Animal Behaviour 83 (2012) 1295–1300

Contents lists available at SciVerse ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Commentary

# Do rapid assays predict repeatability in labile (behavioural) traits?

Peter A. Biro*

*Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Waurn Ponds, Australia*

ARTICLE INFO

*Article history:*
Received 20 October 2011
Initial acceptance 11 November 2011
Final acceptance 18 January 2012
Available online 14 March 2012
MS. number: 11-00843R

*Keywords:*
acclimation
boldness
coping
endocrine
habituation
mixed model
personality plasticity
random regression
temperament

Laboratory studies of labile animal traits (e.g. physiology, behaviour) often place animals into a novel testing apparatus to do short-term assays, and later repeat the procedure to evaluate repeatability. Because animals in nature are never forced into these unnatural situations, measured values may not reflect those observed under more familiar (and therefore more natural) conditions. Thus, we implicitly assume rank order differences across individuals are maintained between these two contexts. I repeatedly assayed behavioural traits for young fish (Ward's damselfish, *Pomacentrus wardi*) in their home tanks, and observed significant repeatability once they were acclimated, but observations taken over the first 2 days did not predict behavioural types evident from subsequent observations. This cautionary note indicates that rapid assays of behavioural traits can significantly misclassify individuals. Furthermore, numerous physiological traits are often correlated with behaviour, suggesting caution for physiological studies as well. Future studies should not assume that labile trait assays predict scores under familiar conditions and, more importantly, should test whether scores under familiar laboratory conditions predict those observed in the field.

Although many biologists study animal traits in the laboratory, we ultimately want to understand the causes and consequences of individual trait differences that would be expressed under natural conditions. Studying how animals express labile traits (e.g. those related to physiology or behaviour) in the laboratory makes experiments more tractable, and provides control to isolate a given effect of interest. However, there are potential problems with this approach that seem to have gone largely unnoticed.

Studies often remove individuals directly from the field (Reale et al. 2000; Martin & Reale 2008; Boratynski & Koteja 2009), or from group housing in the laboratory (Ksiazek et al. 2004; Wilson et al. 2009), and then place them into a novel test apparatus to conduct short-term trait assays. Capturing the animals and then forcing them into novel (and presumably highly stressful) situations clearly differs from natural conditions. In nature, many animals can choose whether or not to expose themselves to novel conditions, and as a result they may often occupy habitats/situations that are usually familiar to them and that are not extremely stressful. When novelty is encountered in the wild (e.g. during dispersal, sudden appearance of a new predator), their response to it is not affected by artificially imposed stress, nor is their response constrained by unfamiliar and unnatural laboratory conditions. Therefore, the implicit assumption of any laboratory study of a labile trait, such as behaviour, is that its expression is a good predictor of trait

* Correspondence: P. A. Biro, Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Waurn Ponds, VIC 3216, Australia.
*E-mail address:* pete.biro@deakin.edu.au.

0003-3472/$38.00 2012 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd.
doi:10.1016/j.anbehav.2012.01.036

expression of those same individuals under more familiar, less stressful, and therefore more natural conditions.

To evaluate a labile trait, researchers often repeatedly observe an animal's response to an assay for repeatability (i.e. consistent individual differences). Habituation (a decline in response to novelty or stress) and acclimation (a change in response while adjusting to novelty/stress) are common responses to repeated assays, occurring for hormonal (Romero 2004), physiological (Ellenberg et al. 2009) and behavioural traits (Budaev 1997; Romero 2004; Martin & Reale 2008; Wong et al. 2010). However, the assumption that trait levels measured under novel laboratory conditions are a predictor of those observed under familiar conditions will not be violated so long as individual acclimation responses are similar, and thus individuals maintain their rank order between novel and familiar conditions (Fig. 1a). However, if individuals differ significantly in the form of their acclimation response, the assumption is violated, and therefore rapid assays of labile traits will not predict those under familiar conditions (Fig. 1b). None the less, studies using rapid assays under novel conditions have clearly been informative about performance in the field (e.g. Reale et al. 2000; Boon et al. 2007). However, if the assumption is violated, then we may misclassify at least some of the individuals in a sample, and this could in turn affect our power to detect relationships between an individual's behavioural type and other variables of interest.

