



## Forum

## Do rapid assays predict repeatability in labile (behavioural) traits? A reply to Biro

Hannah A. Edwards<sup>a,\*</sup>, Isabel S. Winney<sup>a</sup>, Julia Schroeder<sup>a,d</sup>, Hannah L. Dugdale<sup>a,b,c</sup>

<sup>a</sup> Department of Animal and Plant Sciences, University of Sheffield, Sheffield, U.K.

<sup>b</sup> Theoretical Biology, Centre for Ecological and Evolutionary Studies, University of Groningen, Groningen, The Netherlands

<sup>c</sup> Behavioural Ecology and Self-organization, Centre for Ecological and Evolutionary Studies, University of Groningen, Groningen, The Netherlands

<sup>d</sup> Evolutionary Biology, Max Planck Institute for Ornithology Seewiesen, Germany

### ARTICLE INFO

#### Article history:

Received 13 July 2012

Initial acceptance 20 August 2012

Final acceptance 13 December 2012

Available online 31 January 2013

MS. number: 12-00541R

#### Keywords:

acclimation  
activity  
boldness  
exploratory behaviour  
habituation  
individual variation  
novel environment  
personality  
reaction norm  
repeatability

When individuals react differently to a stimulus, and these behavioural trait differences are repeatable (consistent) across time and contexts, the variation is known as personality (e.g. [Briffa & Weiss 2010](#)). Using data from successive behavioural assays, [Biro \(2012\)](#) showed that within-individual behavioural responses during initial tests in a relatively unfamiliar environment were not repeatable. In contrast, later tests conducted in the same environment, which was relatively more familiar, produced repeatable behaviours. Biro therefore suggested that a large number of observations ( $\geq 10$ ) per individual are required to acclimate and characterize labile behavioural states rigorously. We believe that Biro's assay unintentionally conflates different types of behavioural traits. This issue is not discussed in the commentary. We also propose that rapid repeated

behavioural assays habituate individuals and the results cannot be extrapolated to assaying in a novel and unfamiliar environment. Finally, we suggest that Biro's recommended number of repeat samples per individual is often not logistically feasible in wild populations.

A behavioural trait can be measured along an axis with high and low estimates of the trait represented as extremes. [Réale et al. \(2007\)](#) suggested that behavioural traits can be divided into five guideline categories: (1) shy–bold: an individual's response to a risky situation in a familiar environment; (2) exploration–avoidance: an individual's response to a novel situation, such as an unfamiliar environment or object; (3) activity: the activity level of an individual in a nonrisky and familiar environment; (4) aggressiveness: an individual's pugnacity towards conspecifics in a social setting; and (5) sociability: an individual's nonaggressive tendency to seek or avoid conspecifics. As all these axes are sensitive to the context in which they are measured, experiments must be designed to ensure that the intended behavioural axis is assayed ([Carter et al. 2012b](#); [Réale et al. 2007](#)).

\* Correspondence: H. A. Edwards, Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, U.K.

E-mail address: [hannah892@hotmail.com](mailto:hannah892@hotmail.com) (H. A. Edwards).

Biro's assays measured boldness and activity in an increasingly familiar context as testing progressed. Therefore, the activity assay conflated exploration–avoidance and activity behaviours. Boldness and activity were measured twice a day over 7 days in Ward's damselfish, *Pomacentrus wardi*, and individuals were ranked in their response to risk in an environment that became increasingly familiar through acclimatization. Boldness was scored via a startle latency response and activity by the number of aquarium crossings and movement time. Boldness was not repeatable between a relatively unfamiliar and a more familiar environment, that is, before versus after three to four observations. Activity was not repeatable in relatively unfamiliar settings but was repeatable in more familiar settings, that is, after four observations. It has been documented that behavioural traits thought to stem from the same behavioural axis may in fact stem from different axes when sampled in different contexts (Carter et al. 2012a). It is therefore important to ensure that behavioural traits and the contexts in which they are expressed are fully defined. Any underlying assumptions affecting the behavioural trait must be supported before testing and considered during interpretation.

Associative learning, causing habituation to an increasingly familiar stimulus, further confounds behavioural trait measurements in changing contexts. There are two points in Biro's study where habituation may have affected the conclusions drawn from the boldness assay: during the initial training period and during repeated rapid assaying. Before the study, Biro had to train some of the damselfish to retreat to a shelter after a simulated attack. Therefore, Biro's boldness scores, after repeated training and sampling, might not reflect boldness, because of habituation to the stimulus. Biro also revealed that averaging test scores may lead to incorrect conclusions about the behavioural traits of individuals. However, he did not consider the likelihood of habituation or that individuals can differ in their rate of habituation and learning (Glowa & Hansen 1994; Gallistel et al. 2004). We therefore consider it vital to consider context, habituation and individual plasticity in habituation responses when designing and drawing conclusions from behavioural assays.

