BIOLOGICAL REVIEWS
Cambridge Philosophical Society

# Animal personality: what are behavioural ecologists measuring?

| | |
|---|---|
| Journal: | *Biological Reviews* |
| Manuscript ID: | BRV-03-2012-0048.R2 |
| Manuscript Type: | Original Article |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | Carter, Alecia; Australian National University, Fenner School of Environment and Society<br>Feeney, William; Australian National University,<br>Marshall, Harry; Institute of Zoology,<br>Cowlishaw, Guy; Institute of Zoology, Zoological Society of London<br>Heinsohn, Robert; Australian National University, Fenner School of Environment and Society |
| Keywords: | behavioural syndromes, boldness, exploration, methodology, animal personality, risk-taking |
| | |

# Animal personality: what are behavioural ecologists measuring?

Alecia J. Carter[1,2,3], William E. Feeney[4], Harry H. Marshall[2,5], Guy Cowlishaw[2] & Robert Heinsohn[1]

[1]The Fenner School of Environment and Society, The Australian National University, Acton, Canberra, ACT 0200 Australia

[2] The Institute of Zoology, Zoological Society of London, Regent's Park, NW1 4RY, London, U.K.

[3] Current address: Large Animal Research Group, Department of Zoology, The University of Cambridge, Cambridge, UK, CB2 3EJ E

[4] Evolution, Ecology and Genetics, The Research School of Biology, The Australian National University, Acton, Canberra, ACT 0200 Australia

[5] Division of Ecology and Evolution, Department of Life Sciences, Imperial College London, Silwood Park, Ascot, Berkshire, SL5 7PY, U.K.

Corresponding author: Alecia J. Carter: Large Animal Research Group, Department of Zoology, The University of Cambridge, Cambridge, UK, CB2 3EJ E: ac854@cam.ac.uk T: +44 1223 769277

# Animal personality: what are behavioural ecologists measuring?

The discovery that an individual may be constrained, and even behave sub-optimally, because of its personality type has fundamental implications for understanding individual- to group-level processes. Despite recent interest in the study of animal personalities within behavioural ecology, the field is fraught with conceptual and methodological difficulties inherent in any young discipline. We review the current agreement of definitions and methods used in personality studies across taxa and systems, and find that current methods risk misclassifying traits. Fortunately, these problems have been faced before by other similar fields during their infancy, affording important opportunities to learn from past mistakes. We review the tools that were developed to overcome similar methodological problems in psychology. These tools emphasise the importance of attempting to measure animal personality traits using multiple tests and the care that needs to be taken when interpreting correlations between personality traits or their tests. Accordingly, we suggest an integrative theoretical framework that incorporates these tools to facilitate a robust and unified approach in the study of animal personality.

Keywords: animal personality, behavioural syndromes, boldness, exploration, methods, risk-taking

CONTENTS

## I. INTRODUCTION

Behavioural ecology is the study of the ecological and evolutionary bases for animal behaviour and emphasises the individual as the fundamental unit of analysis. Each sub-discipline in behavioural ecology uses rigorous evolutionary reasoning to explain how animals use behaviour to deal with their intrinsic and extrinsic environments. Such thinking has made behavioural ecology a fast-paced field (Caro & Sherman, 2011). However, the use of optimization theory in this thinking has led to the assumption that variation in animal behaviour is either centred on a single adaptive optimum, or on two or more co-existing evolutionarily stable strategies (Weiss & Adams, In press). Variation around these optima has been traditionally viewed as noise (Mather & Anderson, 1993). Recent research in behavioural ecology challenges this view and proposes a 'new' concept that can explain variation in behaviour including sub-optimal tendencies: animal personality (Réale et al., 2007). Animal personality refers to between-individual differences in behaviour that persist through time (Biro & Stamps, 2008; alternatively, behavioural type: Sih et al., 2004b). A related concept, behavioural syndromes, goes one step further, referring to individual-level differences in correlations between personality traits or behaviours (table 1; see also Stamps & Groothuis, 2010). Here we will refer to animal personality in its broadest sense (as 'behavioural variation between individuals'), encompassing the concept of behavioural syndromes throughout. Further, we use the term personality 'trait' in the behavioural ecological sense as a particular aspect of an individual's behavioural repertoire, such as aggression or boldness (for a complete glossary of terms used in this article, see table 1).

Several theories have been proposed to explain the apparently suboptimal behavioural tendencies associated with animal personalities (Dingemanse & Wolf, 2010; Wolf & Weissing, 2010). For example, theories based on cost-benefit trade-offs predict that a bolder individual may receive a benefit by outcompeting conspecifics to gain greater access to resources (Pruitt, Riechert & Jones, 2008; Short & Petren, 2008), but bolder animals may also take more risks making them more

susceptible to predation (Bremner-Harrison, Prodohl & Elwood, 2004; Carter, Goldizen & Tromp, 2010). Animal personalities are associated with differences in fitness and are partly heritable, which raises important questions about how variation in personality is maintained in natural populations (reviewed in Bell, 2007). Despite this, there is still relatively little known about the causes and consequences of animal personality, from either proximate or ultimate perspectives.

Behavioural ecologists are currently asking three broad questions regarding animal personality: 1) Why do individuals vary consistently in their behaviour, in some contexts to the point of sub-optimality? 2) If selection 'pushes' behaviour toward one or more optimal strategies, how is further variation in animal behaviour maintained? 3) Why do cross-context behavioural correlations (behavioural syndromes) occur? Empiricists interested in any of these broad theoretical questions must first measure behaviours to establish whether they exhibit between-individual differences. Despite a recent surge in popularity within behavioural ecology (Réale et al., 2010), the study of animal personalities is fraught with conceptual and methodological difficulties inherent in any young field. Perhaps most telling is the confusion about how personality traits are defined and measured (see below). This confusion can lead to mislabelling traits and misinterpreting results, putting the development of animal personality theory at risk. This in turn raises two fundamental questions: Are researchers who are attempting to test the same personality traits in different taxa actually measuring the same thing? And if not, does this have implications for comparing the results of animal personality studies? The methodological and conceptual problems facing the field of animal personality theory are by no means novel. In fact, many concerns within the behavioural ecology approach to animal personality have been raised in other fields previously (Weiss & Adams, In press). However, no proposed methodological or conceptual frameworks offer solutions to the myriad issues identified. Here, we will review the current issues raised regarding the definitions and methods used to measure animal personality across taxa and systems, and demonstrate that the tools to overcome these problems have been developed previously in the psychological literatures. Our aim is to highlight teething problems in the field, and by heeding the lessons learned elsewhere, encourage a unified

approach to future animal personality studies through the use of research tools that have been successfully used elsewhere.

## II. PROBLEMS MEASURING PERSONALITY TRAITS

### (1) Many tests for one trait

Boldness is one of the most commonly measured personality traits (Conrad et al., 2011), but is perhaps the trait with least consensus over its definition. For instance, boldness has been interpreted as being the propensity to take risks, especially in novel situations (Coleman & Wilson, 1998; Toms, Echevarria & Jouandot, 2010), whereas Réale et al. (2007) defined boldness as an individual's response to a risky situation alone, and excluded reactions to novel situations and stimuli altogether.