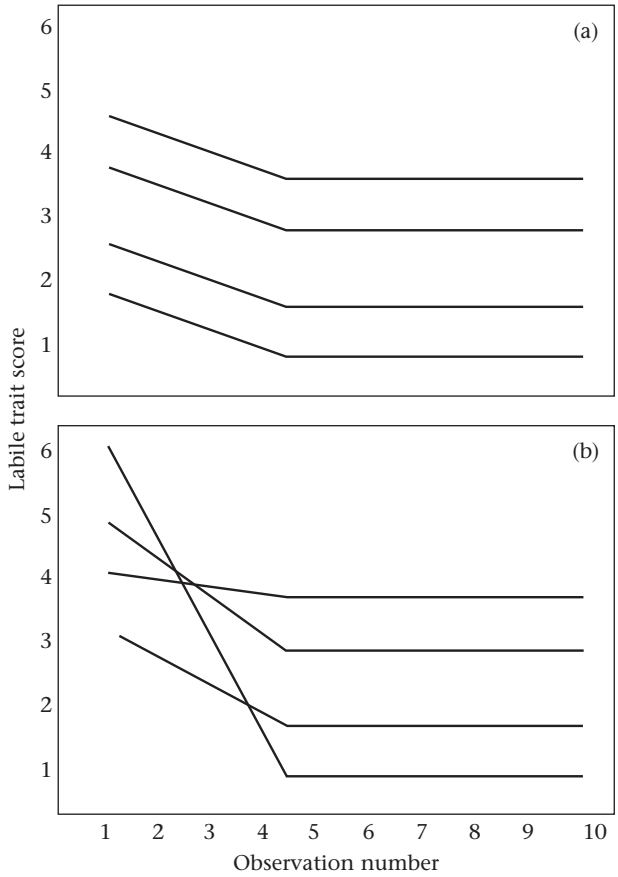


**Figure 1.** Hypothetical changes in a labile trait across repeated observations (e.g. locomotor activity or stress hormone response in an initially novel environment). (a) Individuals that consistently differ in their average levels of a trait, and follow a similar pattern of acclimation. In this scenario, even just a few initial (rapid) assays of naïve animals would yield a good estimate of the trait differences between individuals under familiar (and more natural) conditions, and we would detect significant repeatability both during and after the acclimation period. By contrast, (b) illustrates differences in acclimation, whereby the rank order of relatively naïve individuals would give an incorrect assessment of trait differences of animals under familiar conditions. In this second scenario, we would probably not detect repeatability during the acclimation phase, but would detect it after. Acclimation responses over time are illustrated as a two-phase process for ease of illustration, and to facilitate comparison with some of the results of the present study.

Surprisingly, it seems that this assumption has not been tested directly (but see Martin & Reale 2008; Rodríguez-Prieto et al. 2010, 2011). Perhaps this is so because testing the assumption would require numerous observations conducted on many individuals, and the need to span periods that include acclimation and post-acclimation intervals (Fig. 1). Indeed, studies rarely measure a labile trait more than twice to assess repeatability (Nespolo & Franco 2007; Bell et al. 2009), which is insufficient to test this assumption and might even explain why reported repeatability values are often low (Nespolo & Franco 2007; Williams 2008; Bell et al. 2009).

I repeatedly assayed behavioural responses of young fish housed in home tanks to test whether or not observations made under forced novel conditions predict behavioural traits under familiar and (presumably) less stressful conditions, conditions that are likely to be most similar to what animals in nature experience, most of the time. By extension, results from this study may also have relevance for a variety of physiological traits, such as metabolism and endocrine hormone levels, because they are often correlated with behavioural traits (e.g. Carlson 1986; Gosling 2001; Sih et al. 2004; Overli et al. 2005; Careau et al. 2008, 2010; Sih & Bell 2008; Williams 2008; Biro & Stamps 2010). This study thus represents an important first step towards determining whether rapid assays of labile traits are informative of what we might observe in the field.

## METHODS

I performed the experiment in a temperature-controlled laboratory at Lizard Island Research Station, located on the northern Great Barrier Reef, Australia (14°41′S, 145°27′E). I captured large numbers (ca. 100) of larval Ward's damselfish, *Pomacentrus wardi*, that were in the process of settling to the reef, using light traps anchored just outside the reef crest (Meekan et al. 2001). Fish were caught overnight, and at dawn were brought back to the laboratory by boat in a large aerated bin, where they were held together in a 100-litre aquarium with fresh flow-through sea water at ambient temperature (ca. 28 °C) and live *Artemia* food until focal animals were selected later that morning (see below). At that point it was evident that all fish had undergone metamorphosis, indicated by the adoption of juvenile coloration and shape. Fish not used in the experiment were released at noon onto the reef adjacent to where they were captured. All research was conducted under permits from the Great Barrier Reef Marine Park Authority and James Cook University Animal Ethics Committee.

