



Animal personality: what are behavioural ecologists measuring?

Journal:	<i>Biological Reviews</i>
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 Feeny, William; Australian National University, Marshall, Harry; Institute of Zoology, Cowlshaw, Guy; Institute of Zoology, Zoological Society of London Heinsohn, Robert; Australian National University, Fenner School of Environment and Society
Keywords:	behavioural syndromes, boldness, exploration, methodology, animal personality, risk-taking

SCHOLARONE™
Manuscripts

Only

1
2
3
4 1 **Animal personality: what are behavioural ecologists**
5
6
7 2 **measuring?**
8
9
10 3

11
12
13 4 Alecia J. Carter^{1,2,3}, William E. Feeney⁴, Harry H. Marshall^{2,5}, Guy Cowlshaw² & Robert
14
15 5 Heinsohn¹
16
17
18
19 6

20
21 7
22
23
24 8 ¹The Fenner School of Environment and Society, The Australian National University, Acton,
25
26 9 Canberra, ACT 0200 Australia

27
28
29 10 ² The Institute of Zoology, Zoological Society of London, Regent's Park, NW1 4RY, London, U.K.

30
31
32 11 ³ Current address: Large Animal Research Group, Department of Zoology, The University of
33
34 12 Cambridge, Cambridge, UK, CB2 3EJ E

35
36
37 13 ⁴ Evolution, Ecology and Genetics, The Research School of Biology, The Australian National
38
39 14 University, Acton, Canberra, ACT 0200 Australia

40
41
42 15 ⁵ Division of Ecology and Evolution, Department of Life Sciences, Imperial College London, Silwood
43
44 16 Park, Ascot, Berkshire, SL5 7PY, U.K.

45
46
47 17 Corresponding author: Alecia J. Carter: Large Animal Research Group, Department of Zoology, The
48
49 18 University of Cambridge, Cambridge, UK, CB2 3EJ E: ac854@cam.ac.uk T: +44 1223 769277
50
51
52 19

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

20 **Animal personality: what are behavioural ecologists measuring?**

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

33
34 Keywords: animal personality, behavioural syndromes, boldness, exploration, methods, risk-taking

35

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

36	CONTENTS
37	I. Introduction
38	II. Problems measuring personality traits
39	1. Many tests for one trait
40	2. One test for many traits
41	III. Lessons from other fields
42	1. An important caveat
43	2. Test validity
44	a. Reliability
45	b. Convergent and discriminant validities
46	c. Other validities
47	3. Jingle-jangle fallacies and trait definitions
48	4. Interpreting tests
49	5. Measurement considerations
50	IV. How to measure personality traits
51	V. Conclusions
52	

1
2
3 53 I. INTRODUCTION
4
5

6 54 Behavioural ecology is the study of the ecological and evolutionary bases for animal behaviour and
7
8 55 emphasises the individual as the fundamental unit of analysis. Each sub-discipline in behavioural
9
10 56 ecology uses rigorous evolutionary reasoning to explain how animals use behaviour to deal with their
11
12 57 intrinsic and extrinsic environments. Such thinking has made behavioural ecology a fast-paced field
13
14 58 (Caro & Sherman, 2011). However, the use of optimization theory in this thinking has led to the
15
16 59 assumption that variation in animal behaviour is either centred on a single adaptive optimum, or on
17
18 60 two or more co-existing evolutionarily stable strategies (Weiss & Adams, In press). Variation around
19
20 61 these optima has been traditionally viewed as noise (Mather & Anderson, 1993). Recent research in
21
22 62 behavioural ecology challenges this view and proposes a ‘new’ concept that can explain variation in
23
24 63 behaviour including sub-optimal tendencies: animal personality (Réale et al., 2007). Animal
25
26 64 personality refers to between-individual differences in behaviour that persist through time (Biro &
27
28 65 Stamps, 2008; alternatively, behavioural type: Sih et al., 2004b). A related concept, behavioural
29
30 66 syndromes, goes one step further, referring to individual-level differences in correlations between
31
32 67 personality traits or behaviours (table 1; see also Stamps & Groothuis, 2010). Here we will refer to
33
34 68 animal personality in its broadest sense (as ‘behavioural variation between individuals’),
35
36 69 encompassing the concept of behavioural syndromes throughout. Further, we use the term personality
37
38 70 ‘trait’ in the behavioural ecological sense as a particular aspect of an individual’s behavioural
39
40 71 repertoire, such as aggression or boldness (for a complete glossary of terms used in this article, see
41
42 72 table 1).
43
44
45
46
47

73

48 74 Several theories have been proposed to explain the apparently suboptimal behavioural tendencies
49
50 75 associated with animal personalities (Dingemanse & Wolf, 2010; Wolf & Weissing, 2010). For
51
52 76 example, theories based on cost-benefit trade-offs predict that a bolder individual may receive a
53
54 77 benefit by outcompeting conspecifics to gain greater access to resources (Pruitt, Riechert & Jones,
55
56 78 2008; Short & Petren, 2008), but bolder animals may also take more risks making them more
57
58
59
60

1
2
3 79 susceptible to predation (Bremner-Harrison, Prodohl & Elwood, 2004; Carter, Goldizen & Tromp,
4
5 80 2010). Animal personalities are associated with differences in fitness and are partly heritable, which
6
7 81 raises important questions about how variation in personality is maintained in natural populations
8
9 82 (reviewed in Bell, 2007). Despite this, there is still relatively little known about the causes and
10
11 83 consequences of animal personality, from either proximate or ultimate perspectives.
12
13