Tests of boldness reflect the confusion inherent in its definition. Boldness has been tested by quantifying behavioural responses to novel objects, responses to a novel environment, and responses to predation risk (Toms et al., 2010). However, these three types of test are not necessarily comparable and demonstrate a lack of standardised tests for quantifying the behaviour (Budaev & Brown, 2011;Conrad et al., 2011; Toms et al., 2010; Carter et al., In press). Fox et al. (2009), for example, measured both novel object exploration and exploration of a novel environment in their study of mountain chickadees (*Poecile gambeli*). The authors did not find a correlation between the traits and highlighted that they were not interchangeable, as suggested by other studies (see references in Fox et al., 2009). A similar result was found in a study of pumpkinseed sunfish (*Lepomis gibbosus*); reaction to a threatening novel stimulus did not correlate with response to a novel food (Coleman & Wilson, 1998). Both studies suggested that the behaviour tested was specific to the context that it was tested in. Dingemanse et al. (2007) encountered a similar problem when attempting to test shyness-boldness in sticklebacks (*Gasterosteus aculeatus*) using antipredator behaviour (response to risk). They compared the exploratory behaviour of sticklebacks when exposed to a

predator housed in an adjacent compartment and to their behaviour when exposed to an empty compartment. The authors found no significant differences and relabelled this behaviour as a measure of "exploration-avoidance of an altered environment". This study not only highlights the importance of using controls, but also that incorrectly labelling personality traits may be a common problem. We know of only one study that has directly investigated the relationship between multiple tests of boldness in the same individuals with the intent of describing the validity of the tests. Burns (2008; but see also Bergvall et al., 2011, Carter et al., In press) measured the responses of individual guppies (*Poecilia reticulata*) to three experiments intended to measure boldness: an open-field test, an emergence test and a novel-object test (table 2). The open-field and emergence tests correlated with each other, but neither of these tests correlated with the results of the novel-object test, leading Burns (2008) to conclude that the novel object test should not be used to assess boldness in guppies.

A further important consideration is that a test for one species/taxon is not necessarily appropriate as a test for another (Weiss & Adams, In press). For example, while some animals may perceive a greater risk of predation in an open as opposed to closed habitat (Blumstein & Daniel, 2003), the same may not be true for species that are predated primarily in closed habitats (Whittingham et al., 2004). Although comparability between studies is desirable, if a test is to be adapted to a new system, every effort should be made to make the test as ecologically relevant as possible.

**(2) One test for many traits**

Boldness is commonly studied but remains ill-defined. Different investigative methods of boldness do not always correlate, indicating that 'boldness' might encompass several distinct behavioural traits. However, the reverse is also true. Many traits can be measured with one test. The open-field test is frequently used to measure activity-exploration (for example, Boyer et al., 2010) or boldness (for example, Brown & Braithwaite, 2004) and involves quantifying aspects of an animal's behaviour after being introduced to an open and novel environment. This simple method is thus used frequently but

there are fundamental problems with its interpretation in different circumstances. An individual can be introduced to the open-field by force (by placing it in the environment with no opportunity to escape), or be offered the flexibility to explore the open-field freely (with access to a refuge, by having its home cage placed in or connecting to the environment) (Crusio, 2001; Walsh & Cummins, 1976). However, behaviour in forced *versus* free exploration contexts may not necessarily correspond to the same traits (Kavanau, 1967; Lester, 1968; Misslin & Cigrang, 1986). Free open-field tests are more likely to measure voluntary exploration/curiosity and information-gathering behaviour, while forced open-field tests are more likely to measure fear or anxiety (or both) (Misslin & Cigrang, 1986). In a review of the use of open field tests, Walsh and Cummins (1976) highlighted that the test has been interpreted as measuring emotionality, fear, gregariousness (if more than one individual was used) and exploration. Taken together, these studies indicate that care needs to be taken when using the open-field test, and other tests where protocol differences exist, to test personality traits, especially when comparing across multiple studies.

One test can simultaneously be influenced by and thus measure two or more personality traits (Réale et al., 2007). This 'overlap' can become especially clear when multiple measurements are taken and data reduction techniques, such as factor or principal components analysis, are used (Gorsuch, 1983). If the test measured only one trait, then only one factor/component explaining substantial variation in the measurements would be identified. However, it is more often the case that two or more of the resultant factors/components have this explanatory power, such that two or three unrelated traits will emerge (assuming that each factor/component represents a trait) (for example, see Carter et al., 2011). Although a test that directly measures a targeted trait may be a desirable goal, in reality a test will likely be influenced by multiple traits at the same time (Réale et al., 2007). We deal with this issue further in the next section.

## III. LESSONS FROM OTHER FIELDS

If the multiple 'standard' tests for boldness are not comparable and one test is able to measure two traits concurrently, which tests should animal personality researchers use and how should their results be interpreted? Fortunately these types of problems have been faced before by other fields during their infancy, affording excellent opportunities to adopt their solutions and avoid the likely stagnation associated with inertia in updating methods and concepts (Lockard, 1971; see also Beach, 1950; Boice, 1971; Hodos & Campbell, 1969). Here we consider the use of psychometrics, the theory of psychological measurement, which has helped personality research in the fields of comparative psychology, the psychological study of animal behaviour, and differential psychology, the investigation of individual differences in behaviour (John, Robins & Pervin, 2008). We first outline an important caveat when considering psychometric applications to the study of animal personality within behavioural ecology. We then review important tools and theories that could be applied to, and should not be overlooked, within behavioural ecology.

## (1) An important caveat

Three important points regarding the differences between the psychological and behavioural ecological literatures that may impede communication between the two fields are as follows. First, there are important differences in terminology (Koski, 2011; Uher, 2011). Uher (2011) outlines these differences in detail; one clear example is the term 'trait'. The term is used in behavioural ecology (and by us in this article) to mean a measured aspect of an individual's behaviour (e.g. the rate a behaviour is observed) while in psychology the use is more abstract and describes a construct (see below). Further confusion is added because these terminologies are often debated within each field (for example, see Lay, 1973). Second, psychologists have remarked that behavioural ecology has a narrow, restricted and incomprehensive view of personality variation (Uher, 2011; Weiss & Adams, In press). However, this view relates to differences in approach and research goals: behavioural ecologists take a reductionist approach to animal personality whereas psychologists take a more holistic approach (for a longer discussion of these differences, see Koski, 2011; Uher, 2011; Weiss &

Adams, In press). Finally, there is much historical (and perhaps contemporary) conceptual debate within the psychometric literature (see definitions in Campbell & Fiske, 1959; Cronbach & Meehl, 1955; and discussion in Uher, 2011). While these differences may seem daunting, the psychological and psychometric literatures nonetheless provide an established discussion of measurement problems currently systemic in animal personality research in behavioural ecology.