I randomly selected 30 individuals with similar body size (mean standard length = 12.9 mm, range 12.7–13.5 mm) and placed each fish into its own plastic aquarium by noon that same day. Each aquarium (25 × 16 cm and 17 cm high, filled to a depth of 10 cm) contained a layer of sand on the bottom and a small 'T'-shaped plastic pipe connector placed against the far wall, such that the three openings faced forward. Aquaria were visually isolated from one another and from the observer using plastic sheeting. Fish were fed recently hatched (<24 h old) live *Artemia* nauplii up to three times per day throughout the experiment to ensure ad libitum food, visually confirmed to be swimming about the aquaria. Without moving aquaria, I used a siphon to change 80% of the water at the end of every second day with fresh, temperature-adjusted sea water. Artificial light was provided on a 13:11 h light:dark regime matching outside light conditions. Fish were euthanized (confirmed by cessation of opercular beats for 5 min) using an

overdose of the anaesthetic Aqui-S (200 mg/litre), because these fish were to be further used for otolith analysis.

The morning following introduction of fish to the home tanks, I trained the few fish that did not make use of the shelter when startled. I did this by startling them until they learned to retreat to the shelter following this simulated predatory attack (described below). All fish were startled on each pass, to standardize the frequency of exposures. All fish used the shelter after three training passes over a 20 min period (I did not record which fish required training). Fish were then left undisturbed for 3 h prior to the first recorded behavioural observations in the afternoon (1400 hours). Water temperature was maintained at an average of 29.0 °C (range 28.4–29.5 °C, $N = 780$); the temperature in each aquarium was measured after morning and afternoon behavioural observations.

I quantified fish activity and boldness within their home tanks twice each day: once in the morning (0900–1130 hours) and once in the afternoon (1400–1630 hours). Within each observation period, I allowed approximately 1 h ± 15 min to elapse between scoring one behavioural trait (activity) and another (boldness) for that same individual; activity was always measured first, and fish were always observed in the same order. I observed fish over a 7-day period, yielding 13 (boldness) and 11 (activity) observations per individual (total sample size ca. $N = 780$ per trait). A single afternoon observation was made on day 1; other commitments reduced time available to gather the full sample of activity assays.

Activity was simultaneously estimated in two ways: (1) by quantifying the number of times the fish crossed over the midpoint of the aquarium during a 2 min observation period, and (2) as the proportion of time spent active during that period; movement was defined as a displacement of more than 0.5 times its body length. Although it is possible this relatively short observation might not be representative of fish activity over longer intervals, this seemed not to be a major problem as individuals consistently differed in their patterns of activity following acclimation (see Results). Boldness was estimated as the latency for fish to emerge from shelter following a simulated predation attempt, whereby I rapidly dipped an aquarium dip net handle into the centre of the aquarium. Fish not emerging from shelter within 3 min were assigned a latency value of 180 s.

Visual inspection of individual data plots of boldness over time revealed rather abrupt changes in boldness occurred during the initial three to five observations. I therefore specified a general linear model with mixed effects (fixed and random; Proc Mixed, SAS Institute, Cary, NC, U.S.A.) incorporating an interaction effect between observation number and a break point in the data. I did so by creating a coded dummy variable for a transition point and tested the following model:

$$Y = \text{int} + \text{obs} + \text{break} + \text{obs} \times \text{break}$$

where $Y$ is an observation of a given behaviour for a particular individual, 'int' is the intercept, 'obs' is the observation number, and 'break' specifies a shift in behaviour (e.g. Fig 1b illustrates a break point between observation number 4 and 5). I fitted several likely break points (between 2 and 3, 3 and 4 and 4 and 5) to determine the most likely one given the data (see below). All of these factors, including the interaction effect, were specified as random effects, resulting in a complex model with random slopes and intercepts that generates individual-specific intercepts and slopes (i.e. so-called 'BLUPs', best linear unbiased predictors) both before and after this transition point. Note that this method generates individual-specific predictions, but does not fit individual-specific parameters. Rather, it allows a parameter to have an associated variance that is estimated from repeated measurements on a set of individuals (see Singer & Willett 2003; West et al. 2010). I started with a saturated model that included all random effects (variance parameters and all covariance parameters describing correlations between them) and compared model fit to progressively simpler models (see Supplementary material for annotated model code). I also incorporated temperature, time of day (morning versus afternoon) and size as covariates (fixed factors), but none of these effects were significant (all $P > 0.2$), and consequently were removed from the model before refitting. I used the Kenward–Roger method to calculate degrees of freedom for the fixed effects, using a type III approach. Models were fitted by maximum likelihood (ML, not restricted ML) because I wished to compare fit between models containing different random and fixed effect structures, and to identify the most likely transition ('break') point given the data (Singer & Willett 2003). Model fit was assessed using AICc values, comparing a model with and without a given effect, whereby smaller values indicate better fit and two competing models that differ by less than a few AICc units (typically $\Delta\text{AICc} < 4$) are considered equally likely, and therefore not significantly different (Burnham & Anderson 1998). To facilitate comparison with other studies that do not consider individual differences in acclimation, I also calculated repeatability values for data during the acclimation period (prior to the break point), after and for the entire series of observations. Repeatability ($r$) was estimated as the variance in individual intercept values as a proportion of the total variance ($\text{var}_{\text{int}}/\text{var}_{\text{int}} + \text{var}_{\text{resid}}$; model $Y = \text{int} + \text{obs}$, with only the intercept specified as a random effect (Singer & Willett 2003). This repeatability measure tells us the proportion of variance accounted for by individual differences. Again, I used AICc values to compare models with and without the random intercept effect.

Unlike the boldness data, neither measure of activity showed any obvious pattern of acclimation. Therefore, I fitted mixed models ($Y = \text{int} + \text{obs}$) specifying only the intercept as a random effect. To facilitate comparison with other studies, and the analysis on boldness data, I calculated repeatability values during and after the acclimation period determined in the previous analysis for the boldness data. The number of crosses of the aquarium mid-line was $\log(x + 1)$ transformed, and proportion of time spent moving was arcsine square-root transformed. These two measures of activity provide somewhat redundant information and so I present graphs of the former measure of activity only.

## RESULTS

Patterns of acclimation in boldness dramatically differed across individuals (Fig. 2). As a result, initial assays were not rank order consistent with later observations because individuals consistently differed in responses over time, as evidenced by significant random effects for each of the parameters, including the interaction between the break point and observation number ($\Delta\text{AICc} = 13$ comparing models with and without the random interaction effect; Fig. 2). Presence of this interaction confirmed that rank order was not maintained across apparent familiar and unfamiliar conditions. Most individuals showed large differences in their patterns of behaviour during the first three to four observations (ca. 2 days), followed by relatively stable patterns of behaviour in subsequent observations (Fig. 2). The data indicated a significant shift in behaviour between observations 3 and 4, although a shift between observations 4 and 5 was almost equally likely ($\Delta\text{AICc} = 1$). Inspection of model fits to the data indicated this model did a good job of capturing the obvious patterns in the data (Fig. 2). None the less, for subsequent analysis of repeatability I divided the data set into observations 1–4 and 5–13 to be conservative. As a result of the large interindividual differences in behavioural patterns observed during the first four observations, latency was not

