

## COMMENTARY

# Demystifying animal ‘personality’ (or not): why individual variation matters to experimental biologists

Dominique G. Roche<sup>1,\*</sup>, Vincent Careau<sup>2</sup> and Sandra A. Binning<sup>1</sup>

## ABSTRACT

Animal ‘personality’, defined as repeatable inter-individual differences in behaviour, is a concept in biology that faces intense controversy. Critics argue that the field is riddled with terminological and methodological inconsistencies and lacks a sound theoretical framework. Nevertheless, experimental biologists are increasingly studying individual differences in physiology and relating these to differences in behaviour, which can lead to fascinating insights. We encourage this trend, and in this Commentary we highlight some of the benefits of estimating variation in (and covariation among) phenotypic traits at the inter- and intra-individual levels. We focus on behaviour while drawing parallels with physiological and performance-related traits. First, we outline some of the confusion surrounding the terminology used to describe repeatable inter-individual differences in behaviour. Second, we argue that acknowledging individual behavioural differences can help researchers avoid sampling and experimental bias, increase explanatory power and, ultimately, understand how selection acts on physiological traits. Third, we summarize the latest methods to collect, analyse and present data on individual trait variation. We note that, while measuring the repeatability of phenotypic traits is informative in its own right, it is only the first step towards understanding how natural selection and genetic architecture shape intra-specific variation in complex, labile traits. Thus, understanding how and why behavioural traits evolve requires linking repeatable inter-individual behavioural differences with core aspects of physiology (e.g. neurophysiology, endocrinology, energy metabolism) and evolutionary biology (e.g. selection gradients, heritability).

**KEY WORDS:** Behaviour, Phenotypic variation, Physiology, Reaction norm, Repeatability, Temperament

## Introduction

Anyone working with live animals can hardly dispute the fact that striking behavioural differences exist among individuals. This variation can dramatically, and often predictably, influence how individuals interact with their environment, and can affect the outcome of biotic interactions such as predation, competition, parasitism, cooperation and mate choice (Wolf and Weissing, 2012). However, it has taken over a century for researchers working on non-human animals to appreciate the magnitude as well as the ecological and evolutionary importance of behavioural variation. Similarly, research on inter-individual phenotypic variation (see

Glossary), including consistent individual differences in physiological traits, endocrine responses and performance measures, has been sporadic throughout the second half of the last century (e.g. Tryon, 1942; Huntingford, 1976; Bennett, 1980; Stevenson-Hinde et al., 1980; Arnold and Bennett, 1984; Clark and Ehlinger, 1987; Wilson et al., 1993, 1994). Bennett’s (1987) suggestion that physiological ecologists move away from the ‘tyranny of the golden mean’, and instead focus on understanding the functional basis and eco-evolutionary implications of performance differences among individuals instigated a shift in research focus that has since been echoed across various fields in biology (Gould, 1985; Boake, 1989; Hayes and Jenkins, 1997; Kolok, 1999; Chown, 2001; Bolnick et al., 2003; Lloyd-Smith et al., 2005; Crawford and Oleksiak, 2007). This emphasis on individual-level variation has perhaps been most strongly felt in the field of behavioural ecology. Indeed, the past 10–15 years have witnessed an explosion of studies on animal ‘personality’ (e.g. Gosling, 2001; Sih et al., 2004b; Dingemanse and Réale, 2005; Réale et al., 2007, 2010; Gosling, 2008). However, many experimental biologists continue to treat inter-individual variation as statistical noise surrounding what they consider to be the evolutionarily important signal: the population (or treatment) mean (Careau et al., 2008; Williams, 2008) (Fig. S1). This oversight is unfortunate given that variation provides the raw material for natural selection, which operates at the level of the individual.

Phenotypic variation occurs both at the intra- and inter-individual level, with implications for a species’ ecology and evolution. On the one hand, the ability of individuals to alter their phenotype across contexts (intra-individual variation or plasticity) provides the flexibility to respond to environmental change on short timescales. On the other hand, the consistency of phenotypic differences among individuals across time or contexts influences the efficiency with which natural selection can generate an adaptive response across generations. Understanding both the proximate [e.g. how is genetic architecture (see Glossary) linked to phenotypic variation] and the ultimate (e.g. how does natural selection act on phenotypic variation) causes of intra-specific variation requires that we address questions at the individual level (Box 1). Experimental biologists working at the interface of behaviour, physiology, endocrinology and evolutionary biology are actively needed to help elucidate these relationships.

We believe that experimental biologists should consider measuring individual-level behavioural variation because it can (1) help avoid important biases when studying whole-animal physiology and performance, and (2) have an underlying physiological basis that warrants investigation. Our objective in this Commentary is to highlight key areas of interest and direct the reader to the relevant literature. First, we outline some of the confusion surrounding the terminology in the field of animal personality. Second, we argue that explicitly acknowledging individual behavioural differences can help researchers avoid sampling and experimental bias, increase

<sup>1</sup>Département d’Éco-Éthologie, Institut de Biologie, Université de Neuchâtel, Neuchâtel CH 2000, Switzerland. <sup>2</sup>Canada Research Chair in Functional Ecology, Department of Biology, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5.

\*Author for correspondence (dominique.roche@mail.mcgill.ca)

 D.G.R., 0000-0002-3326-864X

## Glossary

### Adjusted repeatability

Repeatability calculated after controlling for confounding effects, either as fixed or random factors. Assumes that the repeatability is constant for all values of the confounding factor (unlike conditional repeatability; refer to section VI in Nakagawa and Schielzeth, 2010).

### Agreement repeatability (or simply, repeatability)

The proportion of phenotypic variance due to differences among individuals, denoted as  $R$ . A measure of the agreement in absolute measurements (McGraw and Wong, 1996; Nakagawa and Schielzeth, 2010), as opposed to relative measurements. Equivalent to the intra-class correlation coefficient (ICC) in the statistical literature, which is the expected correlation among trait measurements within groups or individuals (i.e. classes) (Sokal and Rohlf, 1995).

### Animal personality

Consistent or repeatable inter-individual differences in behaviour across time and contexts (Réale et al., 2007). Variation among individuals in the intercept of their behavioural reaction norm (Dingemanse et al., 2010). Inter-individual variation in behaviour attributable to the combined influences of genetic effects and environmental effects that permanently affect the phenotype of an individual (Dingemanse and Araya-Ajoy, 2015).

### Behavioural syndrome

Suites of behaviours that co-vary across contexts or situations (Sih et al., 2004a).

### Behavioural type

Within a behavioural syndrome (e.g. boldness, aggressiveness), individuals have a behavioural type (e.g. shy versus bold, more versus less aggressive) (Sih et al., 2004b).

### Conditional repeatability

An estimate of repeatability at a given value of a fixed factor. Conditional repeatability changes across levels of a fixed factor, unlike adjusted repeatability (Nakagawa and Schielzeth, 2010; Biro and Stamps, 2015). Not discussed in this Commentary.

### Coping style

Characterizes the behavioural and physiological responses of individuals to a stressful situation (Koolhaas et al., 1999). Often used as a synonym of animal personality. 'Proactive' animals typically exhibit strong responses to stressful stimuli, whereas 'reactive' animals respond more passively.

### Covariate

A discrete or continuous variable included as a predictor (i.e. independent variable) in a model.

### Covariation

Occurs when variation in one trait is correlated with variation in another.

### Genetic architecture

The underlying genetic basis of a phenotypic trait.

### Minimum metabolic rate ( $MR_{min}$ )

The minimum maintenance metabolism of a resting, non-digesting animal; typically referred to as basal metabolic rate (BMR) in homeotherms and standard metabolic rate (SMR) in poikilotherms (Chabot et al., 2016).

### Phenotypic variance (or variability, or variation)

The statistical variance across the values of a phenotypic trait measured in a population.

### Reaction norm

A function describing the relationship between the phenotype (e.g. behaviour, metabolism, locomotor performance) and the environment or time.

### Selection gradient

A measure of a given trait's effect on relative fitness when the effects of all other measured traits are held constant.

### Variable centring

Expressing the observation of a continuous predictor variable as a deviation from its mean value across all observations. Achieved by subtracting the mean from all observations.

### Variance component

Variation (in the phenotype) associated with a random factor, such as individual or territory identity (Dingemanse and Dochtermann, 2013).

explanatory power and, ultimately, help us to understand how selection acts on physiological traits. Third, we present what we believe are the best practices for collecting, analysing and presenting data on individual differences, whether these are differences in behaviour, physiological or performance traits.

## What is (not) animal personality?

Few concepts in the biological sciences currently face as much controversy and lack of consensus over their usefulness as animal personality. Everything from startle responses in sea anemones to social dominance in chimpanzees has been characterized under the personality moniker (Freeman and Gosling, 2010; Stamps et al., 2012). Although animal personality is now engrained as a mainstream concept in behavioural ecology, the field has its share of critics. Many researchers ignore or dismiss animal personality as a useful concept, perhaps because the field is mostly theory-driven but without a strong conceptual framework (David and Dall, 2016), suffers from a lack of empirical studies (DiRienzo and Montiglio, 2015) and is laden with terminological inconsistencies (Réale et al., 2007; Carter et al., 2013). For example, terms such as temperament, behavioural syndrome,

behavioural type and coping style (see Glossary) are often used interchangeably with personality (Réale et al., 2010). For the non-specialist, navigating this breadth of literature and semantics, as well as determining when behaviours should not be considered as personality traits, can be daunting, and may contribute to the dismissive attitude of many experimental biologists regarding the importance of individual behavioural variation.

Despite these criticisms, proponents argue that personality research has greatly improved our understanding of how and when behavioural differences within populations might be adaptive (i.e. result from natural selection; Réale et al., 2007; Wolf and Weissing, 2012). Indeed, studying inter-individual variation in behaviour can be highly informative when addressing questions of interest to experimental biologists (Boxes 1 and 2). Here, we highlight key elements to consider when studying inter-individual differences in behaviour.

## Broad-sense versus narrow-sense personality

The field of animal personality seeks to characterize repeatable differences among individuals in one or more behaviours, and,

**Box 1. What questions can we ask at the individual level?**

Here, we use hypoxia tolerance to provide concrete examples for each question, since it is a phenotypic trait commonly studied by comparative physiologists, and it can affect behaviour (e.g. Domenici et al., 2007, 2013; Pollock et al., 2007).