**(2) Test validity**

Validity refers to the degree to which a test measures the targeted trait (Burns, 2008; see also Réale et al., 2007). A frequently recurring critique of personality psychology involves the validity of the tests that are used to measure the trait of interest (Duckworth & Kern, 2011). One of the first ways to remedy this problem is to use multiple measurements for multiple traits, and investigate correlations between the measurements (Campbell & Fiske, 1959). However, it is important to be aware that correlations between two measurements could be influenced by shared method variance (Campbell & Fiske, 1959). That is, systematic variation in the type of method used to obtain the measurements could cause tests using similar methods to be more correlated than tests using different methods. This is a particular problem for behavioural ecologists, as most of the methods used are similar (behavioural observations). Method variance can be explicitly estimated by using hierarchically nested models, however, when more than one method is used (Widaman, 1985). Three other key aspects to trait measurement are particularly relevant to behavioural ecologists: Reliability, and Convergent and Discriminant validities (table 3) (Burns, 2008).

*(a) Reliability*

Reliability refers to the consistency of a measure through time, across contexts or across raters/observers. It estimates whether there is agreement between repeated tests of the same nature. Reliability differs from validity because a test may be reliable but not be valid. A common analogy

used to demonstrate this difference (Nunnally 1978) is an archer (the researcher) trying to hit a target (the trait) with a bow and arrow (the test). Validity refers to how close to the centre of the target the arrows land while reliability refers to how close together the arrows are clustered. Reliability is the first psychometric test employed in psychological research to assess a test's performance (Gosling, 2001). In animal personality studies this approach is common practice, and is known as consistency or, more commonly, repeatability (Bell, Hankison & Laskowski, 2009). In fact, behaviours must be repeatable to be considered personality traits within behavioural ecology (Réale et al., 2007; Sih, Bell & Johnson, 2004a).

*(b) Convergent and discriminant validity*

Convergent and discriminant validities rest at either end of a spectrum for validating trait tests (Campbell & Fiske, 1959). Convergent validity investigates whether two tests actually measure the same trait (that is, the measurements from both tests should correlate). Conversely, discriminant validity investigates whether two tests that are employed to measure different traits actually measure different traits (that is, the measurements from the tests should not correlate). In psychology, each measurement from a test is referred to as a trait-method unit and researchers are encouraged to use a multi-trait, multi-method approach to describe variation in personality (Campbell & Fiske, 1959). While it is possible in psychology to use multiple methods (such as behavioural scoring and self-assessment), behavioural ecologists are frequently restricted to one method (behavioural observations), but can use multiple tests. Studies that predict but fail to find a correlation between two tests usually conclude that the trait is context specific. An example would be Fox *et al.*'s (2009) study that failed to find a correlation between response to a novel object and response to a novel environment. However, we could alternatively conclude that the two tests measures different personality traits (Burns, 2008; Carter et al. In press), and so are highly discriminant (Campbell & Fiske, 1959; for alternative interpretations, see Cronbach & Meehl, 1955). Without the use of multiple tests for a given trait it is impossible to measure how convergent or discriminant the tests are.

This can result in, at the very least, confusion over what trait is actually being measured and, in the worst case, measurement of a trait that is interpreted incorrectly. Lessons from psychometrics and psychology (Duckworth & Kern, 2011; McCrae, 1982) therefore emphasise the importance of attempting to measure animal personality traits using multiple methods, or in the case of behavioural ecology, multiple tests (Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Uher, 2011; Weiss & Adams, In press). In behavioural ecology this is rarely the case (but see Bergvall et al., 2011; Burns, 2008; Carter et al., In press). The use of a multi-trait, multi-test approach (Campbell & Fiske, 1959) would allow assessment of a test's appropriateness and aid in further identification of personality traits and behavioural syndromes.

*(c) Other validities*

Two other validation techniques are commonly discussed in psychological research: face validity and construct validity (Nunnally, 1978). In the first case, a test that appears to measure the trait of interest is said to have face validity (i.e. the correlation between a trait's theoretical definition and the subject's response to the test). Although face validity is fundamental for a test to be used, differential psychologists have noted that it is frequently assumed and infrequently confirmed (Nevo, 1985). For example, for open-field studies, defecation was used as a measure of emotionality (response to fear/anxiety) as it had long been known that defecation/urination occurs during periods of emotional stress (that is, it has face validity; Hall, 1934). After some debate about the use of this test (Archer, 1973), it was later validated in a study that measured both defecation and physiological responses to the open-field test (Denenberg, 1973). It is apparent from psychology that face validity can be confounded with construct validity (see below) leading to the possible failure to identify reliable tests.

In the second case, construct validity, like face validity, is related to a theoretical understanding of personality, and investigates the theoretically predicted relationships between a variety of psychological traits and behaviours. That is, a construct is an *a priori*, theoretical idea of how

behaviours should be inter-correlated (Cronbach & Meehl, 1955; Nunnally, 1978). While behavioural syndromes may similarly investigate correlations between behaviours, construct validity raises an important conceptual point that is frequently overlooked in behavioural ecology (from Nunnally, 1978 p 104): while two behaviours may correlate, this is not proof that either behaviour measures the targeted trait/factor (see Section 3 below).

**(3) Jingle-jangle fallacies and trait definitions**

Confusion about over- or under-labelling traits is known in differential psychology as the jingle-jangle fallacy (Bell, 2007; Gosling, 2001; Uher, 2011). The 'jingle' fallacy refers to a single trait label that inadvertently describes two functionally different traits measured with different tests. The 'jangle' fallacy refers to two labels that actually measure the same trait (Block, 1995). Jingle-jangle fallacies can be pervasive and misleading, and there is a need to identify poor tests as early as possible (Block, 1995; Jacoby, 1978). For example, exploration of a novel object and exploration of a novel environment/open field may be given the same label: exploration. However, as noted above, if exploration of an open field is forced, the test may measure anxiety and not exploration, and we would have committed a jingle fallacy. In such circumstances, to avoid a jingle fallacy, each test might be given a different trait name. But this may generate another problem, namely confusion within animal personality studies and the risk that the two trait names could be seen as synonymous of the same trait: a jangle fallacy.

Winter et al. (1998) recognised the jingle fallacy among personality psychologists who were concurrently, but separately, studying psychological traits and motives (two different concepts within this field). While testing the convergent and discriminant validities of tests is essential, progress in answering proximate and ultimate questions about personality could potentially be gained more efficiently by using key validated tests that assess demonstrably independent traits. Winter et al. (1998) suggested simplifying the field of personality research after asking whether all of the tests

were required to understand personality. Reduction of the required number of tests could be achieved by using a phylogenetically controlled meta-analytical approach to identify those traits, and those tests used to measure them, which have been consistently validated in the literature. These traits could then form the basis of a general framework for the study of animal personality (Duckworth & Kern, 2011).

