ORIGINAL ARTICLE

Bias averted: personality may not influence trappability

Chris J. Jolly [1] · Jonathan K. Webb [2] · Graeme R. Gillespie [1,3] · Nelika K. Hughes [1] · Ben L. Phillips [1]

Abstract

If bold animals are more likely to be trapped than shy animals, we take a biased sample of personalities—a problem for behavioural research. Such a bias is problematic, also, for population estimation using mark-recapture models that assume homogeneity in detection probabilities. In this study, we investigated whether differences in boldness result in differences in detection probability in a native Australian rodent, the grassland melomys (*Melomys burtoni*). During a mark-recapture study of this species, we used modified open field tests to assess the boldness (via emergence, and interaction with a novel object) of melomys trapped on the last night of four trapping nights in each of two trapping sessions. Despite melomys showing repeatable variation in these behavioural traits, neither boldness nor emergence latency had an effect on detection probability, and we found no evidence that detection probability varied between individuals. This result suggests that any neophobia is experienced and resolved in individuals of this species on a scale of minutes, rather than the hours across which traps are made available each night. Our work demonstrates that personality-caused sampling bias may not be inevitable, even in situations where animals are required to respond to novelty to be detected, such as in baited traps. Heterogeneity in personality does not inevitably lead to heterogeneity in detection probability.

Significance statement

Historically, passive traps were assumed a non-biased means of sampling animal populations. Increasingly behavioural ecologists suggest that personality traits, particularly individual boldness, may influence behaviour and, as a consequence, could result in sampling bias. Here, we present a comprehensive example of when animal personality has no effect on detection probability. Despite having distinct personalities, detection probabilities of a native Australian rodent, grassland melomys *Melomys burtoni*, were not influenced by whether they were 'shy' or 'bold'. We provide evidence that heterogeneity in personality does not inevitably lead to heterogeneity in detection probability. Given that population estimation models typically assume homogeneity in detection probability between individuals, if this is a broad phenomenon, consistently similar results may improve our confidence in this assumption.

Communicated by A. G. Ophir

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00265-019-2742-4) contains supplementary material, which is available to authorized users.

✉ Chris J. Jolly
  cjolly1@student.unimelb.edu.au

1   School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia

2   School of Life Sciences, University of Technology Sydney, Broadway, Ultimo, NSW 2007, Australia

3   Flora and Fauna Division, Department of Environment and Natural Resources, Northern Territory Government, Berrimah, NT 0828, Australia

Introduction

Why it is that individual animals differ in their behavioural response to potential risks and novel situations has intrigued scientists for decades (see Réale et al. 2007). Increasingly, it is recognised that individuals of many animal species respond predictably in their behaviour, independent of time and/or situation (Sih et al. 2004, 2012; Réale et al. 2007; Wolf and Weissing 2012). Consistent, or repeatable, individual differences in behavioural patterns are referred to as personality in animals (Sih et al. 2004; Koski 2014). Réale et al. (2007) described personality traits as fitting into five categories: activity, aggressiveness, exploration (response to novel situations), shyness-boldness (response to potentially risky situations) and sociability. This

framework for animal personality has since been widely adopted (Biro and Dingemanse 2009; Carter et al. 2013). Behavioural syndromes refer to when two or more of these personality traits correlate across contexts (Sih et al. 2004, 2012), and behavioural syndromes have been garnering increasing attention in the fields of ecology and evolution. Clearly, if animals demonstrate mal-adaptive personality traits (e.g., inappropriate boldness when exposed to a risk of predation), then they are likely to incur fitness costs (reviewed by Smith and Blumstein 2008), and such mal-adaptive behaviours would be expected to be lost via selection (Dall et al. 2004). Yet, animal populations are often found to comprise a breadth of personalities, and many also show evidence of behavioural syndromes (Sih et al. 2012).

Many studies investigating personality in animals use wild-caught individuals that are then transferred to the laboratory (see Carere et al. 2005; Carter et al. 2013). An often-implicit assumption of these laboratory studies is that they have random samples of individuals from the population. If sampling is biased by animal personality, such systematic bias could undermine studies that attempt to understand the distribution of personality traits in populations (Biro and Dingemanse 2009). Undeniably, the existence of 'trap-bold' and 'trap-shy' individuals in populations of animals is a well-known phenomenon and has been observed in numerous taxa (e.g. feral rabbits (*Oryctolagus cuniculus*): Sunnucks 1998; invasive stoats (*Mustela erminea*): King et al. 2003; Bengal tigers (*Panthera tigris tigris*): Wegge et al. 2004; collared flycatchers (*Ficedula albicollis*): Garamszegi et al. 2009). Despite this, most models used to estimate animal population size assume that all individuals have the same detection probability (Jolly 1965; Seber 1970; but see Dorazio and Royle 2003). As a consequence, unmodeled individual-level variation in trappability typically violates model assumptions and can cause biased population estimates (Carothers 1973; Gilbert 1973; Link 2003; Hwang and Huggins 2005; Cubaynes et al. 2010).

In recent years, a number of studies have sought to directly test whether trappability is affected by boldness. For example, a 2009 study found that boldness varied less and was greater in trapped birds than in free-living birds (Garamszegi et al. 2009). This was followed by findings that flight initiation distances of free-living lizards were consistent between trials and strongly predicted the individual's trappability (Carter et al. 2012). Both studies suggest that animals captured using these techniques would generate a biased sample of personalities in these populations. Inverting the problem, numerous studies have actually used trappability as a measure of boldness (e.g., Wilson et al. 1993; Réale et al. 2000; Boyer et al. 2010; Montiglio et al. 2012; Le Cœur et al. 2015).

Whether it is used as a metric, or treated as a methodological nuisance, personality-driven sampling bias is more than just an abstract problem; it can have real-world implications. From a conservation perspective, if behavioural syndromes exist in a population, and trapped animals are bolder than

untrappable individuals (e.g., in response to a novel predator), this could have profound implications for (a) estimating the impact of a threatening process (Ward-Fear et al. 2019) and (b) the success of reintroductions that tend to only relocate the boldest (and so most predator-prone) animals.

More recently, however, animal personality was not found to inevitably lead to sampling bias (Michelangeli et al. 2016). Michelangeli et al. (2016) showed that despite lizards possessing distinct personalities independent of context (behavioural syndromes: Sih et al. 2004), there was no difference in the personality of skinks caught by three different capture methods (hand catching and noosing [active], and pitfall trapping [passive]). Their results imply that trapping bias may not be as pervasive as suspected and may be strongest in methodologies that require animals to respond to novelty, such as that posed by a baited trap (e.g., Carter et al. 2012).