1. Do individuals differ in their average phenotype (e.g. are there inter-individual differences in a given trait such as tolerance to hypoxia)?
2. How much variance in a given phenotypic trait occurs within individuals (e.g. do individuals vary in their tolerance to hypoxia from day to day, or among subsequent observations)?
3. Do individuals exhibit differences in phenotypic plasticity (e.g. do individuals differ in their metabolic response to increasing hypoxia)?
4. Is there a relationship between an individual's mean phenotype and its phenotypic plasticity for that same phenotype (e.g. is there a correlation between an individual's average activity and how its activity is affected across different oxygen levels)?
5. Are two or more phenotypic traits correlated at the inter-individual level [e.g. do individuals with generally high tolerance to hypoxia have a lower minimum metabolic rate ( $MR_{min}$ ; see Glossary)]?
6. Are two or more phenotypic traits correlated at the within-individual level (e.g. if, on a given day, an individual has a higher tolerance to hypoxia compared with its own average, does it also have a lower  $MR_{min}$  relative to its own average)?
7. Is phenotypic plasticity in two or more traits correlated at the inter-individual level (e.g. do individuals showing the greatest decrease in activity in response to decreasing oxygen levels also show the greatest decrease in metabolic rate in response to the same decrease in oxygen levels)?
8. Are there inter-individual differences in intra-individual variance (i.e. do individuals differ in their 'predictability'; see Fig. 2H)?

Patterns of intra- and inter-individual variation described in questions 1–4 are illustrated theoretically in Fig. 1 and can be answered by means of univariate linear mixed-effects models (LMMs; see 'Measuring repeatability'). Patterns of intra- and inter-individual covariation (see Glossary) described in questions 5–7 can be answered using multivariate LMMs (see 'Measuring intra- and inter-individual correlations'). Question 8 can be answered using double-hierarchical generalized LMMs (Cleasby et al., 2015). Examples and further details regarding these and other questions can be found in Dingemanse and Dochtermann (2013) and Metcalfe et al. (2016).

ultimately, to understand the functional basis and/or the ecological and adaptive significance of such consistent variation. Broadly speaking, the term personality could include any observable behaviour found to be consistently different between individuals (broad-sense personality; Réale et al., 2010) and likely to have ecological consequences. In this case, there is

no reason to use the term 'personality' as opposed to 'repeatable behaviour' (Careau and Garland, 2012). One clear disadvantage of the broad-sense personality approach is the contention over how consistent or repeatable inter-individual differences in behaviour must be over time for the term 'personality' to apply (Carter et al., 2013; Koski, 2014). For example, if 25% of the

**Box 2. Individual differences in behaviour can lead to interesting, multidisciplinary questions**

Studies at the individual level are extremely useful to identify the functional basis of, or relationships among, traits (Bennett, 1987). Whereas comparative studies traditionally examine trait (co)variation at the population or species level, the objective of an individual-level approach is to identify significant correlations between two or more traits at the individual level. This is a very effective first step towards identifying potential functional associations among complex traits, and should be followed-up by experimentation on the ecological, mechanistic, genetic and/or adaptive nature of the relationships found.

Many ecological and evolutionarily relevant traits are measured at the whole-organism level, allowing links to be made with lower-level processes (e.g. organ, tissue and cellular mechanisms) as well as processes occurring at higher levels of biological organisation (e.g. population or community dynamics). Therefore, investigating individual variation facilitates bridging key topics of interest in experimental biology, including biochemistry, physiology, morphology, endocrinology, neurobiology, organismal performance, behaviour, life-history strategies and population ecology (see Bennett, 1987). For example, a very active area of interdisciplinary research involving individual behavioural differences is on potential linkages with energy metabolism (Careau et al., 2008; Biro and Stamps, 2010; Careau and Garland, 2012; Mathot and Dingemanse, 2015). Several other studies have established conceptual links between personality and key topics of interest in ecology and evolutionary biology (see table below).

Topic	Example references
Social dominance	David et al., 2011
Mate choice	Schuett et al., 2010
Habitat or space use	Boon et al., 2008; Boyer et al., 2010
Dispersal	Cote et al., 2010
Social foraging	Kurvers et al., 2010
Social networks	Pike et al., 2008
Cognition	Morton et al., 2013; but see Griffin et al., 2015
Parasitism	Barber and Dingemanse, 2010
Life-history strategies	Biro and Stamps, 2008; Réale et al., 2010
Ontogeny	Stamps and Groothuis, 2010; Wilson and Krause, 2012
Performance	Careau and Garland, 2012; Videliér et al., 2014
Stress response	Cockrem, 2007; Øverli et al., 2007
Immune function	Zylberberg et al., 2014; Lopes, 2016
Oxidative stress	Costantini et al., 2008
Wildlife conservation	McDougall et al., 2006; Conrad et al., 2011; Killen et al., 2016

Considering the vast potential for exploring other exciting research avenues, some experimentalists will naturally be tempted to revisit classic research topics in experimental biology (e.g. motor control, blood flow, tolerance to hypoxia) from an individual perspective. However, in doing so, researchers should strive to synthesize existing theory and provide empirical support for the new conceptual advances they propose (see DiRienzo and Montiglio, 2015). In general, most of the new conceptual links listed in the table have a theoretical basis but all are in dire need of further empirical support.

phenotypic variance in a given behaviour is attributable to inter-individual differences, then the most precise way to capture this information is to say that the repeatability ( $R$ ) of this behaviour is  $R=0.25$  (see ‘Measuring repeatability’). It would seem counter-productive to then engage in a debate about whether this behaviour in this population represents personality or not because it is repeatable at  $R=0.25$ .

The narrow-sense definition of personality is restricted to behaviours measured repeatedly using tests that are standardized across individuals (e.g. open-field test, mirror test, use of refuge under simulated predation risk; Table 1). In this case, measured personality traits typically include activity, exploration, boldness, aggressiveness and sociability (Réale et al., 2007, Réale et al., 2010). Which traits are assessed depends on the experimental conditions under which they are assayed (familiar, novel, safe and/or risky environments; Table 1). However, some authors note that measuring a narrow range of behaviours under such artificial conditions might fail to capture relevant variation shaped by natural selection, or simply have no bearing on fitness-enhancing behaviours in the natural world (Koski, 2014; Niemelä and Dingemanse, 2014). Thus, an important area of research in

behavioural ecology is to validate whether behaviour measured under standardized (but artificial) conditions underlies ecological or fitness-relevant behaviour and life-history traits in nature, such as foraging, mating, parental care, agonistic interactions and dispersal (e.g. Herborn et al., 2010; Cole and Quinn, 2012; Dammhahn and Almeling, 2012; Niemelä et al., 2015).

### Personality and plasticity

A common misconception among researchers is that the presence of personality implies the absence of phenotypic plasticity in behaviour. However, individuals can differ consistently both in their average behaviour (i.e. personality) and in their degree of behavioural plasticity (Fig. 1). Hence, phenotypic variation encompasses both among (inter)- and within (intra)-individual variation. Moreover, personality and plasticity are often correlated (e.g. passive mice adjust their levels of aggression according to social context but aggressive mice do not show context-dependent modulation of their behaviour; Natarajan et al., 2009; see Fig. S2C,D). Thus, the responsiveness of an individual to situational changes (i.e. if and how an individual adjusts its behaviour to the environment) is, itself, an individual trait. Recent

**Table 1. Standard tests commonly used to assess animal behavioural traits, including exploration, activity, boldness, aggressiveness and sociability**

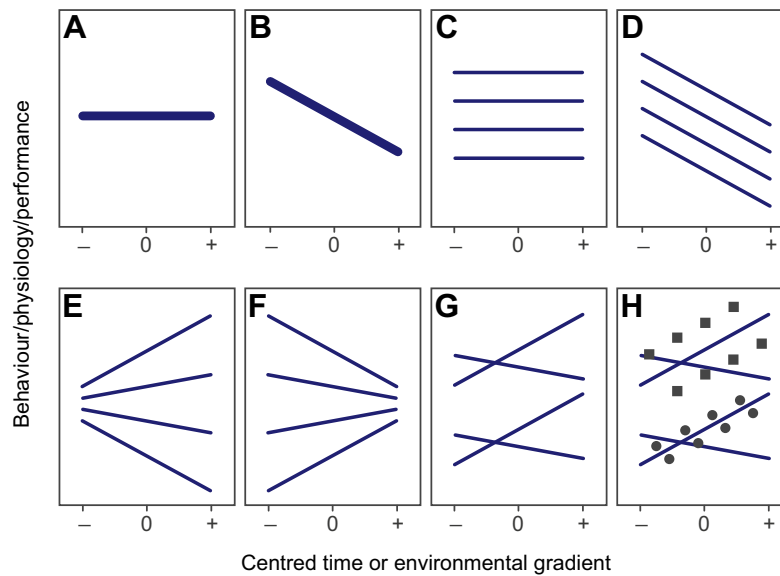
Laboratory tests				
Test	Procedure	Context	Traits assessed	References
Mirror	The animal is presented with a mirror in which it can see its own reflection.	Familiar	Aggressivity, sociability	8, 18
		Novel	Aggressivity, anxiety	7, 9, 21
Open field	The animal is given the opportunity to explore an unfamiliar space (usually a novel environment). Open field and novel environment are often used interchangeably.	Forced	Anxiety, neophobia*	1, 3, 7, 15, 18, 19, 20, 21
		Free	Exploration, boldness	1, 2, 26
Novel object	The animal is presented with an unfamiliar object.	Familiar	Exploration, boldness	1, 2, 3, 17, 18, 20
		Novel	Anxiety, boldness	22, 23, 25, 27
Emergence time	The animal is allowed to emerge from a shelter placed in an unfamiliar environment or after being startled in a familiar environment.	Familiar	Boldness	1, 28
		Novel	Boldness, anxiety	10, 12, 24
Activity	The animal is allowed to move freely in a familiar environment (e.g. its holding enclosure).	Familiar	Activity	11, 24
Proximity to conspecific	The animal's tendency to approach a conspecific in an adjacent compartment is recorded (the control is an empty compartment).	Familiar	Sociability	26
Field tests				
Test	Procedure		Traits assessed	References
Flight initiation distance	The animal's flight initiation distance is recorded as the experimenter or a stimulus approaches it at a constant velocity.		Boldness	4, 5, 6
Predator simulation	The animal is startled using a mock predator or audio playbacks; the experimenter records emergence time from a shelter, time to return to a food patch or resumption of activity.		Boldness	16
Restraining test	The animal is restrained following capture (e.g. by placing it in a transport bag) and its behavioural response quantitatively assessed.		Boldness, docility	14, 21
Activity	The animal's activity patterns are monitored via tracking technology [e.g. radio, acoustic, passive integrated transponder (PIT) tags].		Activity	13
Proximity to conspecifics	The number and duration of an animal's interaction with known conspecifics is quantified (e.g. proximity tags, radio collars, focal observations).		Sociability	13

The trait (or traits) assessed will depend on the context in which the test is carried-out: ‘familiar’ and ‘novel’ refer to whether the animal has had prior experience with the environment or object used in the test; a test is ‘forced’ or ‘free’ depending on whether the animal is given no choice to behave (e.g. placed inside a novel arena) or allowed to behave at will (e.g. the animal can choose to enter a novel arena from a familiar environment such as its holding enclosure).