**(4) Interpreting tests**

Next we raise some issues regarding the interpretation of the results of personality tests. Assuming that reliability has been established, and method variance controlled for, there are two possible interpretations of the relationships between the results of tests of convergent and discriminant validity. These interpretations depend on the focus of the study: whether the tests or the traits being measured are of interest. In the first case, test validation for a single personality trait is the study's goal, and interpretation focuses on whether the two different tests are measuring the same or different traits (the 'personality trait-validation' interpretation, see table 3). In the second case, which is more typical in behavioural ecology, the two different tests are already assumed to measure different personality traits and the goal is to establish a behavioural syndrome, such that interpretation focuses on the relationship between the two tested behaviours (the 'behavioural-syndrome identification' interpretation, table 3). Thus a positive correlation between two tested behaviours can come about because 1) the tests measure the same personality trait (the tests have convergent validity) or 2) the two behaviours are linked by an underlying behavioural syndrome (Burns, 2008). Likewise, a lack of a correlation between two tested behaviours can come about because 3) the tests measure different personality traits (the tests have discriminant validity) or 4) the two behaviours are not linked in a behavioural syndrome (Sih et al., 2004a) (table 3).

If we wished to study exploration as a personality trait, for example, we could perform two tests, exploration of a novel object and exploration of a novel environment/open field, to search for convergent validity. Under the trait focussed interpretation, if these tests did correlate, we could say

that both tests measured exploration, but to be sure, we should perform one more test, such as aggression towards a conspecific that should not correlate with either exploration of a novel environment or of a novel object. Alternatively, if we found that the novel object and novel environment tests did not correlate, but that the novel object and aggression tests did, we could conclude that novel object and aggression measure the same trait, for example, boldness, with convergent validity and discriminant validity from exploration (novel environment) (table 4). Depending on the goal of the study (test validation or identification of a behavioural syndrome) two different interpretations of the results are possible, and care should be taken when investigating correlations between multiple tests.

One possible solution to the problem posed by correlations between test measurements might be found via trait orthogonality (mathematical independence). This concept, primarily employed by differential and human personality researchers, uses factor analysis or principal component analysis (PCA) to identify the orthogonal personality factors that are robust across investigations, samples and time (Goldberg, 1992; Goldberg, 1993). This allows explanation of each factor's observed patterns of variation and can be used to predict various aspects of an individual's behaviour such as job-related performance (Goldberg, 1993). The method has been successfully used to compare personality across species (for example Gosling, Kwan & John, 2003; Weiss et al., 2011) and contexts (for example King, Weiss, & Farmer, 2005) in differential and comparative psychology. While some cross-species studies (Carter & Feeney, 2012; Mettke-Hofmann et al., 2005; Webster, Ward & Hart, 2009) and cross-context studies (Sih, Kats & Maurer, 2003; Sih & Watters, 2005) have used this approach in behavioural ecology, they are rare and factor analysis is not a preferred method. However, we suggest that a factor analytic approach may be appropriate in behavioural ecology to 1) establish independent factors/axes of correlated personality traits (such as a boldness-aggression axis), and 2) investigate how these orthogonal axes impact on individual behavioural ecology, life history productivity and fitness (Biro & Stamps, 2008). Such a factorial approach would also have the benefit of reducing the problems associated with jingle-jangle fallacies.

Réale et al. (2007) recommended using a 'simplified terminology' for animal personality traits that has been widely adopted within animal personality studies (the 'Big Five' animal personality traits: boldness, exploration-avoidance, activity, aggressiveness and sociability). Réale et al. (2007) proposed these terms as categories of traits rather than as traits themselves, warning that the terminology was not exhaustive and it should be reviewed after sufficient research had been conducted. Again, meta-analytical approaches may prove fruitful when attempting to review Réale et al.'s (2007) terminology and factor orthogonality may be a desirable goal. Overall, we see the factorial approach as compatible with existing animal personality research: 1) behavioural syndromes aim to identify correlations between traits, and a factorial approach would identify orthogonal factors (behavioural syndromes) composed of components (traits), and 2) these syndromes/factors could be used to investigate both proximate and ultimate animal personality questions (for further discussion of hierarchical personality taxonomy, see Uher, 2011; Uher et al., 2011).

**(5) Measurement considerations**

While the previous sections have raised theoretical issues with choosing tests and interpreting their results, this section is concerned with decisions about what to measure. These considerations can seem trivial (and are thus frequently overlooked), but may have a substantial effect on the outcome of the study. First, there is the consideration of what method to use to collect the data. Currently there are three different methods used: subjective personality ratings, behavioural coding and experimentation (Jones & Gosling, 2005; Koski, 2011; Uher, 2008, 2011; Uher et al., 2011; Vazire et al., 2007).

Subjective assessments use ratings of multiple items, such as adjectives or behavioural descriptors, by observers familiar with individual animals to describe the dimensions encompassing multiple personality axes. Behavioural codings consist of recording the behaviour of a focal individual

according to a predetermined ethogram, that is, a list of discrete behaviours performed by the species under study during the animal's natural behaviour (Gosling, 2001; Vazire et al., 2007). Experimental assessments similarly record individual patterns of behaviour, but in response to controlled experimental stimuli, to assess variability in a 'limited' number of personality axes such as boldness, aggressiveness or sociability (Nettle & Penke, 2010; Sih et al., 2004b).

Combinations of these approaches are used by personality researchers in the fields of comparative and differential psychology and behavioural ecology, to varying degrees. Behavioural ecologists frequently use behavioural codings of natural behaviour and behaviour during experimentation to assess personality, while comparative psychologists often use ratings and behavioural codings of natural behaviour (hereafter, natural behaviour) and also use experimentation to assess personality (Uher, 2008; Uher & Asendorpf, 2008). Relationships are well established between natural behaviour and experimentation, and between natural behaviour and ratings (for example Carter et al., 2010; Konečná et al., 2008; Kurvers et al., 2010; Pederson, King & Landau, 2005), and there is some evidence that ratings are indicative of experimental responses as well (see Carter et al., 2012 and references therein). While each approach has benefits and drawbacks (Koski, 2011), the high correspondence between measures suggests simultaneous application of these methods may be beneficial (Uher & Asendorpf, 2008), at the very least to test convergent validity (Campbell & Fiske, 1959).

The second measurement consideration regards situational strength (Uher, 2011). A situation in behavioural ecology refers to the conditions at the time of the test (for example, temperature or available resources; Sih et al., 2004b). Situational strength refers to how much an individual's behaviour is influenced by the situation (Uher, 2011). Strong situations may leave little variation between individual's behaviour whereas weaker situations may allow more inter-individual differences to show (Mischel, 1973a,b). Accordingly, strong situations may be inappropriate as

behavioural tests given that little variation may exist between individuals due to floor and ceiling effects (Stamps & Groothuis, 2010; Uher, 2011). Further, cutting off data after a certain time (for example, see Carter et al., 2012) may result in biased personality estimates due to ceiling effects, the implications of which are rarely considered in behavioural ecology (but see Stamps, Briffa & Biro, 2012). For example, are all individual baboons (*Papio ursinus*) that handle a novel food item for longer than 150 seconds equally bold (Carter et al., 2012)?