To test the effects of boldness on detection probability, we studied grassland melomys (*Melomys burtoni*), a medium-sized, granivorous, nocturnal, semi-arboreal rodent native to coastal north-eastern Australia (Begg et al. 1983; Taylor and Tulloch 1985; Kemper et al. 1987; Dyer et al. 2011). Despite their name, grassland melomys occur in a variety of well-watered habitats and are relatively common throughout their range (Begg et al. 1983; Taylor and Tulloch 1985; Kemper et al. 1987). To investigate how boldness may affect detection probability in this native Australian rodent, we measured boldness of melomys using modified open field tests. Open field tests have a long history of being used to effectively assess behaviours such as boldness, exploration and neophobia (Montiglio et al. 2012; Carter et al. 2013; Perals et al. 2017), particularly with rodents in the psychological literature (Walsh and Cummins 1976; Gould et al. 2009). Our aim was to investigate the relationship between boldness and detection probability in this species, and its implications for potentially biasing studies of small- to medium-sized mammals using baited traps. Baited traps (e.g. cage, Elliott and Sherman traps) are widely used for studies of wild mammal populations, yet their potential for introducing sampling bias has yet to be studied. We predicted that bolder individuals would be more willing to enter traps over the monitoring period and would, therefore, have higher detection probabilities, irrespective of sex and weight. We also used simulation to explore our ability to detect real effects of boldness on detection probability given our sample sizes.

Methods

Animal collection

Grassland melomys (*Melomys burtoni*) were collected from Indian Island (known as Kabal by Traditional Owners; 25 km^2), Bynoe Harbour, Northern Territory, Australia (−

12°37′24.60″ S, 130°30′0.72″ E), during three trips occurring in May (site 1) and August (sites 2–7) 2017, and April 2018 (sites 1–7). This island is remote and uninhabited by humans, so all monitoring and behavioural experiments were conducted in the field under near natural conditions. Melomys were caught from seven independent 1 ha (100 m × 100 m) plots spread out across Indian Island (plots spaced from 300 m to 9.5 km apart) using a standard mark-recapture trapping regime designed for a monitoring project (Begg et al. 1983; Kemper et al. 1987). Each site consisted of 100 Elliott traps (Elliott Scientific Equipment, Upwey, Victoria) spaced at 10-m intervals in a 10 × 10 grid (Kemper et al. 1987; Tasker and Dickman 2001). Most trapping grids were open for four nights ($n = 6$); however, the first trapping grid was open for six nights. After four trap nights, it is clear that the majority of the population has been captured at least once (Fig. S1). Traps were placed such that they were under, or close to, cover and tended to be placed in areas likely to be routes of travel for these rodents (e.g. against fallen logs). Traps were baited with peanut butter and rolled oats (Paull et al. 2011). These baits were replaced daily for the duration of the trapping session. Elliott traps were checked for captures early each morning, and all traps were cleared within 2 h of sunrise.

On capture, individual melomys were weighed (g) and their sex was determined via the presence of testes or distance between anal opening and genital opening. Each melomys was given a (Trovan Unique ID100) microchip before release, and, on successive mornings, all were scanned (Trovan LID575 Handheld Reader) for a microchip and any new individuals were given a microchip. These data were collected for each night of trapping, and on the morning following the last night of trapping, all melomys caught were retained for behavioural experiments. Throughout the study, 308 individual melomys were captured and given microchips. Of these 308 individuals, 41% were caught on the final night of trapping and were retained for behavioural trials ($n = 125$). Injury and illness can affect rodent behaviour, and the behaviour of nearby conspecifics via olfactory cues (Arakawa et al. 2011). We suspected that pregnancy may also affect female behaviour via altered hormone levels (Pawluski et al. 2009). For this reason, only large, healthy juveniles ($n = 5$), adult male ($n = 58$), and adult non-visibly pregnant female ($n = 62$) melomys were retained for behavioural experiments.

Melomys were transported within their respective Elliott traps from each study site to base camp the morning after the last night of trapping (mean distance = 548 m; range = 133–1244 m; maximum travel time = 15 min). Traps containing melomys were slowly rolled on their roofs and were maintained in this orientation for the duration of their captivity. This orientation of traps was needed for the behavioural experiments, and inverting them early in the day allowed melomys to become accustomed to it. While held at camp, they were kept in a cool, shaded area throughout the day.

Each melomys was provided with food (peanut butter and rolled oats) and water. Two hours prior to behavioural experiments commencing after dark, all melomys were processed (weighed, sexed and, if new, microchipped) and, at this time, were deprived of food to stimulate feeding and exploratory behaviour (Réale et al. 2007). All melomys were released after undergoing open field testing, and within 3 h of dark. Melomys were not maintained in Elliott traps for any longer than 27 h.

Modified open field tests

We employed modified open field tests (also referred to as emergence tests: see Brown and Braithwaite 2004; López et al. 2005; Carter et al. 2013) to assess boldness in grassland melomys. All open field tests were conducted in opaque-walled experimental arenas (540 mm × 340 mm × 370 mm). Experimental arenas were modified plastic boxes that had an inverted Elliott trap sized hole cut in one end and were illuminated by strings of red LED lights (Fig. 1). Each experimental arena had natural sand as substrate, and a bait ball located both in the centre and along one wall of the arena (Fig. 1). After dark, Elliott traps containing a melomys were inserted into the hole in the side of each experimental arena and melomys were allowed to acclimatise for 10 min. At the start of each trial, Elliott trap doors were locked open—the inverted orientation of the trap prevented them from being triggered closed. Melomys were given 10 min to explore the open field arena. After 10 min, a novel object (plastic bowl) was placed at the end of the arena opposite the Elliot trap (Fig. 1) and melomys were given another 10 min to explore the arena and interact with the novel object. Elliott traps remained open during the entirety of the open field tests, and melomys could shelter and emerge from them under their own volition. The sand substrate was replaced, and arenas washed with seawater and detergent between trials to avoid olfactory contamination by conspecifics. All trials were recorded using a GoPro HERO 3. After each trial, the footage was downloaded to a laptop computer for later playback and data analysis. Most melomys ($n = 95$) were only exposed to a single open field test. A subset of melomys ($n = 30$; males: $n = 12$; females: $n = 18$) from four sites, however, were exposed to repeat trials ($n = 3$) to test for repeatability of behaviours (Bell et al. 2009). Repeat trials were conducted on the same night with an hour interval between trials. Repeating trials on successive nights was impossible because we had no way to ensure animals were recaught, nor did we have the facilities to house animals in captivity while in the field on a remote island. Once trials were complete, each melomys was released at its point of capture.