<sup>1</sup>Carter et al., 2013; <sup>2</sup>Herborn et al., 2010; <sup>3</sup>Verbeek et al., 1994; <sup>4</sup>Carter et al., 2012; <sup>5</sup>Binning et al., 2014; <sup>6</sup>Blumstein, 2003; <sup>7</sup>Ariyomo and Watt, 2012; <sup>8</sup>Balzarini et al., 2014; <sup>9</sup>Farrell and Wilczynski, 2006; <sup>10</sup>Harris et al., 2010; <sup>11</sup>Lopes et al., 2012; <sup>12</sup>Byrnes and Brown, 2016; <sup>13</sup>Wilson et al., 2015; <sup>14</sup>Careau et al., 2015; <sup>15</sup>Dingemanse et al., 2012; <sup>16</sup>Neumann et al., 2013; <sup>17</sup>Koski and Burkart, 2015; <sup>18</sup>Adriaenssens and Johnsson, 2013; <sup>19</sup>Both et al., 2005; <sup>20</sup>Burns, 2008; <sup>21</sup>Petelle et al., 2015; <sup>22</sup>Hamilton et al., 2014; <sup>23</sup>Ou et al., 2015; <sup>24</sup>Biro, 2012; <sup>25</sup>Dammhahn, 2012; <sup>26</sup>Jolles et al., 2015; <sup>27</sup>Kurvers et al., 2009; <sup>28</sup>Mowles et al., 2012.

\*Most papers call the behaviour measured in a forced novel environment tests ‘exploratory behaviour’, but see Carter et al. (2013).





**Fig. 1. Theoretical relationships between phenotypic traits (behaviour, physiology, performance) and time or environmental context examined using an individual-level approach.** The reaction norm approach is useful to inform trait differences among individuals (inter-individual variation) and variation within individuals (intra-individual variation). Personality differences are best thought of as repeatable inter-individual differences in behaviour as measured by the repeatability index '*R*' (Box 3). Lines represent responses of four hypothetical individuals to changes in time or context – for example, levels of aggression in four mice as a function of temperature. (A) All four individuals exhibit the same mean trait levels (same intercepts) and do not respond to changes in time or contexts (zero slope): repeatability and plasticity are nil. (B) Individuals exhibit the same mean trait levels (same intercepts) and respond to time or environmental change in the same way (identical, non-zero slopes): repeatability is nil and plasticity is high. (C) Individuals exhibit different mean trait levels (different intercepts) and do not respond to changes in time or contexts (zero slope): repeatability is high and plasticity is nil. (D) Individuals have different intercepts and respond to time or environmental change in the same way (identical, non-zero slopes): plasticity and (adjusted) repeatability are high, such that the magnitude of trait differences between individuals is maintained across time or contexts. When individuals differ in their levels of plasticity, trait differences (i.e. variation among intercepts at a given time or in a given context) can either be accentuated (E) or attenuated (F). Here, intercepts and slopes differ between individuals: there can be positive slope–intercept covariance (i.e. individuals with a greater intercept have a more positive slope; panel E) or negative slope–intercept covariance (i.e. individuals with a lower intercept have a more positive slope; panel F). (G) Intercepts and slopes differ between individuals and there is no slope–intercept covariance (i.e. no relationship between the slope and the intercept). In E, F and G, plasticity is high and so is (adjusted) repeatability because a large proportion of the total phenotypic variance is due to differences between individuals (i.e. differences in intercepts; see Box 3). For simplicity, residual variation (i.e. unexplained variation) is not shown in A–G. In reality, however, these lines would be fitted through data points with a certain degree of residual variance, which can also vary among individuals. (H) For example, two individuals have the same slope, but exhibit different intra-individual variance (compare circles and squares). Modified from Dingemanse et al. (2010).

evidence suggests that neural plasticity may underlie the physiological (endocrine) and molecular/genetic basis for plasticity versus rigidity in behavioural responses, especially those related to stress (reviewed in Sørensen et al., 2013).

To study behavioural consistency and plasticity, behavioural ecologists have embraced the concept of behavioural reaction norms (BRNs) (Dingemanse et al., 2010) (Fig. 1). The idea is similar to a genetic×environment interaction (G×E) in evolutionary quantitative genetics (Nussey et al., 2007), but instead applied at the individual×environment level (I×E). Once behavioural variation is described as a BRN with an intercept (mean-level individual behaviour, indicative of personality) and a slope (behavioural response to context, indicative of phenotypic plasticity), it becomes possible to examine how personality and phenotypic plasticity are (1) correlated, (2) under selection, and (3) proximally linked through shared underlying mechanisms (Dingemanse et al., 2010; Brommer, 2013; Dingemanse and Dochtermann, 2013). We detail the general reaction norm approach (see Glossary) and provide worked examples with R code in the section 'Measuring repeatability'.

### Why should experimental biologists care about animal personality?

Studying the adaptive evolution of complex traits (e.g. behaviour, metabolic rate) requires the use of techniques and tools developed by quantitative geneticists (Falconer and Mackay, 1996; Lynch and

Walsh, 1998). These tools were developed to measure individual variation (repeatability), transmission of traits from parents to offspring and the strength of natural selection. The advent of personality studies that use approaches from quantitative genetics to link repeatable individual differences in behaviour with core aspects of evolutionary biology [e.g. selection gradients (see Glossary), heritability] has led behavioural ecologists closer to understanding how behaviour evolves (Dingemanse and Réale, 2005). Help from experimental biologists is particularly needed to identify the mechanisms underlying context-dependent correlations between personality and other traits such as metabolic rate, life-history traits and immune function (Box 2). Additionally, experimentalists who familiarize themselves with approaches used in quantitative genetics and personality research will be better positioned to study the ecological and evolutionary consequences of individual variation in physiology, performance and other complex traits (e.g. hypoxia tolerance, stress response). In fact, early examples already demonstrate how physiological and performance traits can be studied from a quantitative genetics perspective (Garland, 1988, 1994; Dohm et al., 2001). In addition to helping researchers adopt new, interdisciplinary approaches to answer evolutionary and ecological questions of interest (Box 2), there are several practical reasons why experimental biologists should consider personality differences among individuals in their research, which we discuss in detail below.

### To control for population sampling bias (including domestication)

Most statistical models assume that the study sample represents a random selection of individuals from the population; therefore, conclusions drawn from a biased study sample might not reflect the range of responses observed in nature or apply to the population as a whole. Sampling biases are likely if using domesticated laboratory populations to address questions about behavioural or physiological trait variation. For example, laboratory conditions rapidly select for certain behavioural and physiological traits as a result of domestication (Lacy et al., 2013). This probably reduces trait variation and inter-individual differences compared with those of wild populations. While laboratory populations are extremely useful for answering mechanistic or process-oriented questions, an even more powerful approach is to seek cross-context validation by combining laboratory and field measurements when feasible (see Niemelä and Dingemanse, 2014).

Sampling biases might also occur in experiments that use wild-caught animals, because some individuals are inherently more ‘catchable’ or ‘trap-happy’ than others (Biro and Dingemanse, 2009; Carter et al., 2012). Thus, differences in the propensity of wild animals to enter traps or take bait selects for certain behavioural types (e.g. Wilson et al., 1993; Garamszegi et al., 2009; Carter et al., 2012; Stuber et al., 2013; Diaz Pauli et al., 2015; Niemelä et al., 2015). This is probably also true of physiological and/or performance traits, although research on this question is lacking (but see Killen et al., 2015). Researchers can reduce these biases by selecting or combining capture methods that are likely to sample different behavioural strategies or physiological abilities (Biro and Dingemanse, 2009). Where possible, researchers should also observe the various responses of individuals to their capture techniques in order to identify those that are particularly susceptible to capture, and modify their techniques accordingly. Incorporating a quantitative measure of individual ‘catchability’ during this process may also allow researchers to evaluate the distribution of the different behaviours or strategies used by individuals, and see how such differences relate to performance or fitness measures obtained in subsequent experiments.

### To estimate personality-related measurement bias in physiological measurements

Individuals often exhibit dramatically different behavioural responses to novel or stressful situations such as handling or introduction to an experimental apparatus. This can lead to differences in the time required to resume normal behaviour (Koolhaas et al., 1999; Carere and van Oers, 2004; Øverli et al., 2007) or the physiological response to a subsequent treatment. These differences can introduce personality-related bias into physiological measurements, as exemplified by the potential influence of personality differences on minimum metabolic rate ( $MR_{\min}$ ) estimates obtained using respirometry (Hayes et al., 1992; Careau et al., 2008). This can be particularly important when experimenters rely on the shape of the metabolic curve (metabolism–time) and the presence or absence of activity to determine  $MR_{\min}$ . In this case, it is possible that  $MR_{\min}$  measurements are inflated by stress for some, but not all, individuals. Such a bias can arise because ‘reactive’ individuals tend to respond to handling and/or novel situations with intense, non-motor behavioural responses, which can persist over long periods of time (Koolhaas et al., 1999). Therefore, while periods of inactivity might truly correspond to resting periods for ‘proactive’ behavioural types, this might not be the case for reactive individuals. Identifying whether inactivity is indicative of resting by reactive animals is difficult when acclimation periods are short because this

reaction might correspond to an intense and prolonged stress response (Careau et al., 2008).