Finally, there has been much debate in psychology about whether the situation or the person is more important during behavioural tests (Mischel, 1968; Mischel, 1973a,b; Mischel, 1999; Mischel, Shoda & Mendoza-Denton, 2002; Tett & Guterman, 2000; see also Uher, 2011), leading to the field of interactionist psychology. Far from always expecting behavioural consistency, interactionist psychologists expect that while some individuals will behave consistently between different situations there are certain situations that can cause different individuals to behave similarly (high situational strength) (Tett & Guterman, 2000). A concept similar to interactionist psychology in behavioural ecology is the behavioural reaction norm (Dingemanse et al., 2010b; Nussey, Wilson & Brommer, 2007) in which individual's responses to different situations or contexts are modelled using random regressions (Nussey et al., 2007; van de Pol & Wright, 2009). This has the potential to provide a useful conceptual framework for understanding interactions between animal identity and the situation interactions where the response variable is the same, however few animal personality researchers have considered individual by situation interactions thus far (for examples, see Betini & Norris, 2012; Dingemanse et al., 2012).

## IV. HOW TO MEASURE PERSONALITY TRAITS

We suggested that the indiscriminate use of 'standard' behavioural tests within animal personality studies may lead to the spurious labelling of personality traits, and have asked whether some studies are actually measuring what they intend to. Further, we explored, in fields such as comparative

psychology, the past experiences of the problems and pitfalls that the animal personality approach in behavioural ecology is now facing. In this section, we incorporate the tools from these other fields into suggestions or goals that can be integrated into future research programs. As an exciting and fast-moving field, animal personality research is potentially at risk from academic 'faddists' (termed Zeitgeister-Shysters: Denenberg 1969; cited in Fetterman, 1986). Therefore, if animal personality research is to maintain its rate of progress, it must develop a robust methodology including multiple trait tests, reliability and validation (Cronbach & Meehl, 1955; Uher, 2011; Weiss & Adams, In press), and continue to clearly identify research questions and hypotheses at their outset. Below we suggest methodological goals that include the tools discussed above, which can be incorporated into future animal personality studies.

*1. Consider test design.* We have reviewed the many conceptual and practical measurement issues associated with different tests. Before starting to collect data, researchers should consider:

a) Which method(s) should be used (subjective personality ratings, behavioural coding or experimentation).

b) Whether the test actually measures the targeted trait in that species? That is, does the test have situational relevance?

c) Situation strength and floor and ceiling effects. Is the stimulus situation too strong? Will having a cut-off in a weak situation affect an individual's position along a personality gradient?

2. *Develop multiple tests for each trait of interest in the study.* Multiple tests of a trait are necessary to establish the reliabilities and validities of the trait and the tests (step 3 below). For each trait that is to be investigated by the research question, we suggest consideration of the following:

a) Does the test have face validity?

b) Have the tests for the trait been used previously? Do the results of the previous studies suggest that these tests are applicable to this study and system?

c) Identify a primary test for a trait, and at least one other that can each be used to test for convergent or discriminant validities.

d) Make explicit predictions about where correlations should occur between the chosen tests if the tests are measuring the targeted traits i.e. consider construct validity before data are collected (see Section III.2.c above).

3. *Validate the tests used in the study.* Validation should comprise:

a) repeatability,

b) Ecological validity (Réale et al. 2007),

c) convergent validity (a correlation between two tests that theoretically measure the same trait) and,

d) discriminant validity (a lack of a correlation between two tests that are hypothesised to measure different traits) for each test.

Alternatively, to avoid Type I errors through testing each measurement against all other measurements (Dochtermann, 2010), principal components analysis or structural equation modelling (Loehlin, 1998; Dingemanse, Dochtermann & Wright, 2010a) can be used to investigate which test measurements load with others (but see Block, 1995 for a criticism of data reduction techniques for this purpose). We emphasise again that a correlation between two tests may come about by either the tests measuring the same trait or the measured traits being linked by an underlying behavioural syndrome, depending on whether a personality trait-validation or behavioural syndromes-identification focussed interpretation is being used (table 3). In this case, careful choice of the validation tests is essential even when investigating behavioural syndromes. Alternatively, a factor analytic

approach may be useful to integrate correlated traits such as a boldness-aggression factor. We can then ask Winter et al.'s (1998) modified questions for the tests:

a. Do the tests measure the same trait with different labels, or are they different traits?

b. How much do the tests overlap? For example, while exploration could be measured using distance travelled in an open field, this measurement may concurrently measure activity. Should activity be controlled for when using this as a measurement of exploration?

c. Are all of the tests required to understand animal personality traits?

If the tests do not load/correlate as predicted, consider the assumptions made.

## V. CONCLUSIONS

1. Throughout this review we have emphasised our view that the behavioural ecological approach to animal personality research is facing methodological and conceptual obstacles that may hinder its progress.

2. Current use of personality tests can be problematic as in some cases different tests may be measuring the same personality trait (many to one) whereas in other cases one test may measure many traits (one to many). This makes their interpretation difficult and limits the scope and comparability of current studies.

3. We have examined lessons learned by psychologists, and suggest tools that could be borrowed from the psychometric literature.

4. We make suggestions for how future studies might use these tools to work towards a more unified and robust model of animal personality. In this framework we suggest the use of multiple tests for measuring personality traits wherever possible, and urge that more consideration be given to interpreting the observed correlations between tests.

5. Despite the obstacles we outlined, we believe the tools exist to strengthen methodology in the field of animal personality research in behavioural ecology, and to further its exciting and rapid progress.

## Acknowledgements

The authors thank Alex Weiss, Sam Gosling and an anonymous reviewer for detailed comments on an earlier version of this work and Culum Brown for reading a preliminary outline. AJC is supported by a Fenner School of Environment and Society Studentship; WEF by a Research School of Biology studentship; HHM by a NERC Open CASE (NE/F013442/1) studentship with ZSL as CASE partner.

## References


ARCHER, J. (1973). Tests for emotionality in rats and mice: a review. *Animal Behaviour* **21,** 205-35.

BEACH, F. A. (1950). The Snark was a Boojum. *American Psychologist* **5,** 115-124.

BELL, A. M. (2007). Future directions in behavioural syndromes research. *Proceedings of the Royal Society Biological Sciences Series B* **274,** 755-761.

BELL, A. M., HANKISON, S. J. & LASKOWSKI, K. L. (2009). The repeatability of behaviour: a meta-analysis. *Animal Behaviour* **77,** 771-783.

BETINI, G. S. & NORRIS, D. R. (2012). The relationship between personality and plasticity in tree swallow aggression and the consequences for reproductive success. *Animal Behaviour* **83,** 137-143.

BIRO, P. A. & STAMPS, J. A. (2008). Are animal personality traits linked to life-history productivity? *Trends in Ecology & Evolution* **23,** 361-368.