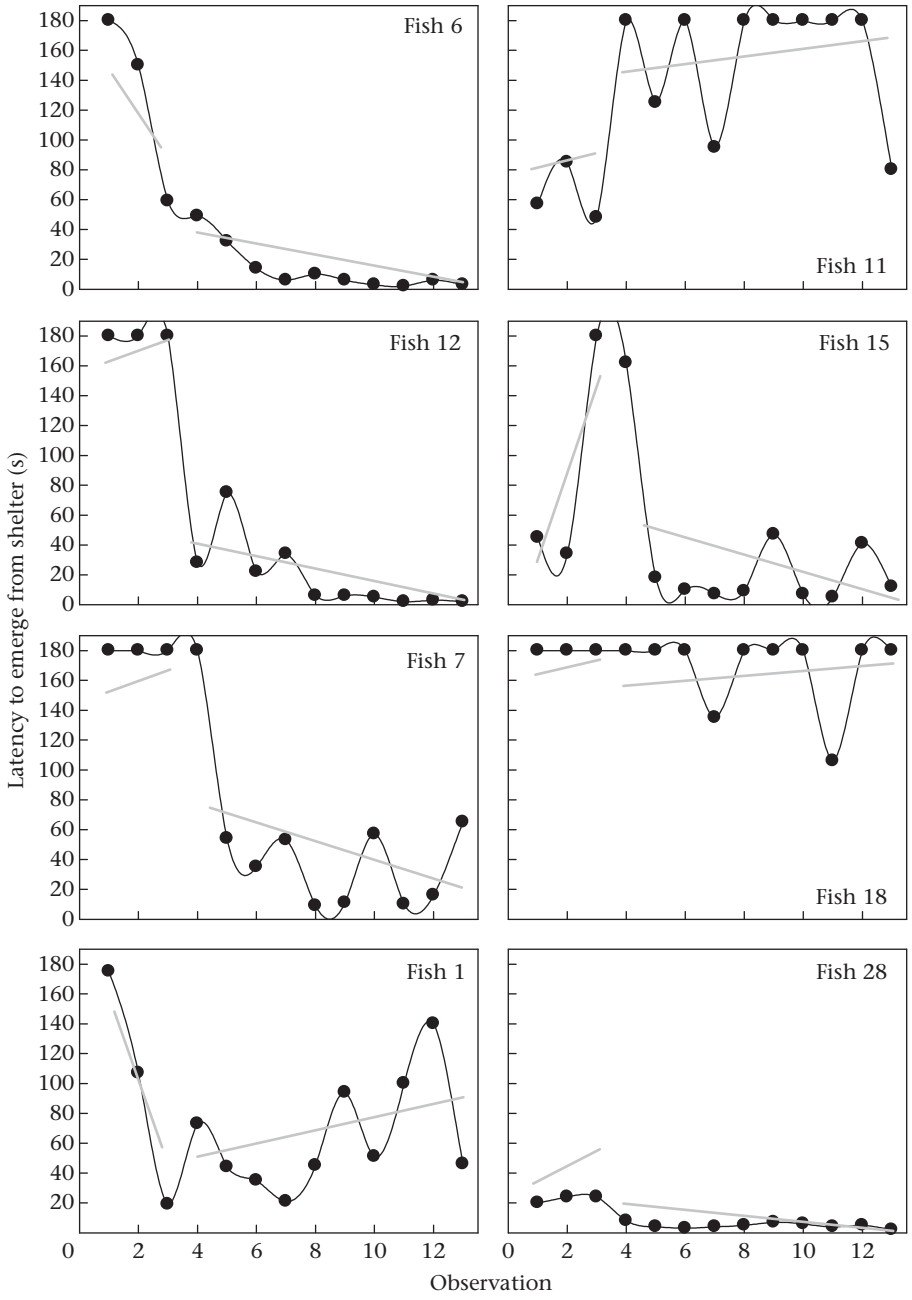


**Figure 2.** Latency to emerge from shelter following a simulated predation attempt (i.e. boldness). The figures show eight individuals selected to illustrate the range of boldness responses over time. Observation 1 represents an afternoon observation on the first day, followed by one morning and one afternoon observation each day thereafter for a total of 7 days. A smoothing spline joins together successive observations to highlight the temporal changes in the behaviour, and the grey line segments represent the individual-specific model predictions (i.e. BLUPs). These results indicate an even more complex acclimation pattern than that initially suggested in Fig. 1a. Note the small but consistent bias in model predictions when observed values are at the maximum of 180 s.

repeatable during this interval ($r = 0.18$, $\Delta AICc = 2$). Following this interval, behaviour was highly repeatable ($r = 0.62$, $\Delta AICc > 100$). Not surprisingly, repeatability was much reduced when I ignored acclimation and used all the observations ($r = 0.38$, $\Delta AICc > 100$).

There was no obvious pattern that would suggest acclimation over time in either measure of activity, and therefore no reason to fit complex models as done for the boldness data (see Fig. 3; graph for time spent moving not presented for reasons of brevity, see Methods). In contrast, there was rather high within-individual variation in activity levels (Fig. 2 versus 3). To facilitate comparison with the boldness data, I also evaluated repeatability during initial (1–4) and subsequent observations (5–13): neither measure of activity was repeatable across observations 1–4 (both $r < 0.15$, $\Delta AICc < 1$ in each case), but repeatability for the number of crosses (0.25) and for proportion of time spent moving (0.35) was significant for subsequent observations ($\Delta AICc > 20$ in each case). When I considered all observations, repeatability was significant but substantially lower for crosses ($r = 0.15$, $\Delta AICc = 24$) and for proportion of time spent moving ($r = 0.24$, $\Delta AICc = 43$).