Although personality-related measurement bias seems likely, there is little empirical evidence on the extent of this problem. However, ignoring obvious differences in individuals’ behavioural responses to experimental procedures will introduce large, unexplained inter-individual variation in physiological or performance parameters. In turn, this variation can modulate treatment effects in experimental studies and/or correlations obtained between multiple traits (Killen et al., 2013). For example, MacKenzie et al. (2009) found that ‘proactive’ and ‘reactive’ carp (*Cyprinus carpio*) exhibit distinct and sometimes opposite patterns of gene expression under control conditions and in response to an immune challenge. Similarly, Rupia et al. (2016) characterized the behavioural and physiological stress responses of olive flounder, *Paralichthys olivaceus*. They found that flounder with ‘bold’ behavioural types responded to acute stress by increasing their metabolic rates, whereas ‘shy’ types decreased their metabolic rates relative to routine. In both cases, failing to account for the dramatic variation in the direction of the stress response between types would mask the effects of the stressor if the population were examined at large. This area of research is in urgent need of further investigation, and experimental biologists are ideally positioned to provide crucial insights via precise, concurrent experimental measurements of multiple behavioural and physiological processes, such as activity, breathing rate, body temperature and metabolic rate.

### To estimate personality-related measurement bias in performance measurements

Variation in motivation among individuals with different behavioural traits can result in underestimates of maximal performance, inflated error variance and biased repeatability (Losos et al., 2002; Adolph and Pickering, 2008; Careau and Garland, 2012; Jornod and Roche, 2015). To measure performance rather than behaviour, tests must therefore force animals to perform at their maximal level. Carefully designing such tests can involve: (1) using stimuli that increase motivation by targeting multiple senses (e.g. stimulate auditory and visual senses simultaneously to induce an escape response), (2) repeating performance trials on individuals to increase the accuracy of performance measurements, and (3) repeating series of trials at different times to quantify intra-individual variance using multilevel mixed-effects models (Cleasby et al., 2015) (Fig. S3).

Instead of focusing only on an individual’s best performance, experimenters can gain much insight from retaining performance measurements across repeated trials. For example, some individuals may be consistently motivated to perform at their maximum level, whereas others may be more variable in their level of motivation, resulting in low and high intra-individual variance, respectively (Fig. S3). If these differences exist, then it becomes interesting to test whether intra-individual variation in performance and/or motivation correlates with behavioural and/or physiological traits (Careau and Garland, 2012). Retaining all repeated performance measures also allows experimenters to test whether individuals differ in how their performance changes across trials, which may be indicative of habituation, fatigue and/or training effects (Fig. S3).

### How should we quantify animal personality?

#### Selecting the right test(s)

A major source of confusion in animal personality research is how personality traits are measured and defined (Carter et al., 2013). First, problems with how behaviours are measured include

determining which personality traits are assessed by a given test. For instance, a single test might assess many traits (i.e. a one-to-many problem; an open field test can be used to measure both exploration and boldness) and different tests might assess the same personality trait (i.e. a many-to-one problem; exploration can be assessed using an open field or novel object test) (Carter et al., 2013) (Table 1). A second, related issue lies in the mislabelling of behaviours assessed in standard tests, resulting in the so-called ‘jingle-jangle fallacy’, where a single behavioural trait is given different labels in different studies (e.g. distance moved in an open field test is said to represent anxiety, boldness and exploration). Alternatively, different traits are labelled as the same (Gosling, 2001; Carter et al., 2013; Koski, 2014); for example, ‘boldness’ has been quantified as the response to a novel object, a novel environment and predation risk (Toms et al., 2010). This source of confusion is especially problematic when researchers attempt to relate personality traits to physiological responses, because the nature of this relationship will vary depending on how the labelled behavioural or physiological trait was measured. For example, a novel object test in an animal’s holding enclosure might assess exploration/curiosity and boldness, whereas the same test in an unfamiliar environment might instead measure fear and/or anxiety (Table 1) (Misslin and Cigrang, 1986; Carter et al., 2013). As such, the magnitude or direction of a correlation between a behavioural and a physiological trait (e.g. distance travelled versus metabolic rate) might differ considerably depending on the context in which the behaviour was assessed. Ignoring this issue will inevitably lead to different predictions regarding the performance or fitness implications associated with such traits.

In general, mislabelling behaviours is problematic, as it impedes our ability to compare behavioural responses across studies, reduces our effectiveness at interpreting the ecological relevance of a measured response and, ultimately, hampers our understanding of how selection acts on behavioural traits in the wild (Carter et al., 2013). To minimize these issues, experimental biologists can simply report what has been measured (e.g. report the distance moved by an animal in a 10 min open-field test) rather than assigning the behaviour a label such as ‘exploration’. Avoiding labels can help prevent the misinterpretation of behaviours assessed using the same test in different contexts. This is easier to achieve when measuring fewer behavioural traits, as would be expected in physiological studies, but difficult when researchers are interested in measuring multiple behavioural traits to look at the existence of independent factors/axes of correlated personality traits such as a boldness–aggression axis, as performed in behavioural ecology studies (see Carter et al., 2013 for suggested approaches in this case).

In addition to considering whether a chosen test really measures the behaviour it is intended to measure (i.e. construct validity; Carter et al., 2013; Koski, 2014), researchers must also consider whether the design of the test and the behaviours measured are ecologically relevant for the species of interest (i.e. ecological validity; Réale et al., 2007; Beckmann and Biro, 2013; Koski, 2014). This is especially true when trying to understand the life-history implications or fitness outcomes associated with a particular behaviour and when comparing studies. Researchers should target traits that are the most ecologically relevant for their study species (Dall and Griffith, 2014), which requires knowledge of the species’ ecology and behavioural repertoire (Koski, 2014). Suggestions of how to achieve this goal are to select traits *a priori* based on sound knowledge of the species’ ecology (Koski, 2014), trial assays on the study species (e.g. Balzarini et al., 2014) and identify poor tests early on (Carter et al., 2013).

### Measuring repeatability

Perhaps the most important feature of personality studies is that behaviour must be measured multiple times. Repeatedly measuring traits in as many individuals as possible allows the calculation of repeatability ( $R$ ), a highly informative population-specific metric to quantify inter-individual phenotypic differences across time or contexts (Box 3).  $R$  provides a standardized estimate of individuality that can be compared across studies; it is also an inherent component of quantitative genetics theory because repeatability sets an upper limit to heritability (i.e. a trait cannot be more heritable than it is repeatable within a population; but see Dohm, 2002). In fact, a close read of the foundational papers on personality and behavioural syndromes (e.g. Sih et al., 2004b; Réale et al., 2007) reveals that the field’s core lies in the application of approaches in quantitative genetics to study the evolution of behaviour (see Boake et al., 2002; Dingemanse and Araya-Ajoy, 2015). Most recently, several studies in the Journal of Experimental Biology have illustrated how these methods can also be used to estimate the repeatability of physiological and performance-related traits and gain important insight into their ecological and evolutionary implications (e.g. Laming et al., 2013; Darveau et al., 2014; Auer et al., 2016; Conradsen et al., 2016).

Calculations to obtain  $R$  are detailed in several excellent reviews and methods articles (Nakagawa and Schielzeth, 2010; Martin et al., 2011; Wolak et al., 2012; Dingemanse and Dochtermann, 2013; Cleasby et al., 2015) but applying these methods can be challenging. Here, we provide a simple worked example for

### Box 3. The repeatability index, $R$

Whenever a trait is measured multiple times on the same set of individuals, the total phenotypic variance ( $V_P$ ) of the population sample can be partitioned into two variance components (see Glossary): inter-individual variance ( $V_{ind}$ ) and intra-individual, or residual, variance ( $V_e$ ). Repeatability ( $R$ ) is the proportion of the total phenotypic variance ( $V_P$ ) attributable to differences between individuals ( $V_{ind}$ ):  $R = V_{ind} / (V_{ind} + V_e)$ , where  $V_P = V_{ind} + V_e$ . Since  $V_e$  corresponds to the non-repeatable fraction of  $V_P$  (i.e. the sum of measurement error and phenotypic variance in response to micro-environmental effects or any unmeasured variable),  $R$  provides a standardized measure of the consistency of phenotypes across time or contexts (Nakagawa and Schielzeth, 2010). Two phenomena can lead to low  $R$  values: high intra-individual variation ( $V_e$ ) and/or low inter-individual variation ( $V_{ind}$ ). Note that inter-individual differences are needed (i.e.  $V_{ind} > 0$ ) for  $R$  to be non-zero (see Fig. 1A,B).

$R$  estimates can be obtained using correlation, ANOVA and linear mixed-effects models (LMMs) (see Wolak et al., 2012 for details). LMMs are currently the method of choice because they (1) allow direct estimates of inter- and intra-individual variance, (2) do not require balanced or complete sampling, and (3) allow calculating  $R$  for traits with non-normal error distributions (Nakagawa and Schielzeth, 2010; Dingemanse and Dochtermann, 2013).

As with other sample statistics (e.g. the mean),  $R$  is an estimate of the true population repeatability and should be accompanied by an estimate of uncertainty (i.e. a standard error or confidence interval; Sokal and Rohlf, 1995). Nakagawa and Schielzeth (2010) propose different methods of calculating confidence intervals (CI) for  $R$ , including a Bayesian approach that uses Markov chain Monte Carlo (MCMC) algorithms and does not require randomization or bootstrapping procedures (these are needed with restricted maximum likelihood methods; REML). Guidelines for determining the sample size (number of individuals,  $n$ ) and number of measurements (number of times a trait is measured on the same individual,  $k$ ) to accurately estimate  $R$  are given by Wolak et al. (2012): the rule of thumb is to increase  $n$  when repeatability is high and increase  $k$  when repeatability is low.