14 84

15
16
17 85 Behavioural ecologists are currently asking three broad questions regarding animal personality: 1)
18
19 86 Why do individuals vary consistently in their behaviour, in some contexts to the point of sub-
20
21 87 optimality? 2) If selection 'pushes' behaviour toward one or more optimal strategies, how is further
22
23 88 variation in animal behaviour maintained? 3) Why do cross-context behavioural correlations
24
25 89 (behavioural syndromes) occur? Empiricists interested in any of these broad theoretical questions
26
27 90 must first measure behaviours to establish whether they exhibit between-individual differences.
28
29 91 Despite a recent surge in popularity within behavioural ecology (Réale et al., 2010), the study of
30
31 92 animal personalities is fraught with conceptual and methodological difficulties inherent in any young
32
33 93 field. Perhaps most telling is the confusion about how personality traits are defined and measured (see
34
35 94 below). This confusion can lead to mislabelling traits and misinterpreting results, putting the
36
37 95 development of animal personality theory at risk. This in turn raises two fundamental questions: Are
38
39 96 researchers who are attempting to test the same personality traits in different taxa actually measuring
40
41 97 the same thing? And if not, does this have implications for comparing the results of animal personality
42
43 98 studies? The methodological and conceptual problems facing the field of animal personality theory
44
45 99 are by no means novel. In fact, many concerns within the behavioural ecology approach to animal
46
47 100 personality have been raised in other fields previously (Weiss & Adams, In press). However, no
48
49 101 proposed methodological or conceptual frameworks offer solutions to the myriad issues identified.
50
51 102 Here, we will review the current issues raised regarding the definitions and methods used to measure
52
53 103 animal personality across taxa and systems, and demonstrate that the tools to overcome these
54
55 104 problems have been developed previously in the psychological literatures. Our aim is to highlight
56
57 105 teething problems in the field, and by heeding the lessons learned elsewhere, encourage a unified
58
59
60

1
2
3 106 approach to future animal personality studies through the use of research tools that have been
4
5 107 successfully used elsewhere.
6
7
8 108

9
10
11 109 II. PROBLEMS MEASURING PERSONALITY TRAITS
12

13
14 110 **(1) Many tests for one trait**
15

16
17 111 Boldness is one of the most commonly measured personality traits (Conrad et al., 2011), but is
18
19 112 perhaps the trait with least consensus over its definition. For instance, boldness has been interpreted as
20
21 113 being the propensity to take risks, especially in novel situations (Coleman & Wilson, 1998; Toms,
22
23 114 Echevarria & Jouandot, 2010), whereas Réale et al. (2007) defined boldness as an individual's
24
25 115 response to a risky situation alone, and excluded reactions to novel situations and stimuli altogether.
26

27 116

28
29
30 117 Tests of boldness reflect the confusion inherent in its definition. Boldness has been tested by
31
32 118 quantifying behavioural responses to novel objects, responses to a novel environment, and responses
33
34 119 to predation risk (Toms et al., 2010). However, these three types of test are not necessarily
35
36 120 comparable and demonstrate a lack of standardised tests for quantifying the behaviour (Budaev &
37
38 121 Brown, 2011; Conrad et al., 2011; Toms et al., 2010; Carter et al., In press). Fox et al. (2009), for
39
40 122 example, measured both novel object exploration and exploration of a novel environment in their
41
42 123 study of mountain chickadees (*Poecile gambeli*). The authors did not find a correlation between the
43
44 124 traits and highlighted that they were not interchangeable, as suggested by other studies (see references
45
46 125 in Fox et al., 2009). A similar result was found in a study of pumpkinseed sunfish (*Lepomis*
47
48 126 *gibbosus*); reaction to a threatening novel stimulus did not correlate with response to a novel food
49
50 127 (Coleman & Wilson, 1998). Both studies suggested that the behaviour tested was specific to the
51
52 128 context that it was tested in. Dingemanse et al. (2007) encountered a similar problem when attempting
53
54 129 to test shyness-boldness in sticklebacks (*Gasterosteus aculeatus*) using antipredator behaviour
55
56 130 (response to risk). They compared the exploratory behaviour of sticklebacks when exposed to a
57
58
59
60

1
2
3 131 predator housed in an adjacent compartment and to their behaviour when exposed to an empty
4
5 132 compartment. The authors found no significant differences and relabelled this behaviour as a measure
6
7 133 of “exploration-avoidance of an altered environment”. This study not only highlights the importance
8
9 134 of using controls, but also that incorrectly labelling personality traits may be a common problem. We
10
11 135 know of only one study that has directly investigated the relationship between multiple tests of
12
13 136 boldness in the same individuals with the intent of describing the validity of the tests. Burns (2008;
14
15 137 but see also Bergvall et al., 2011, Carter et al., In press) measured the responses of individual guppies
16
17 138 (*Poecilia reticulata*) to three experiments intended to measure boldness: an open-field test, an
18
19 139 emergence test and a novel-object test (table 2). The open-field and emergence tests correlated with
20
21 140 each other, but neither of these tests correlated with the results of the novel-object test, leading Burns
22
23 141 (2008) to conclude that the novel object test should not be used to assess boldness in guppies.
24
25

142

26
27
28
29 143 A further important consideration is that a test for one species/taxon is not necessarily appropriate as a
30
31 144 test for another (Weiss & Adams, In press). For example, while some animals may perceive a greater
32
33 145 risk of predation in an open as opposed to closed habitat (Blumstein & Daniel, 2003), the same may
34
35 146 not be true for species that are predated primarily in closed habitats (Whittingham et al., 2004).
36
37 147 Although comparability between studies is desirable, if a test is to be adapted to a new system, every
38
39 148 effort should be made to make the test as ecologically relevant as possible.
40
41

149

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

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

1
2
3 157 there are fundamental problems with its interpretation in different circumstances. An individual can
4
5 158 be introduced to the open-field by force (by placing it in the environment with no opportunity to
6
7 159 escape), or be offered the flexibility to explore the open-field freely (with access to a refuge, by
8
9 160 having its home cage placed in or connecting to the environment) (Crusio, 2001; Walsh & Cummins,
10
11 161 1976). However, behaviour in forced *versus* free exploration contexts may not necessarily correspond
12
13 162 to the same traits (Kavanau, 1967; Lester, 1968; Misslin & Cigrang, 1986). Free open-field tests are
14
15 163 more likely to measure voluntary exploration/curiosity and information-gathering behaviour, while
16
17 164 forced open-field tests are more likely to measure fear or anxiety (or both) (Misslin & Cigrang, 1986).
18
19 165 In a review of the use of open field tests, Walsh and Cummins (1976) highlighted that the test has
20
21 166 been interpreted as measuring emotionality, fear, gregariousness (if more than one individual was
22
23 167 used) and exploration. Taken together, these studies indicate that care needs to be taken when using
24
25 168 the open-field test, and other tests where protocol differences exist, to test personality traits, especially
26
27 169 when comparing across multiple studies.
28
29