## DISCUSSION

My experiment indicated that in the case of boldness, fish displayed unique individual-specific patterns of acclimation that were

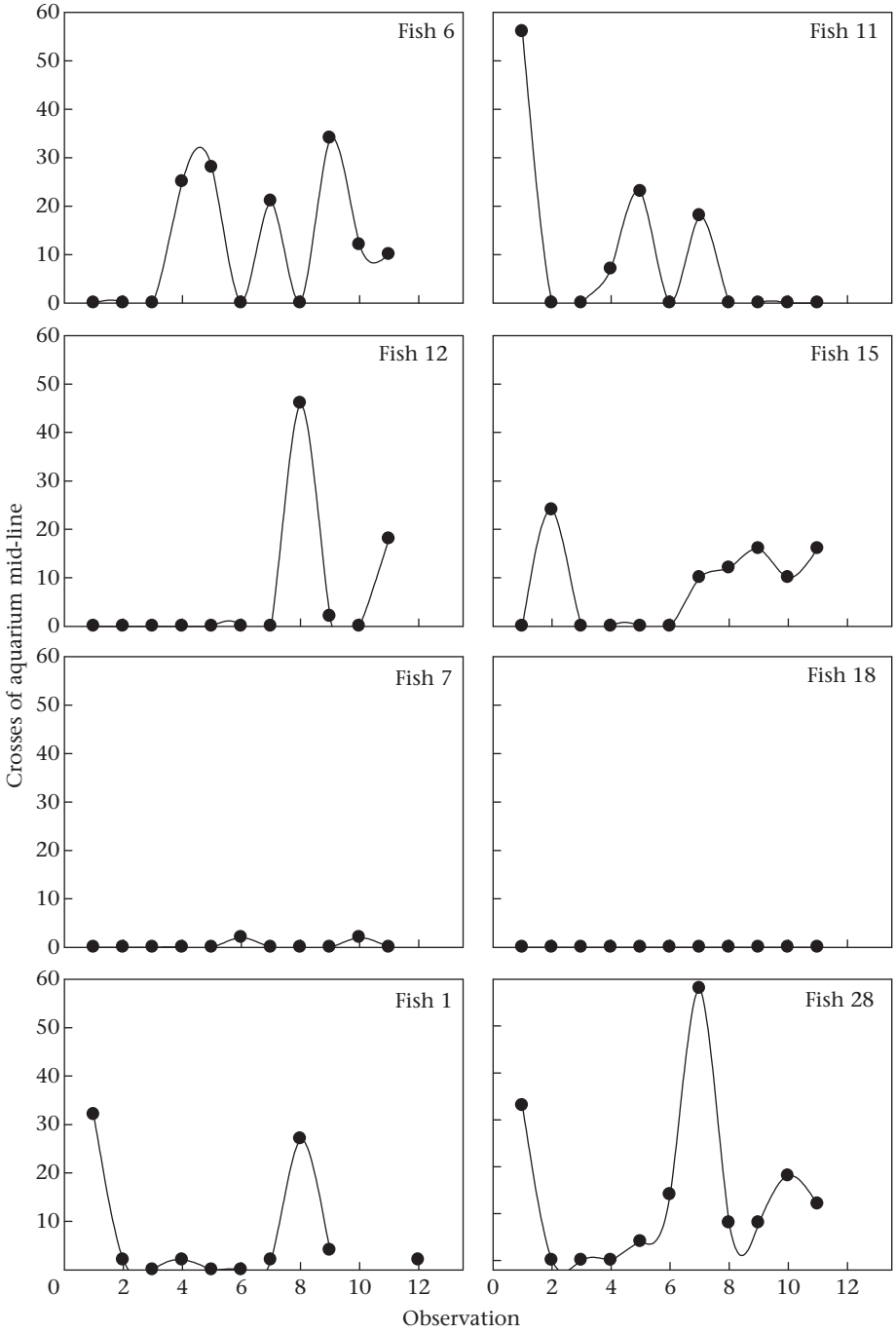


**Figure 3.** Total number of times that the focal fish crossed the aquarium mid-line during an observation. The same individuals as in Fig. 2 are shown to facilitate comparison of individuals across behavioural traits; observation numbers follow exactly the same time course as described in Fig. 2. Note that analysis was done on transformed data.

even more complex than illustrated in Fig. 1 (i.e. some habituated, some were sensitized, and some showed no change). As a result, initial assays were not rank order consistent with those observed when conditions were more familiar. Following an acclimation interval spanning about four observations (over 2 days), repeatability was subsequently fairly high (0.62), among the highest 25% or so of values observed in behavioural studies (Bell et al. 2009). In other words, these two analyses together indicate that response to simulated predation is repeatable, and therefore a trait, but behavioural types could not be properly characterized using a rapid assay approach with a few initial observations.