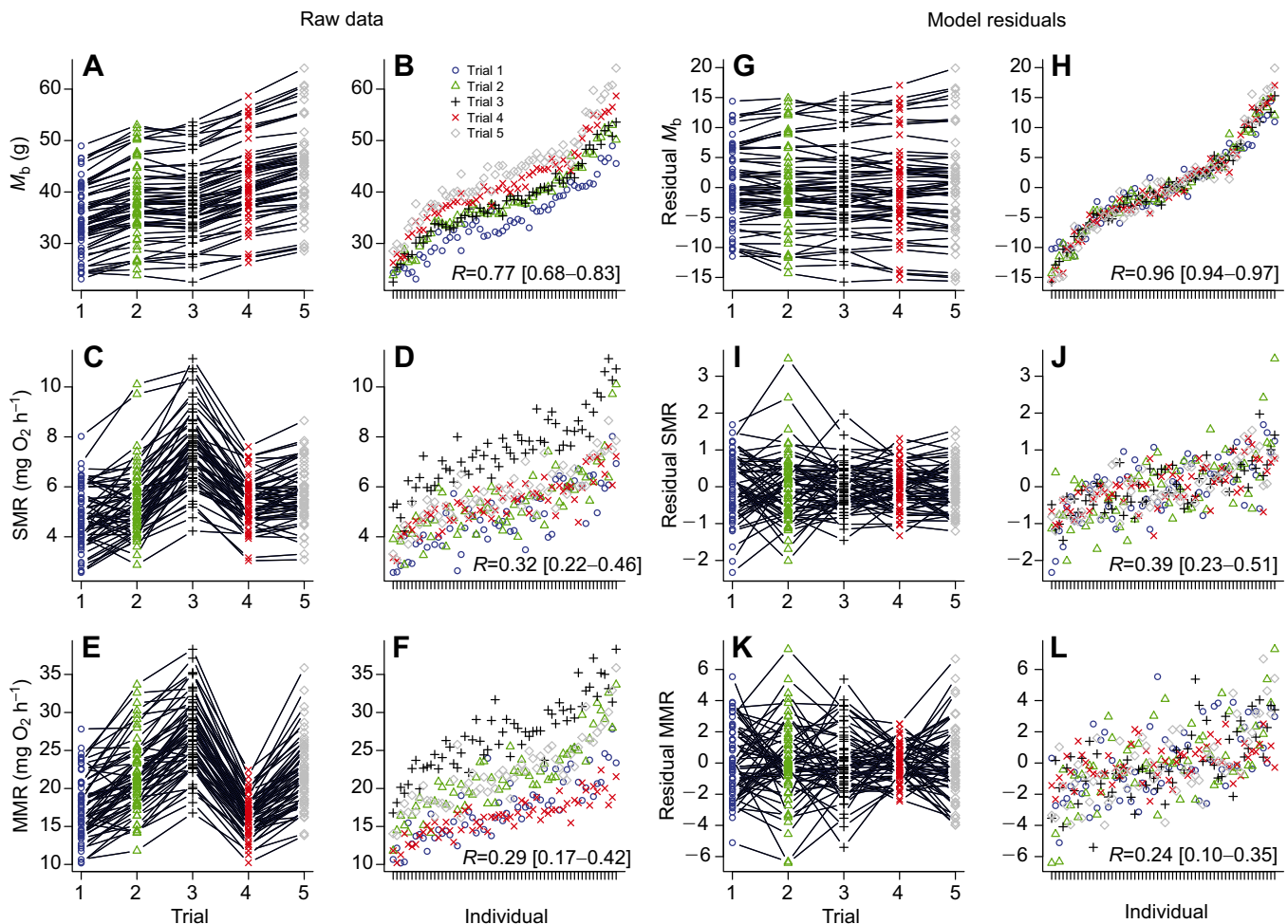


illustrative purposes, which researchers can easily modify and adapt to analyse their own data. We use data from a published study on metabolic traits of fishes (Norin et al., 2016) but the analyses are applicable to any aspect of the phenotype, including behavioural traits. The annotated script is available online (<http://dx.doi.org/10.6084/m9.figshare.3464216>) and it details the computation of  $R$  and its 95% credibility interval (a confidence interval in the Bayesian framework) using the R package MCMCglmm (Hadfield, 2010). For complementary R functions, refer to Nakagawa and Schielzeth (2010) and Wolak et al. (2012).

Norin et al. (2016) took repeated measurements of body mass, standard metabolic rate (SMR) and maximum metabolic rate (MMR) in 60 juvenile barramundi (*Lates calcarifer*), each sequentially exposed to five different environmental treatments consisting of different temperature, salinity and oxygen levels (900 measurements in total). Their objective was to examine whether individuals exhibit differences in their ability to cope with rapidly changing environmental conditions.

The first step to explore these data using an individual-level approach is to compute  $R$  using the raw phenotypic variance [i.e. without controlling for covariates (see Glossary) or fixed-effects]. This  $R$  estimate is termed ‘agreement repeatability’ (see Glossary; Nakagawa and Schielzeth, 2010) and is obtained with a linear mixed-effects model (LMM) comprising only a random factor (fish ID). The results reveal that body mass is highly repeatable across contexts ( $R=0.77$ ,  $CI=0.68–0.83$ ) but that SMR and MMR are only moderately repeatable ( $R=0.32$ ,  $CI=0.22–0.46$  and  $R=0.29$ ,  $CI=0.17–0.42$ ). Fig. 2A–F illustrates the degree of repeatability and variation in these traits using reaction norm plots and caterpillar plots. The reaction norm plots suggest that fish gained mass over the course of the study (Fig. 2A) and that SMR and MMR differ among experimental treatments (Fig. 2C,E).

As a second step, we add time (trials 1–5 specified as a continuous variable) and experimental conditions (three separate factors: temperature, salinity and oxygen) as predictors in the model, with mass as a response variable to control for the variance due to these



**Fig. 2.** Plots illustrating inter- and intra-individual variance using data from Norin et al. (2016). Body mass ( $M_b$ ), standard metabolic rate (SMR) and maximum metabolic rate (MMR) were measured five times each in 60 juvenile barramundi fish (*Lates calcarifer*) under different experimental treatments (trial 1: 35 ppt salinity, 29°C, normoxia; trial 2: 10 ppt salinity; trial 3: 35°C; trial 4: 45% air saturation; trial 5: same as trial 1). Inter- and intra-individual variation is depicted before (A–F) and after correcting for fixed effects (G–L; fixed effects include trial number, treatment and body mass). In both cases, the data are illustrated twice: first with a plot showing individual trajectories over the five trials (trial number was centred in the analysis, so trial 3 becomes the intercept [ $y$  at  $x=0$ ], representing the mean trait across trials; A, C, E and G, I, K), and second with a caterpillar plot showing trait variation across trials (individuals are ordered on the x-axis based on their mean-level trait; B, D, F and H, J, L). In all cases, we estimated repeatability ( $R$ ) and 95% credible intervals (in square brackets) using the R package MCMCglmm (Hadfield, 2010). The R script and data to run the models and create the graphs are available on Figshare (<http://dx.doi.org/10.6084/m9.figshare.3464216>).



fixed effects.  $V_p$  (Box 3) becomes the total phenotypic variance minus the variance accounted for by the fixed effects (time and experimental treatments). As a result, the variance components (see Glossary) in the model change and so does  $R$ : whether  $R$  increases or decreases depends on whether the fixed effects inflate variance at the intra- or inter-individual levels (see Nakagawa and Schielzeth, 2010 for details). Re-calculating  $R$  as such reveals that body mass is highly repeatable across trials ( $R=0.96$ ,  $CI=0.94–0.97$ ). Accordingly, a plot of the model residuals (i.e. after controlling for time and treatment effects) indicates that the majority of the variance in mass occurs among rather than within individuals (Fig. 2G).

Next, we control for the effect of body mass on SMR and MMR (as well as time and treatment effects), to account for the fact that large differences in body mass among individuals inflate repeatability (see Box 3). This is achieved by including these variables as fixed effects in the models. The new adjusted  $R$  estimates are 0.39 for SMR ( $CI=0.23–0.51$ ) and 0.24 ( $CI=0.10–0.35$ ) for MMR (Fig. 2I–L), both of which fall within the range commonly reported for these traits (Auer et al., 2016).  $R$  estimates obtained in this way represent ‘adjusted repeatability’ (see Glossary) as opposed to ‘agreement repeatability’ obtained from LMMs without fixed effects (Nakagawa and Schielzeth, 2010) (refer to the Glossary for an explanation of different repeatability measurements and relevant references).

So far, we have analysed data from Norin et al. (2016) using a random-intercept LMM: individual identity was specified as a random factor in the model such that the  $y$ -intercept for each individual was allowed to vary relative to the population intercept. However, as we saw, fish grew substantially during the experiment (Fig. 2A), and we now wish to test whether individuals differ in their growth trajectories (see step four below). One advantage of LMMs is the ease with which the user can implement a reaction norm framework to simultaneously model variation in the individual phenotypic mean (e.g. personality or, in this case, an individual’s mean mass) and individual responses across time or contexts (i.e. plasticity, or in this case, an individual’s growth rate) (Nussey et al., 2007; Dingemanse et al., 2010; Westneat et al., 2011; Dingemanse and Dochtermann, 2013). Reaction norms are functions that describe the relationship between an individual’s phenotype and the environment (or time) and can be estimated using a random slope LMM (also called a random regression model; Henderson, 1982). In the simplest case of a linear reaction norm, a random slope LMM characterizes each individual with an intercept and a slope. Since the slope of the reaction norm represents the change in phenotype through time or across an environmental gradient, it captures variation in phenotypic plasticity. Variance among intercepts in random slope LMMs is estimated at the point where the covariate is zero. Therefore, to examine differences in mean phenotype, it is important to rescale covariates to a mean of zero such that the intercept represents the average phenotype (this does not apply if there is curvature in the reaction norms) (Morrissey and Liefting, 2016). Mean centring of covariates is also important when estimating the slope–intercept relationship in reaction norms, which allows examining whether individuals with different intercepts have different slopes (see question 4 in Box 1). Typically, random slope LMMs include a covariance term to describe the slope–intercept covariance, which is highly sensitive to where the intercept is located relative to the range over which the data were gathered.

In a fourth step, therefore, we extend our LMM with mass as a response variable to a random-slopes model by including a ‘fish

ID×trial’ interaction term as a random effect. Doing so allows individual slopes to vary with regards to the population slope and reveals that individuals differ significantly in how they grew throughout the study (i.e. growth trajectories differ significantly among individuals in Fig. 2G). Additionally, the slope–intercept correlation is significant and positive ( $r=0.54$ ), indicating that individuals that were heavier on average (i.e. had a greater mean mass across all five trials) grew faster throughout the experiment. Models for SMR and MMR can be extended to random-slopes LMMs in the same way.

### Measuring intra- and inter-individual correlations

Another key advantage of mixed-effects models is the ease with which the user can transition from a univariate to a multivariate analysis. Whereas univariate models are used to calculate  $R$ , multivariate models allow examining relationships between traits at multiple levels of variation simultaneously. If two traits are functionally related, they should be significantly correlated at multiple levels of biological organisation (e.g. within and among individuals, among populations and among species). However, it is also possible that some fundamental trade-off (e.g. energy allocation to different biological functions) applies only at one level of biological organisation, yielding contrasting relationships at different levels (van Noordwijk and de Jong, 1986; Houle, 1991). Under such circumstances, correlations observed within individuals will differ from those observed among individuals (e.g. van de Pol and Wright, 2009). We provide additional detail and a worked example to compare intra- and inter-individual correlations online (<http://dx.doi.org/10.6084/m9.figshare.3464216>).