To measure the boldness of individual melomys, we scored three behaviours typically associated with boldness and neophobia in rodents (Dielenberg and McGregor 2001; McGregor et al. 2002; Réale et al. 2007): whether melomys
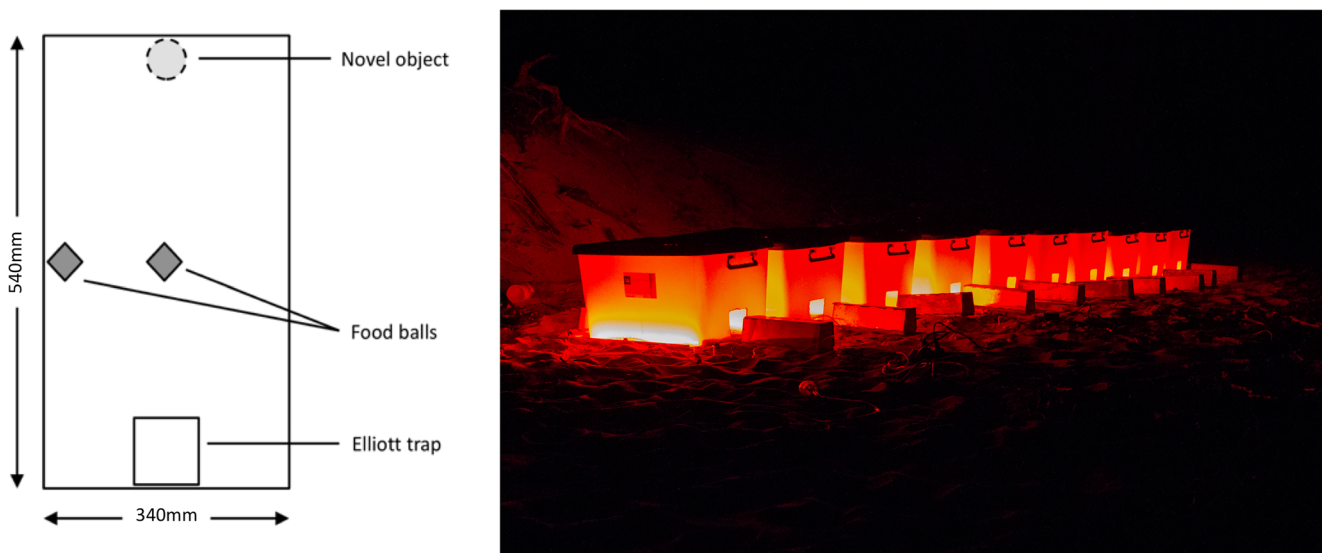
Fig. 1 Open field test experimental setup. Each melomys was allowed to explore the experimental arena for 10 min, after which a novel object (plastic bowl) was placed in each arena, and melomys were given a further 10 min to forage

fully emerged from their Elliott trap hide and entered the open arena during the first 10 min (scored 0 or 1, respectively); whether they fully emerged and entered the trial arena during the second 10 min (scored 0 or 1); and whether they interacted (made contact) with the novel object that was placed in the arena during the second 10 min (scored 0 or 1). Additionally, we recorded emergence latency (seconds) before and after the introduction of the novel object for each trial. For melomys that did not emerge during the 20 min trial, we scored them as having an emergence latency of 1200 s ($n = 47$). Videos were scored by a single observer who was blind to each melomys' origin, identity, behaviour in previous trials, and detection probability. 'Boldness scores' are an index of the response of each melomys to the three measures of boldness. Boldness scores were calculated by summing across the three metrics for each melomys (boldness scores = 0–3). Emergence latency was scored by summing the emergence latency between before (out of 600 s) and after (out of 600 s) the introduction of the novel object.

Statistical analysis

All data analysis was performed using the statistical program R (R Core Team 2019). Behavioural repeatability of boldness scores, emergence latency and novel object interactions through time was assessed using the intraclass correlation coefficient (ICC) using the *rptR* package (Stoffel et al. 2017), for the subset of melomys that were trialled repeatedly. This descriptive statistic describes how strongly multiple repeat measures resemble one another within groups (individual melomys), but it is also influenced by how much individuals differ. The *rptR* package uses generalised linear mixed models to calculate repeatability estimates and allowed us to specify the error structure of our data.

We assumed that individuals caught on the last night of the session were present on the site throughout the session. This assumption is reasonable because each trapping session was relatively brief (between 4 and 6 nights); melomys on Indian Island have very small home ranges (tending to be caught in the same or adjacent traps throughout the trapping period: CJJ et al. unpub. data); we never observed captures of melomys marked at other sites (CJJ et al. unpub. data); and all melomys trialled in open field tests were caught on the final night of trapping (so are clearly present in the population). Mark-recapture analysis estimates the probability of detection given presence, P(detection|presence), and all our animals were present; thus, the proportion of trapping nights each of our animals was observed is a direct estimate of the detection probability used in mark-recapture methods, albeit at an individual, rather than population level. We used the number of capture nights within each trapping session to estimate detection probability for each individual melomys for which we had open field test observations. To do this, we used generalised linear mixed-effects models with binomial errors and a logit link. Since boldness scores were found to be repeatable, only data from the first trial was used in this analysis so that each individual had a single boldness score. We first tested whether boldness score affects detection probability, with session treated as a random effect. This model was run with boldness score defined as both a categorical factor (with four levels: boldness score = 0, 1, 2 and 3) and as a continuous variable (categorical $AIC = 345.39$ vs. continuous $AIC = 342.60$, respectively). Both methods produced similar results, but running the model with boldness score as a continuous variable is the simplest model justifiable (given our question) and also the simplest model with which to develop power analyses (see below), so is the model we report here. Sex

(male, female and unknown [juvenile]) and mass (g) were initially included as fixed effects with and without interaction terms but were removed from the model after they were found to have no effect (sex: $\chi^2 = 139.25$, $df = 3$, $P = 0.41$; mass: $\chi^2 = 125.85$, $df = 107$, $P = 0.96$). We used the same analysis to test whether emergence latency affects detection probability with session treated as a random effect. Again sex and mass were included as fixed effects with and without interaction terms but were removed from the model after they were found to have no effect (sex: $\chi^2 = 139.62$, $df = 3$, $P = 0.85$; mass: $\chi^2 = 126.19$, $df = 107$, $P = 0.86$).