In the case of activity, there were no obvious patterns of acclimation within individuals. Given that within-individual variation was relatively higher for activity assays than for boldness, this may have obscured an acclimation process (if present) during early observations. None the less, there was significant but low repeatability in both measures of activity ($r = 0.25–0.35$) following the first four observations, but not during them. Thus, a rapid assay approach would not have identified activity as a trait.

Together, these results suggest that studies that employ rapid behavioural assays may be misclassifying behavioural types, at least to some extent, and underestimating repeatability. If so,

then this might reduce power to detect relationships between behavioural traits and other variables of interest, such as those related to fitness. By extension, I would expect these results to extrapolate over to physiological traits given that behaviour is often closely correlated with metabolism and endocrine hormone levels (e.g. Carlson 1986; see also Introduction). Of course, this is but one cautionary study, and so we will need more studies to determine the generality of these findings. However, the more domesticated and/or familiar the species is with life in the laboratory, the more likely animals will quickly acclimate and/or habituate to procedures forced upon them. Hence, rapid assays might work perfectly fine in those circumstances, but reseachers should verify that this is so. Similarly, older animals may react less to stressors and be less variable in their behaviour than juveniles (Stamps & Groothuis 2010), making rapid assays sufficiently accurate for adults. Again, we should be careful and not assume this is the case, and I suggest that researchers consider conducting pilot studies to determine appropriate acclimation periods, and/or housing animals in home cages and conducting tests within them to reduce handling and stress. Most importantly, future studies should extend the approach used here and evaluate whether laboratory assays under familiar conditions predict similar traits when observed in the field.

My results also lead me to suggest that future studies gather many more observations per individual than is usually done. I suggest 10 or more observations per animal, to (1) identify adequate acclimation periods if individuals differ in acclimation patterns and (2) rigorously characterize an individual's behavioural and/or physiological type in the face of even moderate within-individual variation (see also Martin et al. 2011) and/or the presence of any systematic temporal changes not related to acclimation (see Fig. 2). Unfortunately, at present we have few data to inform us about the size of within-individual variation in behavioural traits, because few studies collect more than a few data points per animal (see Introduction). However, even with substantial repeated observations, the data on activity rates shown here illustrate that 'significant' repeatability and the behavioural types characterized from these analyses should be cautiously interpreted when within-individual variation is high. This issue becomes more obvious when one considers that a 'moderate' repeatability value, estimated as a correlation coefficient between two measurements (e.g. Pearson $r = 0.3$), is actually a rather weak and inconsistent association (equates to an $R^2$ of only 0.09). Complicating matters further is the possibility that individuals might also consistently differ in their within-individual (i.e. residual) variability (see Figs 2, 3; also Eriksson et al. 2010).

In conclusion, we should not assume that rank order differences across individuals are maintained between (forced) novel and familiar conditions, nor assume that within-individual variation is small and by extension assume that one, two or three rapid assays per individual can suffice to characterize a measured response as a trait. Of course, gathering many samples per animal may not be feasible for some physiological or hormonal assays, but it certainly would be for behaviour.

Thanks to Mark McCormick for assistance with sampling and facilitating the research at Lizard Island through James Cook University, and to Judy Stamps and two anonymous referees for comments on the manuscript. Funding was provided by an ARC Future Fellowship.

## Supplementary Material

Supplementary data related to this article can be found online at doi:10.1016/j.anbehav.2012.01.036.

## References

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

**Biro, P. A. & Stamps, J. A.** 2010. Do consistent individual differences in metabolic rate promote consistent individual differences in behavior? *Trends in Ecology & Evolution*, **25**, 653–659.

**Boon, A. K., Réale, D. & Boutin, S.** 2007. The interaction between personality, offspring fitness and food abundance in North American red squirrels. *Ecology Letters*, **10**, 1094–1104.

**Boratynski, Z. & Koteja, P.** 2009. The association between body mass, metabolic rates and survival of bank voles. *Functional Ecology*, **23**, 330–339.

**Budaev, S. V.** 1997. 'Personality' in the guppy (*Poecilia reticulata*): a correlational study of exploratory behavior and social tendency. *Journal of Comparative Psychology*, **111**, 399–411.

**Burnham, K. P. & Anderson, D. R.** 1998. *Model Selection and Inference: a Practical Information-theoretic Approach*. New York: Springer Verlag.

**Careau, V., Thomas, D., Humphries, M. M. & Reale, D.** 2008. Energy metabolism and animal personality. *Oikos*, **117**, 641–653.

**Careau, V., Réale, D., Humphries, Murray M. & Thomas, Donald W.** 2010. The pace of life under artificial selection: personality, energy expenditure, and longevity are correlated in domestic dogs. *American Naturalist*, **175**, 753–758.