One way of minimizing the potential for the habituation of individuals to novelty might be to use longer intermeasure intervals between assays. In Biro's laboratory study the interval length was 3–17 h. Longer intervals could help reduce the possibility of non-associative learning causing habituation. For example, a shorter interval length between successive exploration tests, across four great tit, *Parus major*, populations, caused birds to become faster explorers (Dingemans et al. 2012). The intermeasure interval length in studies of wild populations is usually less controllable and often considerably longer (potentially years) than in laboratory populations, owing to unpredictable recapture events. Clearly, the experimental design must statistically consider the intersampling interval in the study of both wild and captive populations.

In addition to the effect of interval length, behavioural responses towards the same stimulus may differ when measured in the wild versus the laboratory. Biro correctly highlights that 'the implicit assumption of any laboratory study of a labile trait, such as behaviour, is that its expression is a good predictor of trait expression of those same individuals under more familiar, less stressful, and therefore more natural conditions' (pp. 1295–1296). We generally agree with this notion, although we note that the measurement of the exploration–avoidance axis relies on measuring responses to novel environments. Such tests will naturally involve an element of stress that an individual would not encounter in a familiar environment. The stress response is therefore an element of the exploration response to a novel environment. An empirical study by Herborn et al. (2010) supports this idea by showing that exploration behaviour in a novel environment positively predicted analogous exploration measurements in the wild.

Additionally, individuals sampled in wild populations will have experienced greater variation in their environment than individuals raised in the laboratory. Greater environmental variation coupled with processes such as stimulus generalization may act to habituate wild individuals to stress-imposed responses (Bell et al. 2009; A. Weiss, personal communication). This could be the reason why captive-bred birds and wild birds of the same species react differently to the same novelty tests (E. Schlicht & J. Partecke, personal communication). To support these ideas better, more research is needed to compare behavioural responses of wild and captive animals.

To advance our knowledge of the biology of personality traits we must not only ensure comparability between wild and captive studies but also obtain accurate and precise results. For this, we need to define precisely the quantity and type of data required. Limited resources will inevitably produce a trade-off between the number of individuals measured and the number of measurements per individual. Biro's rapid assay used mixed models to estimate behavioural reaction norms. These models use repeated measures to describe how personalities change over time (Dingemans et al. 2010). Specifically, Biro identified a break point that divided the reaction norm into two periods, probably because of acclimation. Biro recommends that future studies should use at least 10 observations to identify this acclimation point and rigorously characterize the reaction norm. Sensitivity analyses suggest that a ratio of individuals to observations of  $>0.5$  with sample sizes over 200 provides the power to detect individual variation in the slopes and intercepts of reaction norms (Martin et al. 2011; van de Pol 2011). However, to generate accurate and precise estimates, the ratio of individuals to observations should be  $\geq 0.5$ , with 40 or more individuals and a total sample size of at least 1000 observations (Martin et al. 2011; van de Pol 2011). When these criteria are achieved, the ratio of individuals to observations can be biased towards more individuals (van de Pol 2011); however, when repeatability is low the ratio should be biased towards more observations (Garamszegi & Herczeg 2012).

The trade-off between the number of individuals measured and the number of measurements per individual is accentuated in the study of wild animals because of the logistical and administrative constraints. It is often not feasible, and may even be harmful, to collect multiple experimental measurements per individual. A meta-analysis of the repeatability of behavioural measures (see supplementary material in Bell et al. 2009) highlights this possible difficulty. It demonstrated that more repeated measures were usually obtained in the laboratory (mean = 5.7, 95% confidence interval, CI = 4.8–6.6) than in the field (mean = 2.4, 95% CI = 2.3–2.5). The majority of laboratory and field studies sampled individuals just twice, although some studies sampled individuals over 60 times, resulting in a mode of 2 and a mean of 4.4 (95% CI = 3.8–5.0) measures per individual over 759 studies. This implies for wild populations that Biro's suggestion of 10 measures per individual is often logistically unfeasible.

In summary, we believe that Biro's study raises a valid and cautionary note about the misclassification of individual behavioural types and inaccurate estimation of the repeatability of behavioural traits. We highlight that Biro's suggested 10 measures per individual to allow for acclimation is confounded by the habituation effect resulting from associative learning. This consequently conflates behavioural axes because of changing contexts. We question whether Biro's recommendation can improve behavioural trait assessment, especially in wild populations. Carefully designed empirical studies are needed to prevent misclassification of behavioural traits (Carter et al. 2012b; Campbell & Fiske 1959) when traits are correlated and context dependent, and combined field and laboratory studies will help to validate comparability. This

could reveal important patterns and improve our understanding of, for instance, how stable behavioural syndromes are across contexts and time and the impact of habituation and frequency of inter-measure intervals on repeatability estimates.