We were also interested in whether there was an interaction between the night on which a melomys was caught and its boldness score (i.e., were shy melomys more likely to be caught the more nights traps were open?). For each melomys, we scored whether or not an individual was caught each trapping night and fitted a model with a night × boldness interaction term (treating night as a continuous variable). This allowed us to assess whether boldness score affected the relationship between trappability and trap night. We also examined a model without this interaction to determine whether animals habituated to traps over time. We tested both of these queries using generalised linear mixed-effects models with session included as a random effect and with binomial errors and a logit link.

Since behaviours are known to be labile and it should not be assumed that a single behavioural measure is necessarily accurate (Niemelä and Dingemanse 2018), we also subset our data to include only melomys that had completed three open field tests ($n = 30$). This allowed us to more rigorously investigate whether mean boldness scores and/or mean values for emergence latency affected detection probability. Mean values were calculated for each individual by averaging their scores across the three repeat trials they experienced. To do this, we used generalised linear models with binomial errors and a logit link to independently test the effect of these two variables on detection probability.

Additionally, since baited traps require animals to respond to novelty in order to be trapped, we used a generalised linear mixed-effects model with session included as a random effect and with binomial errors and a logit link to test whether one component of our boldness score—whether or not an individual interacted with the novel object during open field tests—affected detection probability.

Our findings throughout this study rely on the assumption that our data are not systematically biased by missing the hardest to catch and/or 'shyest' animals during trapping. We test this assumption directly, by testing whether individual detection (p) is sufficiently variable that we may miss sampling a meaningful proportion of behavioural phenotypes within each population. To test whether there is variation in individual detection (p), we use a generalised linear model with binomial errors and a logit link to compare a null model allowing individual-level variation in detection to a null model without individual-level variation in detection. We then compare these models using a likelihood ratio test. When we conduct this test, we find there is no support for the model allowing variance across individuals in detection probability ($\chi^2 = 577.77$, $df = 1$, $P = 0.65$). This, coupled with a mean per night detection probability of around 0.55 (see results), suggests it is very unlikely that we have an unsampled class of undetectable animals in our study.

For all analyses, P values were obtained by likelihood ratio tests comparing models with and without the effect in question and are presented as F statistics or chi-squared values and P values. Statistical significance was assigned at $\alpha = 0.05$.

Since we failed to reject our null hypothesis and our results were at odds with a number of previous studies that suggest boldness could bias sampling (Wilson et al. 1993; Biro and Dingemanse 2009; Garamszegi et al. 2009; Carter et al. 2012; Stuber et al. 2013; but see Michelangeli et al. 2016), we conducted a power analysis to assess the effect size we could expect to detect given our sample size for each boldness category. To understand how our power to detect an effect changes with effect size, we simulated data (10,000 sets) with sample sizes for each boldness category equivalent to those we obtained, but with varying (linear) effect sizes (see Supplementary Material). These simulated sets were used to evaluate our power to detect effects of varying sizes. All analyses conducted were performed using R (R Core Team 2019).

Results

Melomys showed repeatable behaviour for boldness score, emergence latency and novel object response when individuals responses were compared between the three trials (ICC: boldness score: R [± 95%CI] = 0.67 [0.47, 0.80], $P > 0.001$; emergence time: R [± 95%CI] = 0.73 [0.53, 0.83], $P > 0.001$; novel object: R [± 95%CI] = 0.61 [0.209, 0.974], $P > 0.001$). Despite 'bold' melomys being repeatedly bold between trials (high boldness scores [2–3]) and 'shy' melomys being repeatedly shy between trials (low boldness scores [0–1]), we found no significant overall effect of boldness ($F = 0.077$, $df = 1$, $P = 0.78$; Fig. 2; Table 1) nor emergence latency ($F = 0.337$, $df = 1$, $P = 0.56$; Table 2) on detection probability in melomys. That is, bold melomys were no more likely to be re-caught on successive nights of trapping than were shy individuals. There was also no significant interaction between trap night and boldness score ($F = 0.917$, $df = 1$, $P = 0.34$; Table 3), and similarly, no evidence for habituation (detection systematically changing over time: $F = 198$, $df = 1$, $P = 0.66$). When we subset the data to ensure we were not missing an effect of behaviour that would only be detected from repeated samples of each individual, we found no significant effect of mean boldness ($\chi^2 = 18.10$, $df = 1$, $P = 0.95$; Table 4) nor an effect of

Fig. 2 Estimated mean detection probability (± 95%CI) of grassland melomys (*Melomys burtoni*) in each of four boldness score categories. *n* indicates the number of melomys falling into each boldness category. The black line indicates the minimum linear effect size (expressed as the detection probability) that we could have detected with 80% power given our sample sizes for each category
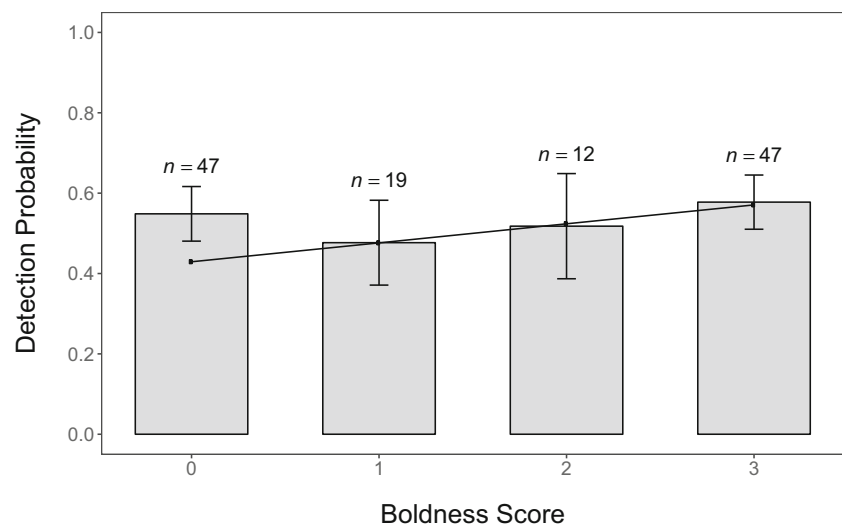
mean emergence latency ($\chi^2 = 18.01$, $df = 1$, $P = 0.93$; Table 5) on detection probability in melomys. Additionally, melomys that interacted with the novel object during open field tests were not more likely to be detected during monitoring than were individuals that did not ($F = 0.101$, $df = 1$, $P = 0.75$).

To understand the effect size we could expect to detect given our data, we simulated 10,000 datasets with sample sizes identical to our dataset within each boldness category, but with varying effect sizes on detection. With these sample sizes, we had sufficient power (80%) to detect an effect of boldness on detection probability with a slope ≥ 0.19 (Fig. 3). For the mean detection probability we observed in our data ($P(\text{detection}|\text{present}) = 0.549$), this is equivalent to detecting a change in detection probability of 0.14, between our shyest and boldest categories (presented as the detectable effect slope in Fig. 3).