30 170

31
32
33 171 One test can simultaneously be influenced by and thus measure two or more personality traits (Réale
34
35 172 et al., 2007). This ‘overlap’ can become especially clear when multiple measurements are taken and
36
37 173 data reduction techniques, such as factor or principal components analysis, are used (Gorsuch, 1983).
38
39 174 If the test measured only one trait, then only one factor/component explaining substantial variation in
40
41 175 the measurements would be identified. However, it is more often the case that two or more of the
42
43 176 resultant factors/components have this explanatory power, such that two or three unrelated traits will
44
45 177 emerge (assuming that each factor/component represents a trait) (for example, see Carter et al., 2011).
46
47 178 Although a test that directly measures a targeted trait may be a desirable goal, in reality a test will
48
49 179 likely be influenced by multiple traits at the same time (Réale et al., 2007). We deal with this issue
50
51 180 further in the next section.
52

53
54 181

55
56
57 182 III. LESSONS FROM OTHER FIELDS
58
59
60

1
2
3 183 If the multiple 'standard' tests for boldness are not comparable and one test is able to measure two
4
5 184 traits concurrently, which tests should animal personality researchers use and how should their results
6
7 185 be interpreted? Fortunately these types of problems have been faced before by other fields during their
8
9 186 infancy, affording excellent opportunities to adopt their solutions and avoid the likely stagnation
10
11 187 associated with inertia in updating methods and concepts (Lockard, 1971; see also Beach, 1950;
12
13 188 Boice, 1971; Hodos & Campbell, 1969). Here we consider the use of psychometrics, the theory of
14
15 189 psychological measurement, which has helped personality research in the fields of comparative
16
17 190 psychology, the psychological study of animal behaviour, and differential psychology, the
18
19 191 investigation of individual differences in behaviour (John, Robins & Pervin, 2008). We first outline an
20
21 192 important caveat when considering psychometric applications to the study of animal personality
22
23 193 within behavioural ecology. We then review important tools and theories that could be applied to, and
24
25 194 should not be overlooked, within behavioural ecology.
26
27
28
29
30

195

31 196 **(1) An important caveat**

32
33
34 197 Three important points regarding the differences between the psychological and behavioural
35
36 198 ecological literatures that may impede communication between the two fields are as follows. First,
37
38 199 there are important differences in terminology (Koski, 2011; Uher, 2011). Uher (2011) outlines these
39
40 200 differences in detail; one clear example is the term 'trait'. The term is used in behavioural ecology
41
42 201 (and by us in this article) to mean a measured aspect of an individual's behaviour (e.g. the rate a
43
44 202 behaviour is observed) while in psychology the use is more abstract and describes a construct (see
45
46 203 below). Further confusion is added because these terminologies are often debated within each field
47
48 204 (for example, see Lay, 1973). Second, psychologists have remarked that behavioural ecology has a
49
50 205 narrow, restricted and incomprehensive view of personality variation (Uher, 2011; Weiss & Adams,
51
52 206 In press). However, this view relates to differences in approach and research goals: behavioural
53
54 207 ecologists take a reductionist approach to animal personality whereas psychologists take a more
55
56 208 holistic approach (for a longer discussion of these differences, see Koski, 2011; Uher, 2011; Weiss &
57
58
59
60

1
2
3 209 Adams, In press). Finally, there is much historical (and perhaps contemporary) conceptual debate
4
5 210 within the psychometric literature (see definitions in Campbell & Fiske, 1959; Cronbach & Meehl,
6
7 211 1955; and discussion in Uher, 2011). While these differences may seem daunting, the psychological
8
9 212 and psychometric literatures nonetheless provide an established discussion of measurement problems
10
11 213 currently systemic in animal personality research in behavioural ecology.
12
13

14 214

15 215 **(2) Test validity**

16
17
18
19 216 Validity refers to the degree to which a test measures the targeted trait (Burns, 2008; see also Réale et
20
21 217 al., 2007). A frequently recurring critique of personality psychology involves the validity of the tests
22
23 218 that are used to measure the trait of interest (Duckworth & Kern, 2011). One of the first ways to
24
25 219 remedy this problem is to use multiple measurements for multiple traits, and investigate correlations
26
27 220 between the measurements (Campbell & Fiske, 1959). However, it is important to be aware that
28
29 221 correlations between two measurements could be influenced by shared method variance (Campbell &
30
31 222 Fiske, 1959). That is, systematic variation in the type of method used to obtain the measurements
32
33 223 could cause tests using similar methods to be more correlated than tests using different methods. This
34
35 224 is a particular problem for behavioural ecologists, as most of the methods used are similar
36
37 225 (behavioural observations). Method variance can be explicitly estimated by using hierarchically
38
39 226 nested models, however, when more than one method is used (Widaman, 1985). Three other key
40
41 227 aspects to trait measurement are particularly relevant to behavioural ecologists: Reliability, and
42
43 228 Convergent and Discriminant validities (table 3) (Burns, 2008).
44
45

46 229

47 48 49 230 *(a) Reliability*

50
51
52 231 Reliability refers to the consistency of a measure through time, across contexts or across
53
54 232 raters/observers. It estimates whether there is agreement between repeated tests of the same nature.
55
56 233 Reliability differs from validity because a test may be reliable but not be valid. A common analogy
57
58
59
60

1
2
3 234 used to demonstrate this difference (Nunnally 1978) is an archer (the researcher) trying to hit a target
4
5 235 (the trait) with a bow and arrow (the test). Validity refers to how close to the centre of the target the
6
7 236 arrows land while reliability refers to how close together the arrows are clustered. Reliability is the
8
9 237 first psychometric test employed in psychological research to assess a test's performance (Gosling,
10
11 238 2001). In animal personality studies this approach is common practice, and is known as consistency
12
13 239 or, more commonly, repeatability (Bell, Hankison & Laskowski, 2009). In fact, behaviours must be
14
15 240 repeatable to be considered personality traits within behavioural ecology (Réale et al., 2007; Sih, Bell
16
17 241 & Johnson, 2004a).