### Conclusion

Animal personality research is a controversial and polarizing topic among biologists, yet the concept has greatly advanced our understanding of the ecological and evolutionary importance of individual phenotypic variation in behaviour. Experimental biologists have much to gain from adopting a similar individual-level approach to (1) quantify and understand variation in organismal performance and physiology (Box 1, Fig. 1), (2) unravel possible mechanistic links between behaviour and physiology (Box 2), and (3) quantify the extent to which traits of interest are repeatable and/or plastic using analytical tools in personality and quantitative genetics research. In doing so, we suggest that experimentalists preferentially avoid the use of specific labels (e.g. boldness) when characterizing behaviour, seek to identify the physiological mechanisms underlying differences in behaviour, and strive to use validated methodologies for collecting, analysing and presenting behavioural data using repeated measurements.

In addition to providing interesting conceptual advances towards understanding how and why traits evolve, an increased focus on individual-level variation has major implications for predicting the effects of environmental changes on wild populations. Continuing the current emphasis on mean treatment effects of environmental stressors (e.g. hypoxia, low pH) while ignoring variation around the mean is a missed opportunity for understanding the traits responsible for conferring resistance or tolerance to a particular stressor (Browman, 2016; Killen et al., 2016). Examining variation in complex traits, including why some individuals are strongly affected by experimental treatments whereas others are not, is a critical first step towards truly understanding how individuals respond and adapt to environmental change, both in the short term and in the long term.

## Acknowledgements

We thank Tommy Norin, Redouan Bshary, Alecia Carter and Mariana Velasque for informative discussions and feedback on the manuscript. Three reviewers provided helpful comments.

## Competing interests

The authors declare no competing or financial interests.

## Author contributions

D.G.R., V.C. and S.A.B. wrote the paper. V.C. wrote the R script to analyse data from Norin et al. (2016).

## Funding

D.G.R. and S.A.B. were supported by postdoctoral fellowships from the Fonds de Recherche du Québec - Nature et Technologies.

## Data availability

Tommy Norin generously provided the data for the examples presented in the paper (available at <http://onlinelibrary.wiley.com/doi/10.1111/1365-2435.12503/abstract>). These data (in long format) and the associated R scripts are available on Figshare at: <http://dx.doi.org/10.6084/m9.figshare.3464216>.

## Supplementary information

Supplementary information available online at <http://jeb.biologists.org/lookup/doi/10.1242/jeb.146712.supplemental>

## References

- Adolph, S. C. and Pickering, T. (2008). Estimating maximum performance: effects of intraindividual variation. *J. Exp. Biol.* **211**, 1336–1343.
- Adriaenssens, B. and Johnsson, J. I. (2013). Natural selection, plasticity and the emergence of a behavioural syndrome in the wild. *Ecol. Lett.* **16**, 47–55.
- Ariyomo, T. O. and Watt, P. J. (2012). The effect of variation in boldness and aggressiveness on the reproductive success of zebrafish. *Anim. Behav.* **83**, 41–46.
- Arnold, S. J. and Bennett, A. F. (1984). Behavioural variation in natural populations. III. Antipredator displays in the garter snake *Thamnophis radix*. *Anim. Behav.* **32**, 1108–1118.
- Auer, S. K., Bassar, R. D., Salin, K. and Metcalfe, N. B. (2016). Repeatability of metabolic rate is lower for animals living under field versus laboratory conditions. *J. Exp. Biol.* **219**, 631–634.
- Balzarini, V., Taborsky, M., Wanner, S., Koch, F. and Frommen, J. G. (2014). Mirror, mirror on the wall: the predictive value of mirror tests for measuring aggression in fish. *Behav. Ecol. Sociobiol.* **68**, 871–878.
- Barber, I. and Dingemanse, N. J. (2010). Parasitism and the evolutionary ecology of animal personality. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 4077–4088.
- Beckmann, C. and Biro, P. A. (2013). On the validity of a single (boldness) assay in personality research. *Ethology* **119**, 937–947.
- Bennett, A. F. (1980). The thermal dependence of lizard behaviour. *Anim. Behav.* **28**, 752–762.
- Bennett, A. F. (1987). Interindividual variability: an underutilized resource. In *New Directions in Ecological Physiology* (ed. M. Feder, A. Bennett, W. Burggren and R. Huey), pp. 147–169. Cambridge: Cambridge University Press.
- Binning, S. A., Barnes, J. I., Davies, J. N., Backwell, P. R. Y., Keogh, J. S. and Roche, D. G. (2014). Ectoparasites modify escape behaviour, but not performance, in a coral reef fish. *Anim. Behav.* **93**, 1–7.
- Biro, P. A. (2012). Do rapid assays predict repeatability in labile (behavioural) traits? *Anim. Behav.* **83**, 1295–1300.
- Biro, P. A. and Dingemanse, N. J. (2009). Sampling bias resulting from animal personality. *Trends Ecol. Evol.* **24**, 66–67.
- Biro, P. A. and Stamps, J. A. (2008). Are animal personality traits linked to life-history productivity? *Trends Ecol. Evol.* **23**, 361–368.
- Biro, P. A. and Stamps, J. A. (2010). Do consistent individual differences in metabolic rate promote consistent individual differences in behavior? *Trends Ecol. Evol.* **25**, 653–659.
- Biro, P. A. and Stamps, J. A. (2015). Using repeatability to study physiological and behavioural traits: ignore time-related change at your peril. *Anim. Behav.* **105**, 223–230.
- Blumstein, D. T. (2003). Flight-initiation distance in birds is dependent on intruder starting distance. *J. Wildl. Manage.* **67**, 852–857.
- Boake, C. R. B. (1989). Repeatability: its role in evolutionary studies of mating behaviour. *Evol. Ecol.* **3**, 173–182.
- Boake, C. R. B., Arnold, S. J., Breden, F., Meffert, L. M., Ritchie, M. G., Taylor, B. J., Wolf, J. B. and Moore, A. J. (2002). Genetic tools for studying adaptation and the evolution of behavior. *Am. Nat.* **160**, S143–S159.
- Bolnick, D. I., Svanbäck, R., Fordyce, J. A., Yang, L. H., Davis, J. M., Hulsey, C. D. and Forister, M. L. (2003). The ecology of individuals: incidence and implications of individual specialization. *Am. Nat.* **161**, 1–28.
- Boon, A. K., Réale, D. and Boutin, S. (2008). Personality, habitat use, and their consequences for survival in North American red squirrels *Tamiasciurus hudsonicus*. *Oikos* **117**, 1321–1328.
- Both, C., Dingemanse, N. J., Drent, P. J. and Tinbergen, J. M. (2005). Pairs of extreme avian personalities have highest reproductive success. *J. Anim. Ecol.* **74**, 667–674.
- Boyer, N., Réale, D., Marmet, J., Pisanu, B. and Chapuis, J.-L. (2010). Personality, space use and tick load in an introduced population of Siberian chipmunks *Tamias sibiricus*. *J. Anim. Ecol.* **79**, 538–547.
- Brommer, J. E. (2013). On between-individual and residual (co)variances in the study of animal personality: are you willing to take the “individual gambit”? *Behav. Ecol. Sociobiol.* **67**, 1027–1032.
- Browman, H. I. (2016). Applying organized scepticism to ocean acidification research. *ICES J. Mar. Sci.* **73**, 529–536.
- Burns, J. G. (2008). The validity of three tests of temperament in guppies (*Poecilia reticulata*). *J. Comp. Psychol.* **122**, 344–356.
- Byrnes, E. E. and Brown, C. (2016). Individual personality differences in Port Jackson sharks *Heterodontus portusjacksoni*. *J. Fish Biol.* **89**, 1142–1157.
- Careau, V. and Garland, T. (2012). Performance, personality, and energetics: correlation, causation, and mechanism. *Physiol. Biochem. Zool.* **85**, 543–571.
- Careau, V., Thomas, D., Humphries, M. M. and Réale, D. (2008). Energy metabolism and animal personality. *Oikos* **117**, 641–653.
- Careau, V., Montiglio, P.-O., Garant, D., Pelletier, F., Speakman, J. R., Humphries, M. M. and Réale, D. (2015). Energy expenditure and personality in wild chipmunks. *Behav. Ecol. Sociobiol.* **69**, 653–661.
- Carere, C. and van Oers, K. (2004). Shy and bold great tits (*Parus major*): body temperature and breath rate in response to handling stress. *Physiol. Behav.* **82**, 905–912.
- Carter, A. J., Heinsohn, R., Goldizen, A. W. and Biro, P. A. (2012). Boldness, trappability and sampling bias in wild lizards. *Anim. Behav.* **83**, 1051–1058.
- Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlshaw, G. and Heinsohn, R. (2013). Animal personality: what are behavioural ecologists measuring? *Biol. Rev.* **88**, 465–475.
- Chabot, D., Steffensen, J. F. and Farrell, A. P. (2016). The determination of standard metabolic rate in fishes. *J. Fish Biol.* **88**, 81–121.
- Chown, S. L. (2001). Physiological variation in insects: hierarchical levels and implications. *J. Insect Physiol.* **47**, 649–660.
- Clark, A. B. and Ehlinger, T. J. (1987). Pattern and adaptation in individual behavioral differences. In *Perspectives in Ethology* (ed. P. P. G. Bateson and P. H. Klopfer), pp. 1–47. Boston, MA: Springer.
- Cleasby, I. R., Nakagawa, S. and Schielzeth, H. (2015). Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. *Methods Ecol. Evol.* **6**, 27–37.
- Cockrem, J. F. (2007). Stress, corticosterone responses and avian personalities. *J. Ornithol.* **148**, 169–178.
- Cole, E. F. and Quinn, J. L. (2012). Personality and problem-solving performance explain competitive ability in the wild. *Proc. R. Soc. B Biol. Sci.* **279**, 1168–1175.
- Conrad, J. L., Weinersmith, K. L., Brodin, T., Saltz, J. B. and Sih, A. (2011). Behavioural syndromes in fishes: a review with implications for ecology and fisheries management. *J. Fish Biol.* **78**, 395–435.
- Conradsen, C., Walker, J. A., Perna, C. and McGuigan, K. (2016). Repeatability of locomotor performance and morphology–locomotor performance relationships. *J. Exp. Biol.* **219**, 2888–2897.
- Costantini, D., Carere, C., Caramaschi, D. and Koolhaas, J. M. (2008). Aggressive and non-aggressive personalities differ in oxidative status in selected lines of mice (*Mus musculus*). *Biol. Lett.* **4**, 119–122.
- Cote, J., Clobert, J., Brodin, T., Fogarty, S. and Sih, A. (2010). Personality-dependent dispersal: characterization, ontogeny and consequences for spatially structured populations. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 4065–4076.
- Crawford, D. L. and Oleksiak, M. F. (2007). The biological importance of measuring individual variation. *J. Exp. Biol.* **210**, 1613–1621.
- Dall, S. R. X. and Griffith, S. C. (2014). An empiricist guide to animal personality variation in ecology and evolution. *Front. Ecol. Evol.* **2**, 3.
- Dammhahn, M. (2012). Are personality differences in a small iteroparous mammal maintained by a life-history trade-off? *Proc. R. Soc. B Biol. Sci.* **279**, 2645–2651.
- Dammhahn, M. and Almeling, L. (2012). Is risk taking during foraging a personality trait? A field test for cross-context consistency in boldness. *Anim. Behav.* **84**, 1131–1139.
- Darveau, C.-A., Billardon, F. and Bélanger, K. (2014). Intraspecific variation in flight metabolic rate in the bumblebee *Bombus impatiens*: repeatability and functional determinants in workers and drones. *J. Exp. Biol.* **217**, 536–544.
- David, M. and Dall, S. R. X. (2016). Unravelling the philosophies underlying ‘animal personality’ studies: a brief re-appraisal of the field. *Ethology* **122**, 1–9.
- David, M., Auclair, Y. and Cézilly, F. (2011). Personality predicts social dominance in female zebra finches, *Taeniopygia guttata*, in a feeding context. *Anim. Behav.* **81**, 219–224.
- Diaz Pauli, B., Wiech, M., Heino, M. and Utne-Palm, A. C. (2015). Opposite selection on behavioural types by active and passive fishing gears in a simulated guppy *Poecilia reticulata* fishery. *J. Fish Biol.* **86**, 1030–1045.