Discussion

Despite melomys having a repeatable behavioural type through time—bold melomys were consistently bold and shy melomys were consistently shy between trials—we detected no effect of this behavioural variation on detection

probability (Fig. 2). That is, bold individuals were no more likely to be trapped than were shy individuals.

Although it is impossible to entirely rule out the possibility that sampling bias could have resulted in us only detecting the boldest individuals and systematically missing the hardest to capture individuals in the population (Biro and Dingemanse 2009; Biro 2013), this possibility is very unlikely in our case. Since our mean per-night detection probability is reasonably high (around 0.55, Fig. 2), and we find no evidence of individual variation in detectability, it is very unlikely that we systematically missed some individuals. Additionally, large variation in the behaviour of tested melomys further suggests it is unlikely we are relying on biased data (see Fig. 2). Consistent with a lack of individual variation in detection, we also found that variation in emergence latency had no effect on detection probability—melomys that rapidly emerged from their shelters were not more detectable than those that took longer to emerge or did not emerge during open field tests. To ensure we had not inaccurately assigned boldness scores and emergence times to individual melomys, we reassessed this question using mean values for these traits (scored from three trials for 30 individuals); however, we detected no effect of mean values for either behavioural measure on detection probability (Tables 4 and 5). We anticipated that these measures of boldness would influence detection

Table 1 Generalised linear mixed-effects model (GLMM) testing whether boldness affects detection probability with session treated as a random effect. This is the simplest model justifiable given our question and is the model used for our power analysis

Fixed effects	Estimate	Standard error	P
Intercept	0.170	0.326	
Boldness	0.018	0.067	0.782
Random effect	**Variance**	**Standard deviation**	
Session	0.178	0.422	

Table 2 Generalised linear mixed-effects model (GLMM) testing whether emergence latency affects detection probability with session treated as a random effect

Fixed effects	Estimate	SE	P
Intercept	0.066	0.387	
Boldness	0.0001	0.0002	0.562
Random effect	**Variance**	**Standard deviation**	
Session	0.823	0.428	

Table 3 Generalised linear mixed-effects model (GLMM) testing whether an interaction between boldness and trap night affects detection, with session treated as a random effect

Fixed effects	Estimate	SE	P
Intercept	− 0.755	0.462	
Boldness	0.177	0.175	0.312
Trap night	0.138	0. 137	0.314
Boldness*trap night	− 0.066	0.069	0.339
Random effect	**Variance**	**Standard deviation**	
Session	0.180	0.424	

Table 5 Generalised linear model (GLM) testing whether mean emergence latency across three open field tests for each individual ($n = 30$) affects detection probability

Fixed effects	Estimate	SE	P
Intercept	$-4.62e^{-2}$	$5.81e^{-1}$	
Boldness	$4.82e^{-5}$	$5.75e^{-4}$	0.933

probability and that bold melomys would be caught more frequently than would shy individuals (Biro and Dingemanse 2009; Garamszegi et al. 2009; Carter et al. 2012; Biro 2013; Stuber et al. 2013; but see Michelangeli et al. 2016), but we were unable to reject the null hypothesis of no effect. Our power analysis implies that we did not simply fail to detect a large effect but that intraspecific differences in boldness really can only have a very small effect (if any) on the detection probability in these rodents (Fig. 2).

Our results are both intriguing and have important implications for animal personality research and population censusing in general. Ours is one of very few studies to demonstrate that trapping bias is not inevitable (Biro and Dingemanse 2009; Michelangeli et al. 2016), despite our use of traps that require animals to respond to novelty (Wilson et al. 1993; Carter et al. 2012; Biro 2013; Stuber et al. 2013). Michelangeli et al. (2016) found strikingly similar results to ours when they tested whether three different capture methods (both active and passive [pitfall trapping]) of lizards (*Lampropholis delicata*) resulted in sampling bias of bolder personality types. In their study, they found that trapping method did not result in sampling bias and concluded that personality-caused sampling bias may be confined to passive trapping methods that require animals to respond to novelty (e.g., baited traps). Here, we demonstrate that personality-caused sampling bias may not even be inevitable in passive sampling that requires response to novelty. In fact, in our open field tests, responses of melomys to a novel object had no effect on their detection probability. This result suggests that neophobia may be experienced and resolved in individuals of this species on a scale of tens of minutes, rather than the hours across which traps are made available each night.

Table 4 Generalised linear model (GLM) testing whether mean boldness across three open field tests for each individual ($n = 30$) affects detection probability

Fixed effects	Estimate	SE	P
Intercept	− 0.011	0.259	
Boldness	0.010	0.173	0.954

Experiments designed to measure animal personality have recently received critical review (Carter et al. 2013), and it is certainly worth considering whether our experiment is appropriate to measure the personality trait we intended to measure. The aim of our study was to measure boldness and to determine whether it influences detection probability, and we can be reasonably confident that our modified open field tests are truly measuring boldness in trialled individuals. Open field tests have been found to test multiple personality traits at once (e.g., exploration and boldness: Walsh and Cummins 1976; Bell et al. 2009; Perals et al. 2017) and have been criticised for this (Carter et al. 2013). Our modification of the open field test design, by allowing melomys to emerge from hiding of their own volition (e.g., latency to emerge), is thought to be a true measure of boldness (Perals et al. 2017). Furthermore, while designing experiments that solely test a single personality trait may be desirable under some circumstances, compound measures are often more ecologically relevant, and designing experiments that evaluate this may be more informative. Certainly, if personality traits are structured into behavioural syndromes, for which there is strong evidence (Sih et al. 2004, 2012), an individual's response to any circumstance (including a behavioural experiment) may almost certainly be influenced by multiple personality traits. Thus, our metrics are broadly accepted metrics of boldness and neophobia, and they are likely ecologically relevant.

Given our metrics are reasonable, it is striking that this behavioural heterogeneity did not result in appreciable detection heterogeneity. In mark-recapture models, it is generally assumed that there is no unaccounted heterogeneity in detection. Behavioural variation is a very likely source of individual-level variation in detectability, and such individual-level variation can cause extreme difficulties in estimating population size in mark-recapture studies, particularly where detection rate is low (Link 2003; Cubaynes et al. 2010). In our case, despite clear variation between individuals in trap-relevant aspects of personality (i.e., boldness and neophobia), this heterogeneity had very little bearing on the animals' trappability; indeed our analysis suggests that there may be little to no individual variation in trappability in our system.