**Carlson, N. R.** 1986. *Physiology of Behavior*. 3rd edn. Needham Heights, Massachusetts: Allyn & Bacon.

**Ellenberg, U., Mattern, T. & Seddon, P. J.** 2009. Habituation potential of yellow-eyed penguins depends on sex, character and previous experience with humans. *Animal Behaviour*, **77**, 289–296.

**Eriksson, C. A., Booth, D. J. & Biro, P. A.** 2010. 'Personality' in two species of temperate damselfish. *Marine Ecology Progress Series*, **420**, 273–276.

**Gosling, S. D.** 2001. From mice to men: what can we learn about personality from animal research? *Psychological Bulletin*, **127**, 45–86.

**Ksiazek, A., Konarzewski, M. & Lapo, I. B.** 2004. Anatomic and energetic correlates of divergent selection for basal metabolic rate in laboratory mice. *Physiological and Biochemical Zoology*, **77**, 890–899.

**Martin, J. G. A. & Reale, D.** 2008. Temperament, risk assessment and habituation to novelty in eastern chipmunks, *Tamias striatus*. *Animal Behaviour*, **75**, 309–318.

**Martin, J. G. A., Nussey, D. H., Wilson, A. J. & Réale, D.** 2011. Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression models. *Methods in Ecology and Evolution*, **4**, 362–374.

**Meekan, M. G., Wilson, S. G., Halford, A. & Retzel, A.** 2001. A comparison of catches of fishes and invertebrates by two light trap designs, in tropical NW Australia. *Marine Biology*, **139**, 373–381.

**Nespolo, R. F. & Franco, M.** 2007. Whole-animal metabolic rate is a repeatable trait: a meta-analysis. *Journal of Experimental Biology*, **210**, 2000–2005.

**Overli, O., Pottinger, T. G. & Winberg, S.** 2005. Behavioral and neuroendocrine correlates of artificial selection for stress responsiveness in rainbow trout: a review. *Integrative and Comparative Biology*, **45**, 463–474.

**Reale, D., Gallant, B. Y., Leblanc, M. & Festa-Bianchet, M.** 2000. Consistency of temperament in bighorn ewes and correlates with behaviour and life history. *Animal Behaviour*, **60**, 589–597.

**Romero, L. M.** 2004. Physiological stress in ecology: lessons from biomedical research. *Trends in Ecology & Evolution*, **19**, 249–255.

**Rodríguez-Prieto, I., Martín, J. & Fernández-Juricic, E.** 2010. Habituation to low-risk predators improves body condition in lizards. *Behavioral Ecology and Sociobiology*, **64**, 1937–1945.

**Rodríguez-Prieto, I., Martín, J. & Fernández-Juricic, E.** 2011. Individual variation in behavioural plasticity: direct and indirect effects of boldness, exploration and sociability on habituation to predators in lizards. *Proceedings of the Royal Society B*, **278**, 266–273.

**Sih, A. & Bell, A. M.** 2008. Insights for behavioral ecology from behavioral syndromes. *Advances in the Study of Behavior*, **38**, 227–281.

**Sih, A., Bell, A. M., Johnson, J. C. & Ziemba, R. E.** 2004. Behavioural syndromes: an integrative overview. *Quarterly Review of Biology*, **79**, 241–277.

**Singer, J. D. & Willett, J. B.** 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

**Stamps, J. A. & Groothuis, T. G. G.** 2010. The development of animal personality: relevance, concepts and perspectives. *Biological Reviews*, **85**, 301–325.

**West, S. G., Ryu, E., Kwok, O. M. & Cham, H.** 2010. Multilevel modeling: current and future applications in personality research. *Journal of Personality*, **79**, 2–50.

**Williams, T. D.** 2008. Individual variation in endocrine systems: moving beyond the 'tyranny of the Golden Mean'. *Philosophical Transactions of the Royal Society B*, **363**, 1687–1698.

**Wilson, A. D. M., Godin, J.-G. J. & Ward, A. J. W.** 2009. Boldness and reproductive fitness correlates in the eastern mosquitofish, *Gambusia holbrooki*. *Ethology*, **116**, 96–104.

**Wong, K., Elegante, M., Bartels, B., Elkhayat, S., Tien, D., Roy, S., Goodspeed, J., Suciu, C., Tan, J., Grimes, C., et al.** 2010. Analyzing habituation responses to novelty in zebrafish (*Danio rerio*). *Behavioural Brain Research*, **208**, 450–457.