- Dingemanse, N. J. and Araya-Ajoy, Y. G. (2015). Interacting personalities: behavioural ecology meets quantitative genetics. *Trends Ecol. Evol.* **30**, 88–97.
- Dingemanse, N. J. and Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *J. Anim. Ecol.* **82**, 39–54.
- Dingemanse, N. J. and Réale, D. (2005). Natural selection and animal personality. *Behaviour* **142**, 1159–1184.
- Dingemanse, N. J., Kazem, A. J. N., Réale, D. and Wright, J. (2010). Behavioural reaction norms: animal personality meets individual plasticity. *Trends Ecol. Evol.* **25**, 81–89.
- Dingemanse, N. J., Bouwman, K. M., van del Pol, M., van Overveld, T., Patrick, S. C., Matthysen, E. and Quinn, J. L. (2012). Variation in personality and behavioural plasticity across four populations of the great tit *Parus major*. *J. Anim. Ecol.* **81**, 116–126.
- DiRienzo, N. and Montiglio, P.-O. (2015). Four ways in which data-free papers on animal personality fail to be impactful. *Front. Ecol. Evol.* **3**, 23.
- Dohm, M. R. D. (2002). Repeatability estimates do not always set an upper limit to heritability. *Funct. Ecol.* **16**, 273–280.
- Dohm, M. R., Hayes, J. P. and Garland, T., Jr. (2001). The quantitative genetics of maximal and basal rates of oxygen consumption in mice. *Genetics* **159**, 267–277.
- Domenici, P., Lefrançois, C. and Shingles, A. (2007). Hypoxia and the antipredator behaviours of fishes. *Philos. Trans. R. Soc. B Biol. Sci.* **362**, 2105–2121.
- Domenici, P., Herbert, N. A., Lefrançois, C., Steffensen, J. F. and McKenzie, D. J. (2013). The effect of hypoxia on fish swimming performance and behaviour. In *Swimming Physiology of Fish: Towards Using Exercise to Farm a Fit Fish in Sustainable Aquaculture* (ed. P. A. Palstra and V. J. Planas), pp. 129–159. Berlin; Heidelberg: Springer Berlin Heidelberg.
- Falconer, D. S. and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Essex, UK: Longman.
- Farrell, W. J. and Wilczynski, W. (2006). Aggressive experience alters place preference in green anole lizards, *Anolis carolinensis*. *Anim. Behav.* **71**, 1155–1164.
- Freeman, H. D. and Gosling, S. D. (2010). Personality in nonhuman primates: a review and evaluation of past research. *Am. J. Primatol.* **72**, 653–671.
- Garamszegi, L. Z., Eens, M. and Török, J. (2009). Behavioural syndromes and trappability in free-living collared flycatchers, *Ficedula albicollis*. *Anim. Behav.* **77**, 803–812.
- Garland, T. (1988). Genetic basis of activity metabolism. I. Inheritance of speed, stamina, and antipredator displays in the garter snake *Thamnophis sirtalis*. *Evolution* **42**, 335–350.
- Garland, T. (1994). Quantitative genetics of locomotor behavior and physiology in a garter snake. In *Quantitative Genetic Studies of Behavioral Evolution* (ed. C. R. B. Boake), pp. 251–276. Chicago: University of Chicago Press.
- Gosling, S. D. (2001). From mice to men: what can we learn about personality from animal research? *Psychol. Bull.* **127**, 45–86.
- Gosling, S. D. (2008). Personality in non-human animals. *Soc. Personal. Psychol. Compass* **2**, 985–1001.
- Gould, S. J. (1985). The median isn't the message. *Discover* **6**, 40–42.
- Griffith, A. S., Guillelme, L. M. and Healy, S. D. (2015). Cognition and personality: an analysis of an emerging field. *Trends Ecol. Evol.* **30**, 207–214.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalised linear mixed models: The MCMCglmm R package. *J. Stat. Soft.* **33**, 1–22.
- Hamilton, T. J., Holcombe, A. and Tresguerres, M. (2014). CO<sub>2</sub>-induced ocean acidification increases anxiety in Rockfish via alteration of GABA<sub>A</sub> receptor functioning. *Proc. R. Soc. B Biol. Sci.* **281**, 20132509.
- Harris, S., Ramnarine, I. W., Smith, H. G. and Pettersson, L. B. (2010). Picking personalities apart: estimating the influence of predation, sex and body size on boldness in the guppy *Poecilia reticulata*. *Oikos* **119**, 1711–1718.
- Hayes, J. P. and Jenkins, S. H. (1997). Individual variation in mammals. *J. Mammal.* **78**, 274–293.
- Hayes, J. P., Speakman, J. R. and Racey, P. A. (1992). Sampling bias in respirometry. *Physiol. Zool.* **65**, 604–619.
- Henderson, C. R. (1982). Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions. *Biometrics* **38**, 623–640.
- Herborn, K. A., Macleod, R., Miles, W. T. S., Schofield, A. N. B., Alexander, L. and Arnold, K. E. (2010). Personality in captivity reflects personality in the wild. *Anim. Behav.* **79**, 835–843.
- Houle, D. (1991). Genetic covariance of fitness correlates: what genetic correlations are made of and why it matters. *Evolution* **45**, 630–648.
- Huntingford, F. A. (1976). The relationship between anti-predator behaviour and aggression among conspecifics in the three-spined stickleback, *Gasterosteus aculeatus*. *Anim. Behav.* **24**, 245–260.
- Jolles, J. W., Fleetwood-Wilson, A., Nakayama, S., Stumpe, M. C., Johnstone, R. A. and Manica, A. (2015). The role of social attraction and its link with boldness in the collective movements of three-spined sticklebacks. *Anim. Behav.* **99**, 147–153.
- Jornod, M. and Roche, D. G. (2015). Inter- vs intra-individual variation and temporal repeatability of escape responses in the coral reef fish *Amblyglyphidodon curacao*. *Biol. Open* **4**, 1395–1399.
- Killen, S. S., Marras, S., Metcalfe, N. B., McKenzie, D. J. and Domenici, P. (2013). Environmental stressors alter relationships between physiology and behaviour. *Trends Ecol. Evol.* **28**, 651–658.
- Killen, S. S., Nati, J. J. H. and Suski, C. D. (2015). Vulnerability of individual fish to capture by trawling is influenced by capacity for anaerobic metabolism. *Proc. R. Soc. B Biol. Sci.* **282**, 20150603.
- Killen, S. S., Adriaenssens, B., Marras, S., Claireaux, G. and Cooke, S. J. (2016). Context dependency of trait repeatability and its relevance for management and conservation of fish populations. *Conserv. Physiol.* **4**, cow007.
- Kolok, A. S. (1999). Interindividual variation in the prolonged locomotor performance of ectothermic vertebrates: a comparison of fish and herpetofaunal methodologies and a brief review of the recent fish literature. *Can. J. Fish. Aquat. Sci.* **56**, 700–710.
- Koolhaas, J. M., Korte, S. M., De Boer, S. F., Van Der Vegt, B. J., Van Reenen, C. G., Hopster, H., De Jong, I. C., Ruis, M. A. W. and Blokhuis, H. J. (1999). Coping styles in animals: current status in behavior and stress-physiology. *Neurosci. Biobehav. Rev.* **23**, 925–935.
- Koski, S. E. (2014). Broader horizons for animal personality research. *Front. Ecol. Evol.* **2**, 70.
- Koski, S. E. and Burkart, J. M. (2015). Common marmosets show social plasticity and group-level similarity in personality. *Sci. Rep.* **5**, 8878.
- Kurvers, R. H. J. M., Eijkelkamp, B., van Oers, K., van Lith, B., van Wieren, S. E., Ydenberg, R. C. and Prins, H. H. T. (2009). Personality differences explain leadership in barnacle geese. *Anim. Behav.* **78**, 447–453.
- Kurvers, R. H. J. M., Van Oers, K., Nolet, B. A., Jonker, R. M., Van Wieren, S. E., Prins, H. H. T. and Ydenberg, R. C. (2010). Personality predicts the use of social information. *Ecol. Lett.* **13**, 829–837.
- Lacy, R. C., Alaks, G. and Walsh, A. (2013). Evolution of *Peromyscus leucopus* mice in response to a captive environment. *PLoS ONE* **8**, e72452.
- Laming, S. R., Jenkins, S. R. and McCarthy, I. D. (2013). Repeatability of escape response performance in the queen scallop, *Aequipecten opercularis*. *J. Exp. Biol.* **216**, 3264–3272.
- Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. and Getz, W. M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359.
- Lopes, P. C. (2016). Why are behavioral and immune traits linked? *Horm. Behav.* <http://dx.doi.org/10.1016/j.yhbeh.2016.09.008>.
- Lopes, P. C., Adelman, J., Wingfield, J. C. and Bentley, G. E. (2012). Social context modulates sickness behavior. *Behav. Ecol. Sociobiol.* **66**, 1421–1428.
- Losos, J. B., Creer, D. A. and Schulte, J. A. II (2002). Cautionary comments on the measurement of maximum locomotor capabilities. *J. Zool.* **258**, 57–61.
- Lynch, M. and Walsh, J. B. (1998). *Genetics and Analysis of Quantitative traits*. Sunderland, Massachusetts: Sinauer Associates.
- MacKenzie, S., Ribas, L., Pilarczyk, M., Capdevila, D. M., Kadri, S. and Huntingford, F. A. (2009). Screening for coping style increases the power of gene expression studies. *PLoS ONE* **4**, e5314.
- Martin, J. G. A., Nussey, D. H., Wilson, A. J. and Réale, D. (2011). Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression models. *Methods Ecol. Evol.* **2**, 362–374.
- Mathot, K. J. and Dingemanse, N. J. (2015). Energetics and behavior: unrequited needs and new directions. *Trends Ecol. Evol.* **30**, 199–206.
- McDougall, P. T., Réale, D., Sol, D. and Reader, S. M. (2006). Wildlife conservation and animal temperament: causes and consequences of evolutionary change for captive, reintroduced, and wild populations. *Anim. Conserv.* **9**, 39–48.
- McGraw, K. O. and Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* **1**, 30–46.
- Metcalfe, N. B., Van Leeuwen, T. E. and Killen, S. S. (2016). Does individual variation in metabolic phenotype predict fish behaviour and performance? *J. Fish Biol.* **88**, 298–321.
- Misslin, R. and Cigrang, M. (1986). Does neophobia necessarily imply fear or anxiety? *Behav. Processes* **12**, 45–50.
- Morrissey, M. B. and Liefing, M. (2016). Variation in reaction norms: statistical considerations and biological interpretation. *Evolution* **70**, 1944–1959.
- Morton, F. B., Lee, P. C. and Buchanan-Smith, H. M. (2013). Taking personality selection bias seriously in animal cognition research: a case study in capuchin monkeys (*Sapajus apella*). *Anim. Cogn.* **16**, 677–684.
- Mowles, S. L., Cotton, P. A. and Briffa, M. (2012). Consistent crustaceans: the identification of stable behavioural syndromes in hermit crabs. *Behav. Ecol. Sociobiol.* **66**, 1087–1094.
- Nakagawa, S. and Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol. Rev.* **85**, 935–956.
- Natarajan, D., de Vries, H., Saaltink, D.-J., de Boer, S. F. and Koolhaas, J. M. (2009). Delineation of violence from functional aggression in mice: an ethological approach. *Behav. Genet.* **39**, 73–90.
- Neumann, C., Agil, M., Widdig, A. and Engelhardt, A. (2013). Personality of wild male crested macaques (*Macaca nigra*). *PLoS ONE* **8**, e69383.
- Niemelä, P. T. and Dingemanse, N. J. (2014). Artificial environments and the study of 'adaptive' personalities. *Trends Ecol. Evol.* **29**, 245–247.