Trappability has often been used as an index for boldness in animal personality studies (Boyer et al. 2010; Montiglio et al. 2012; Le Cœur et al. 2015), but our results suggest that this should not necessarily be assumed (see also Vanden Broecke

Fig. 3 Estimated power to detect an effect (at $\alpha = 0.05$) with fixed sample sizes (boldness score: 0 [$n = 47$]; 1 [$n = 19$]; 2 [$n = 12$]; and 3 [$n = 47$]), calculated over a range of effect sizes
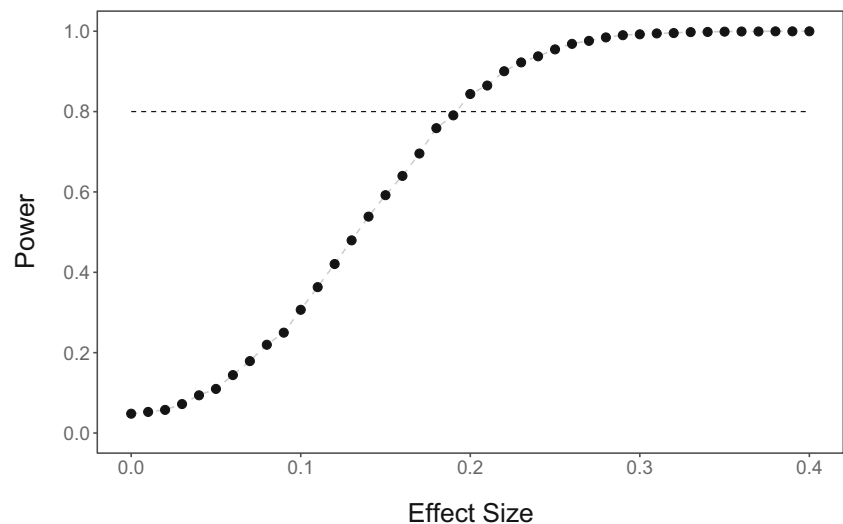
et al. 2018). Certainly, against our metric for boldness (measured in open field trials), there is no correlation with trappability, despite us having considerable power to detect such an effect. We suspect that in this case, any neophobia invoked by the presence of a trap in our system may (a) operate at a time-scale that is irrelevant to trappability and (b) be overwhelmed by the lure of an easy, high-value meal. In our study (as in most studies of wild rodents), traps are both abundant relative to population density (at 10 m intervals) and available to exploit (open) all night. As a result, there is a high probability of individual melomys encountering an open trap and then having hours to investigate it. Although neophobia occurs, it tended to be resolved in minutes in our open field test recordings. Therefore, the time-scale at which neophobia may operate in this species means that variation in this trait may simply missed by trapping. Thus, attention to survey design—ensuring traps are dense enough to avoid trap saturation—may be a general strategy for avoiding a biased sample of personalities. Such considerations also ensure sampling meets the assumptions of most mark-recapture designs (Royle and Dorazio 2008), so well-designed trap-based mark-recapture studies may often avoid personality sampling bias also.

Neophobia may often play out at short time scales, and this seems particularly likely in situations where resources are ephemeral or competition for resources is intense. In our system, in the wet-dry tropics of Northern Australia, for example, rainfall is extremely seasonal—confined to a short wet season—and is very stochastic in both its amount and timing (Taylor and Tulloch 1985). This stochasticity results in resources such as food being extremely unreliable and unpredictable through space and time (Madsen and Shine 1996; Shine and Brown 2008). Such an environment should favour rapid resolution of neophobia, though such environments—in which there is seasonal scarcity—are common. Additionally, predator diversity (Terborgh et al. 2001) and anti-predator responses of prey (Cooper et al. 2014) tend to be reduced on islands, and so selection against boldness may be weaker in our study environment.

Further study investigating personality-induced trapping biases in mainland systems, and in those with steady resource supply (if such can be found), would provide interesting comparisons to the current study.

Although we found no relationship between boldness and trappability, we caution against generalising these results too far. There is evidence that boldness affects trappability/ detection in some species with some trapping designs, and boldness may have other life history consequences. As such, where there are important consequences, the precautionary principle suggests we should assume there to be bias inherent in sampling for each species until proven otherwise. This is of particular concern for the discipline of conservation biology, where sampling bias due to personality could have major implications. At the very least, sampling bias as a result of personality in threatened species could lead to inaccurate population estimates, and inaccurate estimates of impact of threatening processes. Additionally, if behavioural syndromes bias sampling of threatened species to only the boldest individuals in the population, this could lead to reduced success of reintroduction programs. This is of particular concern in Australia, where predation has led to many failed reintroductions (Moseby et al. 2015). Since boldness has been found to incur fitness costs due to reduced survival (Smith and Blumstein 2008), it is of paramount importance that further research is conducted to determine whether reintroduction programs are unintentionally introducing a bias towards collecting and reintroducing only the boldest animals within a threatened population.

Conclusion

The trapping methodology used in this study—baited traps— is one of the most commonly used trapping methodologies for small- to medium-sized mammals worldwide (Tasker and Dickman 2001), and our results suggest that, when deployed

at sufficient density, such traps are an appropriate, unbiased means of sampling behavioural variation in rodents such as the grassland melomys. We found no correlation between boldness and nightly detection probability in melomys. We provide the first evidence that individual differences in animal personality do not bias this method of trapping and provide additional evidence to support the suggestion that trapping bias is not always inevitable (Michelangeli et al. 2016). Baited traps require animals to respond to novelty and overcome neophobia in order to become trapped, and this has led to trappability in such traps being used as a measure of boldness (Wilson et al. 1993; Boyer et al. 2010; Montiglio et al. 2010; Carter et al. 2012; Le Cœur et al. 2015). It may, however, be important to assess species' biology and trapping context (e.g., island tameness [Rodl et al. 2007; Cooper et al. 2014] and an ability to respond to unpredictable resources) when considering whether to use trappability as a reliable measure of boldness. We provide strong evidence to suggest that trappability should only be used as a measure of boldness with caution and additional experiments may be required to test whether trappability is indeed a good measure of this trait. Finally, individual heterogeneity in detection poses a significant impediment to accurate estimates of population size using standard mark-recapture models. We provide evidence that individual-level variation in boldness does not inevitably lead to individual-level detection heterogeneity.

Data availability The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval All applicable international, national and/or institutional guidelines for the use of animals were followed. We obtained permission to access Indian Island and to trap melomys from Indigenous Traditional Owners as part of on-going monitoring of a northern quoll population on the island (Permit to enter & remain on Aboriginal land or sea adjoining Aboriginal land (Research) from Northern Land Council, NT). We obtained permits allowing us to trap melomys from the Parks and Wildlife Commission of the Northern Territory (Permit to Undertake Scientific Research on Wildlife). The University of Melbourne Animal Ethics Committee approved the experimental design (ID number: 1814518).