18
19
20 242

21
22
23 243 *(b) Convergent and discriminant validity*

24
25
26 244 Convergent and discriminant validities rest at either end of a spectrum for validating trait tests
27
28 245 (Campbell & Fiske, 1959). Convergent validity investigates whether two tests actually measure the
29
30 246 same trait (that is, the measurements from both tests should correlate). Conversely, discriminant
31
32 247 validity investigates whether two tests that are employed to measure different traits actually measure
33
34 248 different traits (that is, the measurements from the tests should not correlate). In psychology, each
35
36 249 measurement from a test is referred to as a trait-method unit and researchers are encouraged to use a
37
38 250 multi-trait, multi-method approach to describe variation in personality (Campbell & Fiske, 1959).
39
40 251 While it is possible in psychology to use multiple methods (such as behavioural scoring and self-
41
42 252 assessment), behavioural ecologists are frequently restricted to one method (behavioural
43
44 253 observations), but can use multiple tests. Studies that predict but fail to find a correlation between two
45
46 254 tests usually conclude that the trait is context specific. An example would be Fox *et al.*'s (2009) study
47
48 255 that failed to find a correlation between response to a novel object and response to a novel
49
50 256 environment. However, we could alternatively conclude that the two tests measures different
51
52 257 personality traits (Burns, 2008; Carter et al. In press), and so are highly discriminant (Campbell &
53
54 258 Fiske, 1959; for alternative interpretations, see Cronbach & Meehl, 1955). Without the use of
55
56 259 multiple tests for a given trait it is impossible to measure how convergent or discriminant the tests are.
57
58
59
60

1
2
3 260 This can result in, at the very least, confusion over what trait is actually being measured and, in the
4
5 261 worst case, measurement of a trait that is interpreted incorrectly. Lessons from psychometrics and
6
7 262 psychology (Duckworth & Kern, 2011; McCrae, 1982) therefore emphasise the importance of
8
9 263 attempting to measure animal personality traits using multiple methods, or in the case of behavioural
10
11 264 ecology, multiple tests (Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Uher, 2011; Weiss &
12
13 265 Adams, In press). In behavioural ecology this is rarely the case (but see Bergvall et al., 2011; Burns,
14
15 266 2008; Carter et al., In press). The use of a multi-trait, multi-test approach (Campbell & Fiske, 1959)
16
17 267 would allow assessment of a test's appropriateness and aid in further identification of personality
18
19 268 traits and behavioural syndromes.
20
21

22 269

23
24
25 270 *(c) Other validities*
26

27
28 271 Two other validation techniques are commonly discussed in psychological research: face validity and
29
30 272 construct validity (Nunnally, 1978). In the first case, a test that appears to measure the trait of interest
31
32 273 is said to have face validity (i.e. the correlation between a trait's theoretical definition and the
33
34 274 subject's response to the test). Although face validity is fundamental for a test to be used, differential
35
36 275 psychologists have noted that it is frequently assumed and infrequently confirmed (Nevo, 1985). For
37
38 276 example, for open-field studies, defecation was used as a measure of emotionality (response to
39
40 277 fear/anxiety) as it had long been known that defecation/urination occurs during periods of emotional
41
42 278 stress (that is, it has face validity; Hall, 1934). After some debate about the use of this test (Archer,
43
44 279 1973), it was later validated in a study that measured both defecation and physiological responses to
45
46 280 the open-field test (Denenberg, 1973). It is apparent from psychology that face validity can be
47
48 281 confounded with construct validity (see below) leading to the possible failure to identify reliable tests.
49

50
51 282

52
53
54 283 In the second case, construct validity, like face validity, is related to a theoretical understanding of
55
56 284 personality, and investigates the theoretically predicted relationships between a variety of
57
58 285 psychological traits and behaviours. That is, a construct is an *a priori*, theoretical idea of how
59
60

1
2
3 286 behaviours should be inter-correlated (Cronbach & Meehl, 1955; Nunnally, 1978). While behavioural
4
5 287 syndromes may similarly investigate correlations between behaviours, construct validity raises an
6
7 288 important conceptual point that is frequently overlooked in behavioural ecology (from Nunnally, 1978
8
9 289 p 104): while two behaviours may correlate, this is not proof that either behaviour measures the
10
11 290 targeted trait/factor (see Section 3 below).

12
13
14 291

15 16 17 292 **(3) Jingle-jangle fallacies and trait definitions**

18
19
20 293 Confusion about over- or under-labelling traits is known in differential psychology as the jingle-jangle
21
22 294 fallacy (Bell, 2007; Gosling, 2001; Uher, 2011). The 'jingle' fallacy refers to a single trait label that
23
24 295 inadvertently describes two functionally different traits measured with different tests. The 'jangle'
25
26 296 fallacy refers to two labels that actually measure the same trait (Block, 1995). Jingle-jangle fallacies
27
28 297 can be pervasive and misleading, and there is a need to identify poor tests as early as possible (Block,
29
30 298 1995; Jacoby, 1978). For example, exploration of a novel object and exploration of a novel
31
32 299 environment/open field may be given the same label: exploration. However, as noted above, if
33
34 300 exploration of an open field is forced, the test may measure anxiety and not exploration, and we
35
36 301 would have committed a jingle fallacy. In such circumstances, to avoid a jingle fallacy, each test
37
38 302 might be given a different trait name. But this may generate another problem, namely confusion
39
40 303 within animal personality studies and the risk that the two trait names could be seen as synonymous of
41
42 304 the same trait: a jangle fallacy.

43
44
45 305

46
47
48 306 Winter et al. (1998) recognised the jingle fallacy among personality psychologists who were
49
50 307 concurrently, but separately, studying psychological traits and motives (two different concepts within
51
52 308 this field). While testing the convergent and discriminant validities of tests is essential, progress in
53
54 309 answering proximate and ultimate questions about personality could potentially be gained more
55
56 310 efficiently by using key validated tests that assess demonstrably independent traits. Winter et al.
57
58 311 (1998) suggested simplifying the field of personality research after asking whether all of the tests

1
2
3 312 were required to understand personality. Reduction of the required number of tests could be achieved
4
5 313 by using a phylogenetically controlled meta-analytical approach to identify those traits, and those tests
6
7 314 used to measure them, which have been consistently validated in the literature. These traits could then
8
9 315 form the basis of a general framework for the study of animal personality (Duckworth & Kern, 2011).