This research was supported by a U.K. Natural Environment Research Council (NERC) studentship (H.A.E.), a University of Sheffield Scholarship (I.S.W.), NERC grant NE/F006071/1 to Terry Burke by the Volkswagen foundation (J.S.), a NERC fellowship NE/I021748/1 (H.L.D.) and a Netherlands Organisation for Scientific Research Visitor Travel Grant (H.L.D.). I.S.W. had the idea for the article, H.A.E. wrote it, and H.A.E., I.S.W., J.S. and H.L.D. discussed and revised it. We thank Peter Biro, Alexander Weiss, an anonymous referee and Terry Burke for their helpful comments.

## References

- Bell, A. M., Hankison, S. J. & Laskowski, K. L. 2009. The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, **77**, 771–783, <http://dx.doi.org/10.1016/j.anbehav.2008.12.022>.
- Biro, P. A. 2012. Do rapid assays predict repeatability in labile (behavioural) traits? *Animal Behaviour*, **83**, 1295–1300, <http://dx.doi.org/10.1016/j.anbehav.2012.01.036>.
- Briffa, M. & Weiss, A. A. 2010. Animal personality. *Current Biology*, **20**, R912–R914.
- Campbell, D. T. & Fiske, D. W. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, **56**, 81–105, <http://dx.doi.org/10.1037/h0046016>.
- Carter, A. J., Marshall, H. H., Heinsohn, R. & Cowlshaw, G. 2012a. How not to measure boldness: novel object and antipredator responses are not the same in wild baboons. *Animal Behaviour*, **84**, 603–609, <http://dx.doi.org/10.1016/j.anbehav.2012.06.015>.
- Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlshaw, G. & Heinsohn, R. 2012b. Animal personality: what are behavioural ecologists measuring? *Biological Reviews*, published online 18 December 2012. <http://dx.doi.org/10.1111/brv.12007>.
- Dingemanse, N. J., Kazem, A. J. N., Réale, D. & Wright, J. 2010. Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, **25**, 81–89, <http://dx.doi.org/10.1016/j.tree.2009.07.013>.
- Dingemanse, N. J., Bouwman, K. M., van de Pol, M., Van Overveld, T., Patrick, S. C., Matthysen, E. & Quinn, J. L. 2012. Variation in personality and behavioural plasticity across four populations of the great tit *Parus major*. *Journal of Animal Ecology*, **81**, 116–126, <http://dx.doi.org/10.1111/j.1365-2656.2011.01877.x>.
- Glowa, J. R. & Hansen, C. T. 1994. Differences in response to an acoustic startle stimulus among forty-six rat strains. *Behavior Genetics*, **24**, 79–84, <http://dx.doi.org/10.1007/BF01067931>.
- Gallistel, C. R., Fairhurst, S. & Balsam, P. 2004. The learning curve: implications of a quantitative analysis. *Proceedings of the National Academy of Sciences, U.S.A.*, **101**, 13124–13131, <http://dx.doi.org/10.1073/pnas.0404965101>.
- Garamszegi, L. s. Z. & Herczeg, G. 2012. Behavioural syndromes, syndrome deviation and the within- and between-individual components of phenotypic correlations: when reality does not meet statistics. *Behavioral Ecology and Sociobiology*, **66**, 1651–1658, <http://dx.doi.org/10.1007/s00265-012-1439-8>.
- Herborn, K. A., Macleod, R., Miles, W. T. S., Schofield, A. N. B., Alexander, L. & Arnold, K. E. 2010. Personality in captivity reflects personality in the wild. *Animal Behaviour*, **79**, 835–843, <http://dx.doi.org/10.1016/j.anbehav.2009.12.026>.
- 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*, **2**, 362–374, <http://dx.doi.org/10.1111/j.2041-210X.2010.00084.x>.
- van de Pol, M. 2011. Quantifying individual variation in reaction norms: how study design affects the accuracy, precision and power of random regression models. *Methods in Ecology and Evolution*, **3**, 268–280, <http://dx.doi.org/10.1111/j.2041-210X.2011.00160.x>.
- Réale, D., Reader, S. M., Sol, D., McDougall, P. T. & Dingemanse, N. J. 2007. Integrating animal temperament within ecology and evolution. *Biological Reviews*, **82**, 291–318, <http://dx.doi.org/10.1111/j.1469-185X.2007.00010.x>.