References

Arakawa H, Cruz S, Deak T (2011) From models to mechanisms: odorant communication as a key determinant of social behavior in rodents during illness-associated states. Neurosci Biobehav Rev 35:1916–1928. https://doi.org/10.1016/j.neubiorev.2011.03.007

Begg R, Walsh B, Woerle F, King S (1983) Ecology of *Melomys burtoni*, the grassland Melomys (Rodentia: Muridae) at Cobourg Peninsula, N.T. Wildl Res 10:259–267

Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behaviour: a meta-analysis. Anim Behav 77:771–783

Biro PA (2013) Are most samples of animals systematically biased? Consistent individual trait differences bias samples despite random sampling. Oecologia 171:339–345. https://doi.org/10.1007/s00442-012-2426-5

Biro PA, Dingemanse NJ (2009) Sampling bias resulting from animal personality. Trends Ecol Evol 24:66–67. https://doi.org/10.1016/j.tree.2008.11.001

Boyer N, Réale D, Marmet J, Pisanu B, Chapuis JL (2010) Personality, space use and tick load in an introduced population of Siberian chipmunks *Tamias sibiricus*. J Anim Ecol 79:538–547. https://doi.org/10.1111/j.1365-2656.2010.01659.x

Brown C, Braithwaite VA (2004) Size matters: a test of boldness in eight populations of the poeciliid *Brachyraphis episcopi*. Anim Behav 68:1325–1329. https://doi.org/10.1016/j.anbehav.2004.04.004

Carere C, Drent PJ, Privitera L, Koolhaas JM, Groothuis TGG (2005) Personalities in great tits, *Parus major*: stability and consistency. Anim Behav 70:795–805. https://doi.org/10.1016/j.anbehav.2005.01.003

Carothers AD (1973) The effects of unequal catchability on Jolly-Seber estimates. Biometrics 29:79. https://doi.org/10.2307/2529678

Carter AJ, Heinsohn R, Goldizen AW, Biro PA (2012) Boldness, trappability and sampling bias in wild lizards. Anim Behav 83:1051–1058. https://doi.org/10.1016/j.anbehav.2012.01.033

Carter AJ, Feeney WE, Marshall HH, Cowlishaw G, Heinsohn R (2013) Animal personality: what are behavioural ecologists measuring? Biol Rev 88:465–475. https://doi.org/10.1111/brv.12007

Cooper WE, Pyron RA, Garland T (2014) Island tameness: living on islands reduces flight initiation distance. Proc R Soc B 281:20133019. https://doi.org/10.1098/rspb.2013.3019

Cubaynes S, Pradel R, Choquet R et al (2010) Importance of accounting for detection heterogeneity when estimating abundance: the case of French wolves. Conserv Biol 24:621–626. https://doi.org/10.1111/j.1523-1739.2009.01431.x

Dall SRX, Houston AI, McNamara JM (2004) The behavioural ecology of personality: consistent individual differences from an adaptive perspective. Ecol Lett 7:734–739. https://doi.org/10.1111/j.1461-0248.2004.00618.x

Dielenberg RA, McGregor IS (2001) Defensive behavior in rats towards predatory odors: a review. Neurosci Biobehav Rev 25:597–609. https://doi.org/10.1016/S0149-7634(01)00044-6

Dorazio RM, Royle JA (2003) Mixture models for estimating the size of a closed population when capture rates vary among individuals. Biometrics 59:351–364. https://doi.org/10.1111/1541-0420.00042

Dyer BC, Clarke AR, Fuller SJ (2011) Population dynamics, diet and pest status of the grassland melomys (*Melomys burtoni*) in northern

Queensland sugarcane crops. Wildl Res 38:330. https://doi.org/10.1071/WR10171

Garamszegi LZ, Eens M, Török J (2009) Behavioural syndromes and trappability in free-living collared flycatchers, *Ficedula albicollis*. Anim Behav 77:803–812. https://doi.org/10.1016/j.anbehav.2008.12.012

Gilbert RO (1973) Approximations of the bias in the Jolly-Seber capture-recapture model. Biometrics 29:501. https://doi.org/10.2307/2529173

Gould TD, Dao DT, Kovacsics CE (2009) The open field test. In: Gould TD (ed) Mood and anxiety related phenotypes in mice. Humana Press, Totowa, NJ, pp 1–20

Hwang W-H, Huggins R (2005) An examination of the effect of heterogeneity on the estimation of population size using capture-recapture data. Biometrika 92:229–233. https://doi.org/10.1093/biomet/92.1.229

Jolly GM (1965) Explicit estimates from capture-recapture data with both death and immigration-stochastic model. Biometrika 52:225–247

Kemper CM, Kitchener DJ, Humphreys WF, How RA, Bradley AJ, Schmitt LH (1987) The demography and physiology of *Melomys* sp. (Rodentia: Muridae) in the Mitchell Plateau area, Kimberley, Western Australia. J Zool 212:533–562. https://doi.org/10.1111/j.1469-7998.1987.tb02923.x

King CM, Davis SA, Purdey D, Lawrence B (2003) Capture probability and heterogeneity of trap response in stoats (*Mustela erminea*). Wildl Res 30:611. https://doi.org/10.1071/WR02091

Koski SE (2014) Broader horizons for animal personality research. Front Ecol Evol 2:1–6. https://doi.org/10.3389/fevo.2014.00070

Le Cœur C, Thibault M, Pisanu B, Thibault S, Chapuis JL, Baudry E (2015) Temporally fluctuating selection on a personality trait in a wild rodent population. Behav Ecol 26:1285–1291. https://doi.org/10.1093/beheco/arv074

Link WA (2003) Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. Biometrics 59:1123–1130. https://doi.org/10.1111/j.0006-341X.2003.00129.x

López P, Hawlena D, Polo V, Amo L, Martín J (2005) Sources of individual shy–bold variations in antipredator behaviour of male Iberian rock lizards. Anim Behav 69:1–9. https://doi.org/10.1016/j.anbehav.2004.05.010

Madsen T, Shine R (1996) Seasonal migration of predators and prey-a study of pythons and rats in tropical Australia. Ecology 77:149–156. https://doi.org/10.2307/2265663

McGregor IS, Schrama L, Ambermoon P, Dielenberg RA (2002) Not all 'predator odours' are equal: cat odour but not 2, 4, 5 trimethylthiazoline (TMT; fox odour) elicits specific defensive behaviours in rats. Behav Brain Res 129:1–16