10
11
12 316

13
14
15 317 **(4) Interpreting tests**

16
17 318 Next we raise some issues regarding the interpretation of the results of personality tests. Assuming
18
19 319 that reliability has been established, and method variance controlled for, there are two possible
20
21 320 interpretations of the relationships between the results of tests of convergent and discriminant validity.
22
23 321 These interpretations depend on the focus of the study: whether the tests or the traits being measured
24
25 322 are of interest. In the first case, test validation for a single personality trait is the study's goal, and
26
27 323 interpretation focuses on whether the two different tests are measuring the same or different traits (the
28
29 324 'personality trait-validation' interpretation, see table 3). In the second case, which is more typical in
30
31 325 behavioural ecology, the two different tests are already assumed to measure different personality traits
32
33 326 and the goal is to establish a behavioural syndrome, such that interpretation focuses on the
34
35 327 relationship between the two tested behaviours (the 'behavioural-syndrome identification'
36
37 328 interpretation, table 3). Thus a positive correlation between two tested behaviours can come about
38
39 329 because 1) the tests measure the same personality trait (the tests have convergent validity) or 2) the
40
41 330 two behaviours are linked by an underlying behavioural syndrome (Burns, 2008). Likewise, a lack of
42
43 331 a correlation between two tested behaviours can come about because 3) the tests measure different
44
45 332 personality traits (the tests have discriminant validity) or 4) the two behaviours are not linked in a
46
47 333 behavioural syndrome (Sih et al., 2004a) (table 3).

48
49
50
51 334

52
53
54 335 If we wished to study exploration as a personality trait, for example, we could perform two tests,
55
56 336 exploration of a novel object and exploration of a novel environment/open field, to search for
57
58 337 convergent validity. Under the trait focussed interpretation, if these tests did correlate, we could say
59
60

1
2
3 338 that both tests measured exploration, but to be sure, we should perform one more test, such as
4
5 339 aggression towards a conspecific that should not correlate with either exploration of a novel
6
7 340 environment or of a novel object. Alternatively, if we found that the novel object and novel
8
9 341 environment tests did not correlate, but that the novel object and aggression tests did, we could
10
11 342 conclude that novel object and aggression measure the same trait, for example, boldness, with
12
13 343 convergent validity and discriminant validity from exploration (novel environment) (table 4).
14
15 344 Depending on the goal of the study (test validation or identification of a behavioural syndrome) two
16
17 345 different interpretations of the results are possible, and care should be taken when investigating
18
19 346 correlations between multiple tests.
20
21
22 347
23
24
25 348 One possible solution to the problem posed by correlations between test measurements might be
26
27 349 found via trait orthogonality (mathematical independence). This concept, primarily employed by
28
29 350 differential and human personality researchers, uses factor analysis or principal component analysis
30
31 351 (PCA) to identify the orthogonal personality factors that are robust across investigations, samples and
32
33 352 time (Goldberg, 1992; Goldberg, 1993). This allows explanation of each factor's observed patterns of
34
35 353 variation and can be used to predict various aspects of an individual's behaviour such as job-related
36
37 354 performance (Goldberg, 1993). The method has been successfully used to compare personality across
38
39 355 species (for example Gosling, Kwan & John, 2003; Weiss et al., 2011) and contexts (for example
40
41 356 King, Weiss, & Farmer, 2005) in differential and comparative psychology. While some cross-species
42
43 357 studies (Carter & Feeney, 2012; Mettke-Hofmann et al., 2005; Webster, Ward & Hart, 2009) and
44
45 358 cross-context studies (Sih, Kats & Maurer, 2003; Sih & Watters, 2005) have used this approach in
46
47 359 behavioural ecology, they are rare and factor analysis is not a preferred method. However, we suggest
48
49 360 that a factor analytic approach may be appropriate in behavioural ecology to 1) establish independent
50
51 361 factors/axes of correlated personality traits (such as a boldness-aggression axis), and 2) investigate
52
53 362 how these orthogonal axes impact on individual behavioural ecology, life history productivity and
54
55 363 fitness (Biro & Stamps, 2008). Such a factorial approach would also have the benefit of reducing the
56
57 364 problems associated with jingle-jangle fallacies.
58
59
60

1
2
3 365
4
5

6 366 Réale et al. (2007) recommended using a ‘simplified terminology’ for animal personality traits that
7
8 367 has been widely adopted within animal personality studies (the ‘Big Five’ animal personality traits:
9
10 368 boldness, exploration-avoidance, activity, aggressiveness and sociability). Réale et al. (2007)
11
12 369 proposed these terms as categories of traits rather than as traits themselves, warning that the
13
14 370 terminology was not exhaustive and it should be reviewed after sufficient research had been
15
16 371 conducted. Again, meta-analytical approaches may prove fruitful when attempting to review Réale et
17
18 372 al.’s (2007) terminology and factor orthogonality may be a desirable goal. Overall, we see the
19
20 373 factorial approach as compatible with existing animal personality research: 1) behavioural syndromes
21
22 374 aim to identify correlations between traits, and a factorial approach would identify orthogonal factors
23
24 375 (behavioural syndromes) composed of components (traits), and 2) these syndromes/factors could be
25
26 376 used to investigate both proximate and ultimate animal personality questions (for further discussion of
27
28 377 hierarchical personality taxonomy, see Uher, 2011; Uher et al., 2011).

30
31 378
32
33

34 379 **(5) Measurement considerations**

35
36
37 380 While the previous sections have raised theoretical issues with choosing tests and interpreting their
38
39 381 results, this section is concerned with decisions about what to measure. These considerations can
40
41 382 seem trivial (and are thus frequently overlooked), but may have a substantial effect on the outcome of
42
43 383 the study. First, there is the consideration of what method to use to collect the data. Currently there are
44
45 384 three different methods used: subjective personality ratings, behavioural coding and experimentation
46
47 385 (Jones & Gosling, 2005; Koski, 2011; Uher, 2008, 2011; Uher et al., 2011; Vazire et al., 2007).

48
49
50 386
51

52
53 387 Subjective assessments use ratings of multiple items, such as adjectives or behavioural descriptors, by
54
55 388 observers familiar with individual animals to describe the dimensions encompassing multiple
56
57 389 personality axes. Behavioural codings consist of recording the behaviour of a focal individual
58
59
60

1
2
3 390 according to a predetermined ethogram, that is, a list of discrete behaviours performed by the species
4
5 391 under study during the animal's natural behaviour (Gosling, 2001; Vazire et al., 2007). Experimental
6
7 392 assessments similarly record individual patterns of behaviour, but in response to controlled
8
9 393 experimental stimuli, to assess variability in a 'limited' number of personality axes such as boldness,
10
11 394 aggressiveness or sociability (Nettle & Penke, 2010; Sih et al., 2004b).

