 677–684.
- Mowles, S. L., Cotton, P. A., & Briffa, M. (2012). Consistent crustaceans: The identification of stable behavioural syndromes in hermit crabs. *Behavioral Ecology and Sociobiology*, *66*, 1087–1094.
- Nakagawa, S., & Parker, T. H. (2015). Replicating research in ecology and evolution: Feasibility, incentives, and the cost–benefit conundrum. *BMC Biology*, *13*, 1–6.
- O'Neill, S. J., Williamson, J. E., Tosetto, L., & Brown, C. (2018). Effects of acclimation on behavioural repeatability in two behaviour assays of the guppy *Poecilia reticulata*. *Behavioral Ecology and Sociobiology*, *72*, 1–11.
- Postma, M., & Goedhart, J. (2019). PlotsOfData: A web app for visualizing data together with their summaries. *PLoS Biology*, *17*, Article e3000202.
- van Praag, H., Kempermann, G., & Gage, F. H. (2000). Neural consequences of environmental enrichment. *Nature Reviews Neuroscience*, *1*, 191–198.
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ramsay, K., Kaiser, M. J., & Hughes, R. N. (1997). A field study of intraspecific competition for food in hermit crabs (*Pagurus bernhardus*). *Estuarine, Coastal and Shelf Science*, *44*, 213–220.
- Saunders, D. S. (1997). Insect circadian rhythms and photoperiodism. *Invertebrate Neuroscience*, *3*, 155–164.
- Simon, A. F., Chou, M. T., Salazar, E. D., Nicholson, T., Saini, N., Metchev, S., & Krantz, D. E. (2012). A simple assay to study social behavior in *Drosophila*: Measurement of social space within a group 1. *Genes, Brain and Behavior*, *11*, 243–252.
- Turra, A., & Denadai, M. R. (2003). Daily activity of four tropical intertidal hermit crabs from southeastern Brazil. *Brazilian Journal of Biology*, *63*, 537–544.
- Ward, A. J. W. (2012). Social facilitation of exploration in mosquitofish (*Gambusia holbrooki*). *Behavioral Ecology and Sociobiology*, *66*, 223–230.
- Webster, M. M., & Rutz, C. (2020). How STRANGE are your study animals? *Nature*, *582*, 337–340.
- Webster, M. M., & Ward, A. J. W. (2011). Personality and social context. *Biological Reviews*, *86*, 759–773.
- Wilson, D. S., Coleman, K., Clark, A. B., & Biederman, L. (1993). Shy–bold continuum in pumpkinseed sunfish (*Lepomis gibbosus*): An ecological study of a psychological trait. *Journal of Comparative Psychology*, *107*, 250–260.
- Woodward, M. A., Winder, L. A., & Watt, P. J. (2019). Enrichment increases aggression in zebrafish. *Fishes*, *4*, 22.