BLOCK, J. (1995). A Contrarian View of the Five Factor Approach to Personality Description. *Psychological Bulletin* **117,** 187-215.

BLUMSTEIN, D. T. & DANIEL, J. C. (2003). Red kangaroos (*Macropus rufus*) receive an antipredator benefit from aggregation. *Acta Ethologica* **5,** 95-99.

BOICE, R. (1971). On the fall of comparative psychology. *American Psychologist* **26,** 858-859.

BOYER, N., REALE, D., MARMET, J., PISANU, B. & CHAPUIS, J.-L. (2010). Personality, space use and tick load in an introduced population of Siberian chipmunks *Tamias sibiricus*. *Journal of Animal Ecology* **79,** 538-547.

BREMNER-HARRISON, S. B., PRODOHL, P. A. & ELWOOD, R. W. (2004). Behavioural trait assessment as a release criterion: boldness predicts early death in a reintroduction programme of captive-bred swift fox (*Vulpes velox*). *Animal Conservation* **7,** 313-320.

BROWN, C. & BRAITHWAITE, V. A. (2004). Size matters: a test of boldness in eight populations of the poeciliid Brachyraphis episcopi. *Animal Behaviour* **68,** 1325-1329.

BUDAEV, S. & BROWN, C. (2011). Personality Traits and Behaviour. In *Fish Cognition and Behavior* (ed. C. Brown, K. N. Laland and J. Krause), pp. 135-165. Wiley-Blackwell, Oxford.

BURNS, J. G. (2008). The Validity of Three Tests of Temperament in Guppies (Poecilia reticulata). *Journal of Comparative Psychology* **122,** 344-356.

BUTLER, M. W., TOOMEY, M. B., MCGRAW, K. J. & ROWE, M. (2011). Ontogenetic immune challenges shape adult personality in mallard ducks. *Proceedings of the Royal Society Biological Sciences Series B***,** Available online.

CAMPBELL, D. T. & FISKE, D. W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin* **56,** 81-105.

CARO, T. & SHERMAN, P. W. (2011). Endangered species and a threatened discipline: behavioural ecology. *Trends in Ecology & Evolution* **26,** 111-118.

CARTER, A. J. & FEENEY, W. E. (2012) Taking a Comparative Approach: Analysing Personality as a Multivariate behavioural Response across Species. *PLoS ONE* **7**, e42440.

CARTER, A. J., GOLDIZEN, A. W. & TROMP, S. A. (2010). Agamas exhibit behavioral syndromes: bolder males bask and feed more but may suffer higher predation. *Behavioral Ecology* **21,** 655-661.

CARTER, A. J., MARSHALL, H., HEINSOHN, R. & COWLISHAW, G. (In press). How not to measure boldness: novel object and antipredator responses are not the same in wild baboons. *Animal Behaviour*

CARTER, A. J., MARSHALL, H., HEINSOHN, R. & COWLISHAW, G. (2012). Evaluating animal personalities: do observer assessments and experimental tests measure the same thing? *Behavioral Ecology and Sociobiology* **66,** 153-160.

COLEMAN, K. & WILSON, D. S. (1998). Shyness and boldness in pumpkinseed sunfish: individual differences are context-specific. *Animal Behaviour* **56,** 927-936.

CONRAD, J. L., WEINERSMITH, K. L., BRODIN, T., SALTZ, J. B. & SIH, A. (2011). Behavioural syndromes in fishes: a review with implications for ecology and fisheries management. *Journal of Fish Biology* **78,** 395-435.

CRONBACH, L. J. & MEEHL, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin* **52,** 281-302.

CRUSIO, W. E. (2001). Genetic dissection of mouse exploratory behaviour. *Behavioural Brain Research* **125,** 127-132.

DENENBERG, V. H. (1973). Open-Field Behavior in the Rat: what does it mean? *Annals of the New York Academy of Sciences* **159,** 852-859.

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***,** 116-126.

DINGEMANSE, N. J., DOCHTERMANN, N. & WRIGHT, J. (2010a). A method for exploring the structure of behavioural syndromes to allow formal comparison within and between data sets. *Animal Behaviour* **79,** 439-450.

DINGEMANSE, N. J., KAZEM, A. J. N., RÉALE, D. & WRIGHT, J. (2010b). Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution* **25,** 81-89.

DINGEMANSE, N. J. & WOLF, M. (2010). Recent models for adaptive personality differences: a review. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365,** 3947-3958.

DINGEMANSE, N. J., WRIGHT, J., KAZEM, A. J. N., THOMAS, D. K., HICKLING, R. & DAWNAY, N. (2007). Behavioural syndromes differ predictably between 12 populations of three-spined stickleback. *Journal of Animal Ecology* **76,** 1128-1138.

DOCHTERMANN, N. A. (2010). Behavioral syndromes: carryover effects, false discovery rates, and a priori hypotheses. *Behavioral Ecology* **21,** 437-439.

DUCKWORTH, A. L. & KERN, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality* **45,** 259-268.

FETTERMAN, D. M. (1986). Ethnography and Qualitative Design in Educational-Research - Goetz,Jp, Lecompte,Md. *American Anthropologist* **88,** 764-765.

FOX, R. A., LADAGE, L. D., ROTH, T. C. & PRAVOSUDOV, V. V. (2009). Behavioural profile predicts dominance status in mountain chickadees, *Poecile gambeli*. *Animal Behaviour* **77,** 1441-1448.

FROST, A. J., WINROW-GIFFEN, A., ASHLEY, P. J. & SNEDDON, L. U. (2007). Plasticity in animal personality traits: does prior experience alter the degree of boldness? *Proceedings of the Royal Society Biological Sciences Series B* **274,** 333-339.

GOLDBERG, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment* **4,** 26-42.

GOLDBERG, L. R. (1993). The Structure of Phenotypic Personality Traits. *American Psychologist* **48,** 26-34.

GORSUCH, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

GOSLING, S. D. (2001). From mice to men: What can we learn about personality from animal research? *Psychological Bulletin* **127,** 45-86.

GOSLING, S. D. & JOHN, O. P. (1999). Personality dimensions in nonhuman animals: A cross-species review. *Current Directions in Psychological Science* **8,** 69-75.

GOSLING, S. D., KWAN, V. S. Y. & JOHN, O. P. (2003). A dog's got personality: A cross-species comparative approach to personality judgments in dogs and humans. *Journal of Personality and Social Psychology* **85,** 1161-1169.

HALL, C. S. (1934). Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *Journal of Comparative Psychology* **18,** 385-403.