12
13
14 395

15
16
17 396 Combinations of these approaches are used by personality researchers in the fields of comparative and
18
19 397 differential psychology and behavioural ecology, to varying degrees. Behavioural ecologists
20
21 398 frequently use behavioural codings of natural behaviour and behaviour during experimentation to
22
23 399 assess personality, while comparative psychologists often use ratings and behavioural codings of
24
25 400 natural behaviour (hereafter, natural behaviour) and also use experimentation to assess personality
26
27 401 (Uher, 2008; Uher & Asendorpf, 2008). Relationships are well established between natural behaviour
28
29 402 and experimentation, and between natural behaviour and ratings (for example Carter et al., 2010;
30
31 403 Konečná et al., 2008; Kurvers et al., 2010; Pederson, King & Landau, 2005), and there is some
32
33 404 evidence that ratings are indicative of experimental responses as well (see Carter et al., 2012 and
34
35 405 references therein). While each approach has benefits and drawbacks (Koski, 2011), the high
36
37 406 correspondence between measures suggests simultaneous application of these methods may be
38
39 407 beneficial (Uher & Asendorpf, 2008), at the very least to test convergent validity (Campbell & Fiske,
40
41 408 1959).

42
43
44 409

45
46
47 410 The second measurement consideration regards situational strength (Uher, 2011). A situation in
48
49 411 behavioural ecology refers to the conditions at the time of the test (for example, temperature or
50
51 412 available resources; Sih et al., 2004b). Situational strength refers to how much an individual's
52
53 413 behaviour is influenced by the situation (Uher, 2011). Strong situations may leave little variation
54
55 414 between individual's behaviour whereas weaker situations may allow more inter-individual
56
57 415 differences to show (Mischel, 1973a,b). Accordingly, strong situations may be inappropriate as
58
59
60

1
2
3 416 behavioural tests given that little variation may exist between individuals due to floor and ceiling
4
5 417 effects (Stamps & Groothuis, 2010; Uher, 2011). Further, cutting off data after a certain time (for
6
7 418 example, see Carter et al., 2012) may result in biased personality estimates due to ceiling effects, the
8
9 419 implications of which are rarely considered in behavioural ecology (but see Stamps, Briffa & Biro,
10
11 420 2012). For example, are all individual baboons (*Papio ursinus*) that handle a novel food item for
12
13 421 longer than 150 seconds equally bold (Carter et al., 2012)?
14
15
16 422

17
18
19 423 Finally, there has been much debate in psychology about whether the situation or the person is more
20
21 424 important during behavioural tests (Mischel, 1968; Mischel, 1973a,b; Mischel, 1999; Mischel, Shoda
22
23 425 & Mendoza-Denton, 2002; Tett & Guterman, 2000; see also Uher, 2011), leading to the field of
24
25 426 interactionist psychology. Far from always expecting behavioural consistency, interactionist
26
27 427 psychologists expect that while some individuals will behave consistently between different situations
28
29 428 there are certain situations that can cause different individuals to behave similarly (high situational
30
31 429 strength) (Tett & Guterman, 2000). A concept similar to interactionist psychology in behavioural
32
33 430 ecology is the behavioural reaction norm (Dingemanse et al., 2010b; Nussey, Wilson & Brommer,
34
35 431 2007) in which individual's responses to different situations or contexts are modelled using random
36
37 432 regressions (Nussey et al., 2007; van de Pol & Wright, 2009). This has the potential to provide a
38
39 433 useful conceptual framework for understanding interactions between animal identity and the situation
40
41 434 interactions where the response variable is the same, however few animal personality researchers have
42
43 435 considered individual by situation interactions thus far (for examples, see Betini & Norris, 2012;
44
45 436 Dingemanse et al., 2012).
46
47
48 437

49 50 51 438 IV. HOW TO MEASURE PERSONALITY TRAITS 52

53
54 439 We suggested that the indiscriminate use of 'standard' behavioural tests within animal personality
55
56 440 studies may lead to the spurious labelling of personality traits, and have asked whether some studies
57
58 441 are actually measuring what they intend to. Further, we explored, in fields such as comparative
59
60

1
2
3 442 psychology, the past experiences of the problems and pitfalls that the animal personality approach in
4
5 443 behavioural ecology is now facing. In this section, we incorporate the tools from these other fields
6
7 444 into suggestions or goals that can be integrated into future research programs. As an exciting and fast-
8
9 445 moving field, animal personality research is potentially at risk from academic ‘faddists’ (termed
10
11 446 Zeitgeist-Shysters: Denenberg 1969; cited in Fetterman, 1986). Therefore, if animal personality
12
13 447 research is to maintain its rate of progress, it must develop a robust methodology including multiple
14
15 448 trait tests, reliability and validation (Cronbach & Meehl, 1955; Uher, 2011; Weiss & Adams, In
16
17 449 press), and continue to clearly identify research questions and hypotheses at their outset. Below we
18
19 450 suggest methodological goals that include the tools discussed above, which can be incorporated into
20
21 451 future animal personality studies.

22
23
24 452 *1. Consider test design.* We have reviewed the many conceptual and practical measurement
25
26 453 issues associated with different tests. Before starting to collect data, researchers should
27
28 454 consider:

- 29
30
31 455 a) Which method(s) should be used (subjective personality ratings, behavioural coding
32
33 456 or experimentation).
34
35
36 457 b) Whether the test actually measures the targeted trait in that species? That is, does the
37
38 458 test have situational relevance?
39
40
41 459 c) Situation strength and floor and ceiling effects. Is the stimulus situation too strong?
42
43 460 Will having a cut-off in a weak situation affect an individual’s position along a
44
45 461 personality gradient?