Michelangeli M, Wong BBM, Chapple DG (2016) It's a trap: sampling bias due to animal personality is not always inevitable. Behav Ecol 27:62–67. https://doi.org/10.1093/beheco/arv123

Montiglio P-O, Garant D, Thomas D, Réale D (2010) Individual variation in temporal activity patterns in open-field tests. Anim Behav 80:905–912. https://doi.org/10.1016/j.anbehav.2010.08.014

Montiglio P-O, Garant D, Pelletier F, Réale D (2012) Personality differences are related to long-term stress reactivity in a population of wild eastern chipmunks, *Tamias striatus*. Anim Behav 84:1071–1079. https://doi.org/10.1016/j.anbehav.2012.08.010

Moseby KE, Carthey AJR, Schroeder T (2015) The influence of predators and prey naivety on reintroduction success: current and future directions. In: Armstrong DP, Hayward MW, Moro D, Seddon PJ (eds) Advances in reintroduction biology of Australia and New Zealand Fauna. CSRIO Publishing, Collingwood, Victoria, pp 29–42

Niemelä PT, Dingemanse NJ (2018) On the usage of single measurements in behavioural ecology research on individual differences. Anim Behav 145:99–105. https://doi.org/10.1016/j.anbehav.2018.09.012

Paull DJ, Claridge AW, Barry SC (2011) There's no accounting for taste: bait attractants and infrared digital cameras for detecting small to medium ground-dwelling mammals. Wildl Res 38:188. https://doi.org/10.1071/WR10203

Pawluski JL, Brummelte S, Barha CK, Crozier TM, Galea LA (2009) Effects of steroid hormones on neurogenesis in the hippocampus of the adult female rodent during the estrous cycle, pregnancy, lactation and aging. Front Neuroendocrinol 30:343–357. https://doi.org/10.1016/j.yfrne.2009.03.007

Perals D, Griffin AS, Bartomeus I, Sol D (2017) Revisiting the open-field test: what does it really tell us about animal personality? Anim Behav 123:69–79. https://doi.org/10.1016/j.anbehav.2016.10.006

R Core Team (2019) R: a language and environment for statistical computing. R Foundation for statistical computing, Vienna, http://www.R-project.org

Réale D, Gallant BY, Leblanc M, Festa-Bianchet M (2000) Consistency of temperament in bighorn ewes and correlates with behaviour and life history. Anim Behav 60:589–597. https://doi.org/10.1006/anbe.2000.1530

Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ (2007) Integrating animal temperament within ecology and evolution. Biol Rev 82:291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x

Rodl T, Berger S, Michael Romero L, Wikelski M (2007) Tameness and stress physiology in a predator-naive island species confronted with novel predation threat. Proc R Soc Lond B 274:577–582. https://doi.org/10.1098/rspb.2006.3755

Royle JA, Dorazio RM (2008) Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. Academic Press, San Diego

Seber GAF (1970) The effects of trap response on tag recapture estimates. Biometrics 26:13. https://doi.org/10.2307/2529040

Shine R, Brown GP (2008) Adapting to the unpredictable: reproductive biology of vertebrates in the Australian wet-dry tropics. Philos Trans R Soc B 363:363–373. https://doi.org/10.1098/rstb.2007.2144

Sih A, Bell A, Johnson JC (2004) Behavioral syndromes: an ecological and evolutionary overview. Trends Ecol Evol 19:372–378. https://doi.org/10.1016/j.tree.2004.04.009

Sih A, Cote J, Evans M, Fogarty S, Pruitt J (2012) Ecological implications of behavioural syndromes. Ecol Lett 15:278–289. https://doi.org/10.1111/j.1461-0248.2011.01731.x

Smith BR, Blumstein DT (2008) Fitness consequences of personality: a meta-analysis. Behav Ecol 19:448–455. https://doi.org/10.1093/beheco/arm144

Stoffel MA, Nakagawa S, Schielzeth H (2017) rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods Ecol Evol 8:1639–1644

Stuber EF, Araya-Ajoy YG, Mathot KJ, Mutzel A, Nicolaus M, Wijmenga JJ, Mueller JC, Dingemanse NJ (2013) Slow explorers take less risk: a problem of sampling bias in ecological studies. Behav Ecol 24:1092–1098. https://doi.org/10.1093/beheco/art035

Sunnucks P (1998) Avoidance of novel objects by rabbits (*Oryctolagus cuniculus L.*). Wildl Res 25:273. https://doi.org/10.1071/WR97038

Tasker E, Dickman C (2001) A review of Elliott trapping methods for small mammals in Australia. Aust Mammal 23:77. https://doi.org/10.1071/AM01077

Taylor JA, Tulloch D (1985) Rainfall in the wet-dry tropics: extreme events at Darwin and similarities between years during the period 1870–1983 inclusive. Austral Ecol 10:281–295. https://doi.org/10.1111/j.1442-9993.1985.tb00890.x

Terborgh J, Lopez L, Nunez P, Rao M, Shahabuddin G, Orihuela G, Riveros M, Ascanio R, Adler GH, Lambert TD, Balbas L (2001) Ecological meltdown in predator-free forest fragments. Science 294:1923–1926

Vanden Broecke B, Borremans B, Mariën J, Makundi RH, Massawe AW, Leirs H, Hughes NK (2018) Does exploratory behavior or activity in

a wild mouse explain susceptibility to virus infection? Curr Zool 64: 585–592. https://doi.org/10.1093/cz/zox053

Walsh RN, Cummins RA (1976) The open-field test: a critical review. Psychol Bull 83:482–504

Ward-Fear G, Rangers B, Pearson D, Bruton M, Shine R (2019) Sharper eyes see shyer lizards: collaboration with indigenous peoples can alter the outcomes of conservation research. Conserv Lett (published online, https://doi.org/10.1111/conl.12643)

Wegge P, Pokheral CP, Jnawali SR (2004) Effects of trapping effort and trap shyness on estimates of tiger abundance from camera trap studies. Anim Conserv 7:251–256. https://doi.org/10.1017/S1367943004001441

Wilson DS, Coleman K, Clark AB, Biederman L (1993) Shy-bold continuum in pumpkinseed sunfish (*Lepomis gibbosus*): an ecological study of a psychological trait. J Comp Psychol 107:250–260. https://doi.org/10.1037//0735-7036.107.3.250

Wolf M, Weissing FJ (2012) Animal personalities: consequences for ecology and evolution. Trends Ecol Evol 27:452–461. https://doi.org/10.1016/j.tree.2012.05.001