HODOS, W. & CAMPBELL, C. B. (1969). Scala Naturae - Why There Is No Theory in Comparative Psychology. *Psychological Review* **76,** 337-&.
JACOBY, J. (1978). Consumer Research - State of Art Review. *Journal of Marketing* **42,** 87-96.
JOHN, O. P., ROBINS, R. P. & PERVIN, L. A. (2008). *Handbook of Personality: Theory & Research*. The Guilford Press, New York, NY.
JONES, A. C. & GOSLING, S. D. (2005). Temperament and personality in dogs (Canis familiaris): A review and evaluation of past research. *Applied Animal Behaviour Science* **95,** 1-53.
KAVANAU, J. L. (1967). Behavior of captive white-footed mice. *Science* **155,** 1623-39.
KING, J. E., WEISS, A. & FARMER, K. H. (2005). A chimpanzee (*Pan troglodytes*) analogue of cross-national generalization of personality structure: Zoological parks and an African sanctuary. *Journal of Personality*, **73**, 389-410.
KONEČNÁ, M., LHOTA, S., WEISS, A., URBANEK, T., ADAMOVA, T. & PLUHACEK, J. (2008). Personality in Free-Ranging Hanuman Langur (*Semnopithecus entellus*) Males: Subjective Ratings and Recorded Behavior. *Journal of Comparative Psychology* **122,** 379-389.
KOSKI, S. E. (2011). How to Measure Animal Personality and Why Does It Matter? Integrating the Psychological and Biological Approaches to Animal Personality. In *From Genes to Animal Behaviour: Social structures, personalities, communication by color*, vol. 0. *Primatology Monographs* (ed. M. Inoue-Murayama, S. Kawamura and A. Weiss), pp. 115-136. Springer Japan.
KURVERS, R. H. J. M., EIJKELENKAMP, B., VAN OERS, K., VAN LITH, B., VAN WIEREN, S. E., YDENBERG, R. C. & PRINS, H. H. T. (2009). Personality differences explain leadership in barnacle geese. *Animal Behaviour* **78,** 447-453.
KURVERS, R. H. J. M., PRINS, H. H. T., VAN WIEREN, S. E., VAN OERS, K., NOLET, B. A. & YDENBERG, R. C. (2010). The effect of personality on social foraging: shy barnacle geese scrounge more. *Proceedings of the Royal Society B-Biological Sciences* **277,** 601-608.
LAY, C. (1973). Some notes on the concept of cross-situational consistency. In *Personality at the crossroads: Current issues in interactional psychology* (ed. D. Magnusson and N. S. Endler), pp. 143-146. John Wiley & Sons, New York.
LESTER, D. (1968). Curiosity and the role of the role of the individual in the imposition of change. *Psychol Rep* **22,** 109-12.
LOCKARD, R. B. (1971). Reflections on the fall of comparative psychology: is there a message for us all? *American Psychologist* **26,** 168-179.
LOEHLIN, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3rd ed.). Lawrence Erlbaum Associates, Mahwah, NJ.
LOPEZ, P., HAWLENA, D., POLO, V., AMO, L. & MARTIN, J. (2005). Sources of individual shy-bold variations in antipredator behaviour of male Iberian rock lizards. *Animal Behaviour* **69,** 1-9.
MATHER, J. A. & ANDERSON, R. C. (1993). Personalities of Octopuses (*Octopus rubescens*). *Journal of Comparative Psychology* **107,** 336-340.
MCCRAE, R. R. (1982). Consensual Validation of Personality Traits - Evidence from Self-Reports and Ratings. *Journal of Personality and Social Psychology* **43,** 293-303.
METTKE-HOFMANN, C., EBERT, C., SCHMIDT, T., STEIGER, S. & STIEB, S. (2005). Personality traits in resident and migratory warbler species. *Behaviour* **142,** 1357-1375.
MILLER, K. A., GARNER, J. P. & MENCH, J. A. (2005). The test-retest reliability of four behavioural tests of fearfulness for quail: a critical evaluation. *Applied Animal Behaviour Science* **92,** 113-127.
MISCHEL, W. (1968). *Personality and assessment*. John Wiley and Sons, New York, NY.
MISCHEL, W. (1973a). The interaction of person and situation. In *Personality at the crossroads: Current issues in interactional psychology* (ed. D. Magnusson and N. S. Endler), pp. 333-352. John Wiley & Sons, New York.

MISCHEL, W. (1973b). Toward a Cognitive Social Learning Reconceptualization of Personality. *Psychological Review* **80,** 252-283.
MISCHEL, W. (1999). Implications of person-situation interaction: Getting over the field's borderline personality disorder. *European Journal of Personality* **13,** 455-461.
MISCHEL, W., SHODA, Y. & MENDOZA-DENTON, R. (2002). Situation-behavior profiles as a locus of consistency in personality. *Current Directions in Psychological Science* **11,** 50-54.
MISSLIN, R. & CIGRANG, M. (1986). Does Neophobia Necessarily Imply Fear or Anxiety. *Behavioural Processes* **12,** 45-50.
NETTLE, D. & PENKE, L. (2010). Personality: bridging the literatures from human psychology and behavioural ecology. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365,** 4043-4050.
NEVO, B. (1985). Face Validity Revisited. *Journal of Educational Measurement* **22,** 287-293.
NUNNALLY, J. C. (1978). *Psychometric Theory*. McGraw-Hill Book Company, New York.
NUSSEY, D. H., WILSON, A. J. & BROMMER, J. E. (2007). The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology* **20,** 831-844.
PEDERSON, A. K., KING, J. E. & LANDAU, V. I. (2005). Chimpanzee (*Pan troglodytes*) personality predicts behavior. *Journal of Research in Personality* **39,** 534-549.
PETER, J. P. (1979). Reliability - Review of Psychometric Basics and Recent Marketing Practices. *Journal of Marketing Research* **16,** 6-17.
PRUITT, J. N., RIECHERT, S. E. & JONES, T. C. (2008). Behavioural syndromes and their fitness consequences in a socially polymorphic spider, Anelosimus studiosus. *Animal Behaviour* **76,** 871-879.
RÉALE, D., DINGEMANSE, N. J., KAZEM, A. J. N. & WRIGHT, J. (2010). Evolutionary and ecological approaches to the study of personality. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365,** 3937-3946.
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.
SHORT, K. H. & PETREN, K. (2008). Boldness underlies foraging success of invasive Lepidodactylus lugubris geckos in the human landscape. *Animal Behaviour* **76,** 429-437.
SIH, A., BELL, A. & JOHNSON, J. C. (2004a). Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution* **19,** 372-378.
SIH, A., BELL, A. M., JOHNSON, J. C. & ZIEMBA, R. E. (2004b). Behavioral syndromes: An integrative overview. *Quarterly Review of Biology* **79,** 241-277.
SIH, A., KATS, L. B. & MAURER, E. F. (2003). Behavioural correlations across situations and the evolution of antipredator behaviour in a sunfish-salamander system. *Animal Behaviour* **65,** 29-44.
SIH, A. & WATTERS, J. V. (2005). The mix matters: behavioural types and group dynamics in water striders. *Behaviour* **142,** 1417-1431.
STAMPS, J. & GROOTHUIS, T. G. G. (2010). The development of animal personality: relevance, concepts and perspectives. *Biological Reviews* **85,** 301-325.
TETT, R. P. & GUTERMAN, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality* **34,** 397-423.
TOMS, C. N., ECHEVARRIA, D. J. & JOUANDOT, D. J. (2010). A methodological review of personality-related studies in fish: Focus on the shy-bold axis of behaviour. *International Journal of Comparative Psychology* **23,** 1-25.
UHER, J. (2008). Comparative personality research: Methodological approaches. *European Journal of Personality* **22,** 427-455.