46
47
48 462 *2. Develop multiple tests for each trait of interest in the study.* Multiple tests of a trait are
49
50 463 necessary to establish the reliabilities and validities of the trait and the tests (step 3 below).
51
52 464 For each trait that is to be investigated by the research question, we suggest consideration of
53
54 465 the following:

- 55
56
57 466 a) Does the test have face validity?
58
59
60

- 1
2
3 467 b) Have the tests for the trait been used previously? Do the results of the previous studies
4
5 468 suggest that these tests are applicable to this study and system?
6
7
8 469 c) Identify a primary test for a trait, and at least one other that can each be used to test for
9
10 470 convergent or discriminant validities.
11
12
13 471 d) Make explicit predictions about where correlations should occur between the chosen tests
14
15 472 if the tests are measuring the targeted traits i.e. consider construct validity before data are
16
17 473 collected (see Section III.2.c above).
18
19
20 474 3. *Validate the tests used in the study.* Validation should comprise:
21
22
23 475 a) repeatability,
24
25 476 b) Ecological validity (Réale et al. 2007),
26
27
28 477 c) convergent validity (a correlation between two tests that theoretically measure the
29
30 478 same trait) and,
31
32
33 479 d) discriminant validity (a lack of a correlation between two tests that are hypothesised
34
35 480 to measure different traits) for each test.
36
37

38 481 Alternatively, to avoid Type I errors through testing each measurement against all other
39
40 482 measurements (Dochtermann, 2010), principal components analysis or structural equation
41
42 483 modelling (Loehlin, 1998; Dingemans, Dochtermann & Wright, 2010a) can be used to
43
44 484 investigate which test measurements load with others (but see Block, 1995 for a criticism
45
46 485 of data reduction techniques for this purpose). We emphasise again that a correlation
47
48 486 between two tests may come about by either the tests measuring the same trait or the
49
50 487 measured traits being linked by an underlying behavioural syndrome, depending on
51
52 488 whether a personality trait-validation or behavioural syndromes-identification focussed
53
54 489 interpretation is being used (table 3). In this case, careful choice of the validation tests is
55
56 490 essential even when investigating behavioural syndromes. Alternatively, a factor analytic
57
58
59
60

1
2
3 491 approach may be useful to integrate correlated traits such as a boldness-aggression factor.

4
5 492 We can then ask Winter et al.'s (1998) modified questions for the tests:

6
7
8 493 a. Do the tests measure the same trait with different labels, or are they different traits?

9
10
11 494 b. How much do the tests overlap? For example, while exploration could be measured using
12 495 distance travelled in an open field, this measurement may concurrently measure activity.

13
14 496 Should activity be controlled for when using this as a measurement of exploration?

15
16
17
18 497 c. Are all of the tests required to understand animal personality traits?

19
20 498 If the tests do not load/correlate as predicted, consider the assumptions made.

21
22
23 499

24
25
26 500 V. CONCLUSIONS

27
28
29 501 1. Throughout this review we have emphasised our view that the behavioural ecological approach
30 502 to animal personality research is facing methodological and conceptual obstacles that may
31 503 hinder its progress.

32
33
34
35
36 504 2. Current use of personality tests can be problematic as in some cases different tests may be
37 505 measuring the same personality trait (many to one) whereas in other cases one test may
38 506 measure many traits (one to many). This makes their interpretation difficult and limits the
39 507 scope and comparability of current studies.

40
41
42
43
44 508 3. We have examined lessons learned by psychologists, and suggest tools that could be borrowed
45 509 from the psychometric literature.

46
47
48
49 510 4. We make suggestions for how future studies might use these tools to work towards a more
50 511 unified and robust model of animal personality. In this framework we suggest the use of
51 512 multiple tests for measuring personality traits wherever possible, and urge that more
52 513 consideration be given to interpreting the observed correlations between tests.

1
2
3 514 5. Despite the obstacles we outlined, we believe the tools exist to strengthen methodology in the
4
5 515 field of animal personality research in behavioural ecology, and to further its exciting and rapid
6
7 516 progress.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Review Only

1
2
3 517 Acknowledgements
4
5

6 518 The authors thank Alex Weiss, Sam Gosling and an anonymous reviewer for detailed
7

8 519 comments on an earlier version of this work and Culum Brown for reading a preliminary
9

10 520 outline. AJC is supported by a Fenner School of Environment and Society Studentship; WEF
11