- Niemelä, P. T., Lattenkamp, E. Z. and Dingemanse, N. J. (2015). Personality-related survival and sampling bias in wild cricket nymphs. *Behav. Ecol.* **26**, 936–946.
- Norin, T., Malte, H. and Clark, T. D. (2016). Differential plasticity of metabolic rate phenotypes in a tropical fish facing environmental change. *Funct. Ecol.* **30**, 369–378.
- Nussey, D. H., Wilson, A. J. and Brommer, J. E. (2007). The evolutionary ecology of individual phenotypic plasticity in wild populations. *J. Evol. Biol.* **20**, 831–844.
- Ou, M., Hamilton, T. J., Eom, J., Lyall, E. M., Gallup, J., Jiang, A., Lee, J., Close, D. A., Yun, S.-S. and Brauner, C. J. (2015). Responses of pink salmon to CO<sub>2</sub>-induced aquatic acidification. *Nature Clim. Change* **5**, 950–955.
- Øverli, Ø., Sørensen, C., Pulman, K. G. T., Pottinger, T. G., Korzan, W., Summers, C. H. and Nilsson, G. E. (2007). Evolutionary background for stress-coping styles: relationships between physiological, behavioral, and cognitive traits in non-mammalian vertebrates. *Neurosci. Biobehav. Rev.* **31**, 396–412.
- Petelle, M. B., Martin, J. G. A. and Blumstein, D. T. (2015). Heritability and genetic correlations of personality traits in a wild population of yellow-bellied marmots (*Marmota flaviventris*). *J. Evol. Biol.* **28**, 1840–1848.
- Pike, T. W., Samanta, M., Lindström, J. and Royle, N. J. (2008). Behavioural phenotype affects social interactions in an animal network. *Proc. R. Soc. B Biol. Sci.* **275**, 2515–2520.
- Pollock, M. S., Clarke, L. M. J. and Dubé, M. G. (2007). The effects of hypoxia on fishes: from ecological relevance to physiological effects. *Environ. Rev.* **15**, 1–14.
- Réale, D., Reader, S. M., Sol, D., McDougall, P. T. and Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biol. Rev.* **82**, 291–318.
- Réale, D., Dingemanse, N. J., Kazem, A. J. N. and Wright, J. (2010). Evolutionary and ecological approaches to the study of personality. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 3937–3946.
- Rupia, E. J., Binning, S. A., Roche, D. G. and Lu, W. (2016). Fight-flight or freeze-hide? Personality and metabolic phenotype mediate physiological defence responses in flatfish. *J. Anim. Ecol.* **85**, 927–937.
- Schuett, W., Tregenza, T. and Dall, S. R. X. (2010). Sexual selection and animal personality. *Biol. Rev.* **85**, 217–246.
- Sih, A., Bell, A. and Johnson, J. C. (2004a). Behavioral syndromes: an ecological and evolutionary overview. *Trends Ecol. Evol.* **19**, 372–378.
- Sih, A., Bell, A. M., Johnson, J. C. and Ziemba, R. E. (2004b). Behavioral syndromes: an integrative overview. *Q. Rev. Biol.* **79**, 241–277.
- Sokal, R. R. and Rohlf, F. J. (1995). *Biometry: The Principles and Practice of Statistics in Biological Research*, 3rd edn. New York: W.H. Freeman and Company.
- Sørensen, C., Johansen, I. B. and Øverli, Ø. (2013). Neural plasticity and stress coping in teleost fishes. *Gen. Comp. Endocrinol.* **181**, 25–34.
- Stamps, J. A. and Groothuis, T. G. G. (2010). Developmental perspectives on personality: implications for ecological and evolutionary studies of individual differences. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 4029–4041.
- Stamps, J. A., Briffa, M. and Biro, P. A. (2012). Unpredictable animals: individual differences in intraindividual variability (IIV). *Anim. Behav.* **83**, 1325–1334.
- Stevenson-Hinde, J., Stillwell-Barnes, R. and Zunz, M. (1980). Subjective assessment of rhesus monkeys over four successive years. *Primates* **21**, 66–82.
- Stuber, E. F., Araya-Ajoy, Y. G., Mathot, K. J., Mutzel, A., Nicolaus, M., Wijmenga, J. J., Mueller, J. C. and Dingemanse, N. J. (2013). Slow explorers take less risk: a problem of sampling bias in ecological studies. *Behav. Ecol.* **24**, 1092–1098.
- Toms, C. N., Echevarria, D. J. and Jouandot, D. J. (2010). A methodological review of personality-related studies in fish: focus on the shy-bold axis of behavior. *Int. J. Comp. Psychol.* **23**, 1–25.
- Tryon, R. C. (1942). Individual differences. In *Comparative Psychology* (ed. F. A. Moss), pp. 330–365. New York: Prentice-Hall.
- van de Pol, M. and Wright, J. (2009). A simple method for distinguishing within-versus between-subject effects using mixed models. *Anim. Behav.* **77**, 753–758.
- van Noordwijk, A. J. and de Jong, G. (1986). Acquisition and allocation of resources: their influence on variation in life history tactics. *Am. Nat.* **128**, 137–142.
- Verbeek, M. E. M., Drent, P. J. and Wiepkema, P. R. (1994). Consistent individual differences in early exploratory behaviour of male great tits. *Anim. Behav.* **48**, 1113–1121.
- Videler, M., Bonneaud, C., Cornette, R. and Herrel, A. (2014). Exploration syndromes in the frog *Xenopus (Silurana) tropicalis*: correlations with morphology and performance? *J. Zool.* **294**, 206–213.
- Westneat, D. F., Hatch, M. I., Wetzel, D. P. and Ensminger, A. L. (2011). Individual variation in parental care reaction norms: integration of personality and plasticity. *Am. Nat.* **178**, 652–667.
- Williams, T. D. (2008). Individual variation in endocrine systems: moving beyond the 'tyranny of the Golden Mean'. *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 1687–1698.
- Wilson, A. D. M. and Krause, J. (2012). Personality and metamorphosis: is behavioral variation consistent across ontogenetic niche shifts? *Behav. Ecol.* **23**, 1316–1323.
- Wilson, D. S., Coleman, K., Clark, A. B. and Biederman, L. (1993). Shy-bold continuum in pumpkinseed sunfish (*Lepomis gibbosus*): an ecological study of a psychological trait. *J. Comp. Psychol.* **107**, 250–260.
- Wilson, D. S., Clark, A. B., Coleman, K. and Dearstyne, T. (1994). Shyness and boldness in humans and other animals. *Trends Ecol. Evol.* **9**, 442–446.
- Wilson, A. D. M., Brownscombe, J. W., Krause, J., Krause, S., Gutowsky, L. F. G., Brooks, E. J. and Cooke, S. J. (2015). Integrating network analysis, sensor tags, and observation to understand shark ecology and behavior. *Behav. Ecol.* **26**, 1577–1586.
- Wolak, M. E., Fairbairn, D. J. and Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods Ecol. Evol.* **3**, 129–137.
- Wolf, M. and Weissing, F. J. (2012). Animal personalities: consequences for ecology and evolution. *Trends Ecol. Evol.* **27**, 452–461.
- Zylberberg, M., Klasing, K. C. and Hahn, T. P. (2014). In house finches, *Haemorrhous mexicanus*, risk takers invest more in innate immune function. *Anim. Behav.* **89**, 115–122.