UHER, J. (2011). Individual behavioral phenotypes: An integrative meta-theoretical framework. Why "behavioral syndromes" are not analogs of "personality". *Developmental Psychobiology* **53,** 521-548.

UHER, J. & ASENDORPF, J. B. (2008). Personality assessment in the Great Apes: Comparing ecologically valid behavior measures, behavior ratings, and adjective ratings. *Journal of Research in Personality* **42,** 821-838.

UHER, J. Personality in nonhuman primates. What can we learn from human personality psychology? *Personality and temperament in nonhuman primates* (eds. A. Weiss, J. E. King, & L. Murray), pp. 41-76. Springer, New York, NY.

Personality and Temperament in Nonhuman Primates. *Developments in Primatology: Progress and Prospects* (ed. L. Barrett), pp. 41-76. Springer New York.

VAN DE POL, M. V. & WRIGHT, J. (2009). A simple method for distinguishing within- versus between-subject effects using mixed models. *Animal Behaviour* **77,** 753-758.

VAZIRE, S., GOSLING, S. D., DICKEY, A. S. & SCHAPIRO, S. J. (2007). Measuring personality in nonhuman animals. In *Handbook of Research Methods in Personality Psychology* (ed. R. W. Robins, R. C. Fraley and R. F. Krueger), pp. 190-206. Guilford, New York.

WALSH, R. N. & CUMMINS, R. A. (1976). The Open-Field Test - A Critical Review. *Psychological Bulletin* **83,** 482-504.

WEBSTER, M. M., WARD, A. J. W. & HART, P. J. B. (2009). Individual boldness affects interspecific interactions in sticklebacks. *Behavioral Ecology and Sociobiology* **63,** 511-520.

WEISS, A. & ADAMS, M. J. (In press). Differential behavioral ecology. In *Animal personalities: Behavior, physiology and evolution*, (ed. C. Carere and D. Maestripieri). University of Chicago Press, Chicago, IL.

WEISS, A., ADAMS, M. J., WIDDIG, A. & GERALD, M. S. (2011). Rhesus Macaques (*Macaca mulatta*) as Living Fossils of Hominoid Personality and Subjective Well-Being. *Journal of Comparative Psychology* **125,** 72-83.

WHITTINGHAM, M. J., BUTLER, S. J., QUINN, J. L. & CRESSWELL, W. (2004). The effect of limited visibility on vigilance behaviour and speed of predator detection: implications for the conservation of granivorous passerines. *Oikos* **106,** 377-385.

WINTER, D. G., JOHN, O. P., STEWART, A. J., KLOHNEN, E. C. & DUNCAN, L. E. (1998). Traits and motives: Toward an integration of two traditions in personality research. *Psychological Review* **105,** 230-250.

WOLF, M. & WEISSING, F. J. (2010). An explanatory framework for adaptive personality differences. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365,** 3959-3968.

**Table 1**
Glossary of terms.

**Table 2**

Definitions of animal personality tests mentioned throughout and a non-exhaustive list what they have been used to measure.

**Table 3**

Alternative interpretations of a correlation/lack of correlation between two tests/tested traits given a behaviour syndrome identification interpretation or a personality trait-validation interpretation. Numbering refers to the main text.

**Table 4**

When assessing the validity of a test, the validity should be checked using multiple other tests (multi-trait, multi-test approach; see text). 'Convergent' represents a correlation between the traits, 'discriminant' represents no correlation. The bottom diagonal demonstrates the predicted correlations between tests if the test (e.g. novel object test) measured, in this example, exploration. The top diagonal represents the predicted correlations between tests if the test (e.g. novel object test) measured, in this example, boldness.

781 Table 1

| Term | Definition |
|---|---|
| A behaviour | A mutually exclusive aspect of a behavioural repertoire that can be quantified in a test or as part of an activity budget |
| Behavioural syndromes | Correlations between two or more personality traits through time or across contexts. |
| Factor | A group of frequently correlating personality traits that ideally is orthogonal to other traits. For example, boldness and aggression could usually be included in one factor. Factors are normally identified through interpretation of output from data-reduction techniques, e.g. principal component analysis, PCA. |
| Label | A term used to describe a personality trait |
| Measurement | A value that is taken from a test that is used to quantify an aspect of a personality trait (such as latency to enter the open-field) |
| Overlap | The concept that one test can simultaneously measure two or more personality traits |
| Personality | Between-individual differences in behaviour that persist through time. |
| (Personality) trait | A specific aspect of a behavioural repertoire that shows between-individual variation and within-individual consistency (such as boldness, aggression, activity). |
| Validity | The degree to which a test measures the targeted trait |

Table 2

| Test | Execution | Used to measure |
|---|---|---|
| Open field | An individual is introduced into an arena, usually novel. Can be 'forced' or 'free' if the individual is given no choice to enter or allowed to enter the arena at will, respectively | Emotionality, fear, gregariousness and exploration (Walsh & Cummins, 1976), |
| Novel object | An individual is introduced to a novel object | Boldness (Frost et al., 2007; Kurvers et al., 2009), exploration (Réale et al., 2007) |
| Novel environment | A modified version of the open field test; an individual is introduced into an unfamiliar environment, but the environment may include novel stimuli, or familiar stimuli arranged in a novel manner | Exploration-avoidance (Dingemanse et al., 2007), activity (Butler et al., 2011) |
| Emergence test | A modified version of the open field test if the individual is emerging into a novel environment; individuals are either introduced to a novel environment/open field (arguably a measurement from an open field test) or are startled in their home cages and their latencies to emerge from a shelter or resume normal behaviour are recorded | Boldness (Brown & Braithwaite, 2004; Lopez et al., 2005), fearfulness (Miller, Garner & Mench, 2005) |

Table 3

| **Result:** | **Personality trait-validation interpretation** | **Behavioural syndrome-identification interpretation** |
|---|---|---|
| Correlation found between tests or measurements load on the same factor | 1. The test shows **convergent validity**: the test/s allow(s) the measurement of the same personality trait | 2. The behaviour is considered **context general**: the same behaviour is expressed in both situations (suggesting the existence of a syndrome) |
| No correlation found between tests or measurements load on different factors | 3. The test shows **discriminant validity**: the test/s measure/s different personality traits | 4. The behaviour is considered **context specific**: the expression of the behaviour depends on the context (suggesting no syndrome exists) |

787 Table 4

| | | | | |
|---|---|---|---|---|
| | Test | Novel Object | Open Field | Aggression towards conspecific |
| Exploration | Novel object | - | Discriminant | Convergent |
| | Open field | Convergent | - | Discriminant |
| | Aggression towards conspecific | Discriminant | Discriminant | - |

Boldness

788

789