12 521 by a Research School of Biology studentship; HHM by a NERC Open CASE
13

14 522 (NE/F013442/1) studentship with ZSL as CASE partner.
15

16
17 523
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Review Only

524 References

- 525 ARCHER, J. (1973). Tests for emotionality in rats and mice: a review. *Animal Behaviour* **21**,
526 205-35.
- 527 BEACH, F. A. (1950). The Snark was a Boojum. *American Psychologist* **5**, 115-124.
- 528 BELL, A. M. (2007). Future directions in behavioural syndromes research. *Proceedings of the*
529 *Royal Society Biological Sciences Series B* **274**, 755-761.
- 530 BELL, A. M., HANKISON, S. J. & LASKOWSKI, K. L. (2009). The repeatability of behaviour: a meta-
531 analysis. *Animal Behaviour* **77**, 771-783.
- 532 BETINI, G. S. & NORRIS, D. R. (2012). The relationship between personality and plasticity in tree
533 swallow aggression and the consequences for reproductive success. *Animal*
534 *Behaviour* **83**, 137-143.
- 535 BIRO, P. A. & STAMPS, J. A. (2008). Are animal personality traits linked to life-history
536 productivity? *Trends in Ecology & Evolution* **23**, 361-368.
- 537 BLOCK, J. (1995). A Contrarian View of the Five Factor Approach to Personality Description.
538 *Psychological Bulletin* **117**, 187-215.
- 539 BLUMSTEIN, D. T. & DANIEL, J. C. (2003). Red kangaroos (*Macropus rufus*) receive an
540 antipredator benefit from aggregation. *Acta Ethologica* **5**, 95-99.
- 541 BOICE, R. (1971). On the fall of comparative psychology. *American Psychologist* **26**, 858-859.
- 542 BOYER, N., REALE, D., MARMET, J., PISANU, B. & CHAPUIS, J.-L. (2010). Personality, space use and
543 tick load in an introduced population of Siberian chipmunks *Tamias sibiricus*. *Journal*
544 *of Animal Ecology* **79**, 538-547.
- 545 BREMNER-HARRISON, S. B., PRODOHL, P. A. & ELWOOD, R. W. (2004). Behavioural trait assessment
546 as a release criterion: boldness predicts early death in a reintroduction programme
547 of captive-bred swift fox (*Vulpes velox*). *Animal Conservation* **7**, 313-320.
- 548 BROWN, C. & BRAITHWAITE, V. A. (2004). Size matters: a test of boldness in eight populations of
549 the poeciliid *Brachyraphis episcopi*. *Animal Behaviour* **68**, 1325-1329.
- 550 BUDAEV, S. & BROWN, C. (2011). Personality Traits and Behaviour. In *Fish Cognition and*
551 *Behavior* (ed. C. Brown, K. N. Laland and J. Krause), pp. 135-165. Wiley-Blackwell,
552 Oxford.
- 553 BURNS, J. G. (2008). The Validity of Three Tests of Temperament in Guppies (*Poecilia*
554 *reticulata*). *Journal of Comparative Psychology* **122**, 344-356.
- 555 BUTLER, M. W., TOOMEY, M. B., MCGRAW, K. J. & ROWE, M. (2011). Ontogenetic immune
556 challenges shape adult personality in mallard ducks. *Proceedings of the Royal Society*
557 *Biological Sciences Series B*, Available online.
- 558 CAMPBELL, D. T. & FISKE, D. W. (1959). Convergent and Discriminant Validation by the
559 Multitrait-Multimethod Matrix. *Psychological Bulletin* **56**, 81-105.
- 560 CARO, T. & SHERMAN, P. W. (2011). Endangered species and a threatened discipline:
561 behavioural ecology. *Trends in Ecology & Evolution* **26**, 111-118.
- 562 CARTER, A. J. & FEENEY, W. E. (2012) Taking a Comparative Approach: Analysing Personality as
563 a Multivariate behavioural Response across Species. *PLoS ONE* **7**, e42440.
- 564 CARTER, A. J., GOLDIZEN, A. W. & TROMP, S. A. (2010). Agamas exhibit behavioral syndromes:
565 bolder males bask and feed more but may suffer higher predation. *Behavioral*
566 *Ecology* **21**, 655-661.
- 567 CARTER, A. J., MARSHALL, H., HEINSOHN, R. & COWLISHAW, G. (In press). How not to measure
568 boldness: novel object and antipredator responses are not the same in wild
569 baboons. *Animal Behaviour*
- 570 CARTER, A. J., MARSHALL, H., HEINSOHN, R. & COWLISHAW, G. (2012). Evaluating animal
571 personalities: do observer assessments and experimental tests measure the same
572 thing? *Behavioral Ecology and Sociobiology* **66**, 153-160.
- 573 COLEMAN, K. & WILSON, D. S. (1998). Shyness and boldness in pumpkinseed sunfish: individual
574 differences are context-specific. *Animal Behaviour* **56**, 927-936.

- 1
2
3 575 CONRAD, J. L., WEINERSMITH, K. L., BRODIN, T., SALTZ, J. B. & SIH, A. (2011). Behavioural
4 576 syndromes in fishes: a review with implications for ecology and fisheries
5 577 management. *Journal of Fish Biology* **78**, 395-435.
6 578 CRONBACH, L. J. & MEEHL, P. E. (1955). Construct Validity in Psychological Tests. *Psychological*
7 579 *Bulletin* **52**, 281-302.
8 580 CRUSIO, W. E. (2001). Genetic dissection of mouse exploratory behaviour. *Behavioural Brain*
9 581 *Research* **125**, 127-132.
10 582 DENENBERG, V. H. (1973). Open-Field Behavior in the Rat: what does it mean? *Annals of the*
11 583 *New York Academy of Sciences* **159**, 852-859.
12 584 DINGEMANSE, N. J., BOUWMAN, K. M., VAN DE POL, M., VAN OVERVELD, T., PATRICK, S. C., MATTHYSEN,
13 585 E. & QUINN, J. L. (2012). Variation in personality and behavioural plasticity across four
14 586 populations of the great tit *Parus major*. *Journal of Animal Ecology*, 116-126.
15 587 DINGEMANSE, N. J., DOCHTERMANN, N. & WRIGHT, J. (2010a). A method for exploring the
16 588 structure of behavioural syndromes to allow formal comparison within and between
17 589 data sets. *Animal Behaviour* **79**, 439-450.
18 590 DINGEMANSE, N. J., KAZEM, A. J. N., RÉALE, D. & WRIGHT, J. (2010b). Behavioural reaction norms:
19 591 animal personality meets individual plasticity. *Trends in Ecology & Evolution* **25**, 81-
20 592 89.
21 593 DINGEMANSE, N. J. & WOLF, M. (2010). Recent models for adaptive personality differences: a
22 594 review. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**,
23 595 3947-3958.
24 596 DINGEMANSE, N. J., WRIGHT, J., KAZEM, A. J. N., THOMAS, D. K., HICKLING, R. & DAWNAY, N. (2007).
25 597 Behavioural syndromes differ predictably between 12 populations of three-spined
26 598 stickleback. *Journal of Animal Ecology* **76**, 1128-1138.
27 599 DOCHTERMANN, N. A. (2010). Behavioral syndromes: carryover effects, false discovery rates,
30 600 and a priori hypotheses. *Behavioral Ecology* **21**, 437-439.
31 601 DUCKWORTH, A. L. & KERN, M. L. (2011). A meta-analysis of the convergent validity of self-
32 602 control measures. *Journal of Research in Personality* **45**, 259-268.
33 603 FETTERMAN, D. M. (1986). Ethnography and Qualitative Design in Educational-Research -
34 604 Goetz, Jp, Lecompte, Md. *American Anthropologist* **88**, 764-765.
35 605 FOX, R. A., LADAGE, L. D., ROTH, T. C. & PRAVOSUDOV, V. V. (2009). Behavioural profile predicts
36 606 dominance status in mountain chickadees, *Poecile gambeli*. *Animal Behaviour* **77**,
37 607 1441-1448.
38 608 FROST, A. J., WINROW-GIFFEN, A., ASHLEY, P. J. & SNEDDON, L. U. (2007). Plasticity in animal
39 609 personality traits: does prior experience alter the degree of boldness? *Proceedings*
40 610 *of the Royal Society Biological Sciences Series B* **274**, 333-339.
41 611 GOLDBERG, L. R. (1992). The development of markers for the big-five factor structure.
42 612 *Psychological Assessment* **4**, 26-42.
43 613 GOLDBERG, L. R. (1993). The Structure of Phenotypic Personality Traits. *American Psychologist*
44 614 **48**, 26-34.
45 615 GORSUCH, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
46 616 GOSLING, S. D. (2001). From mice to men: What can we learn about personality from animal
47 617 research? *Psychological Bulletin* **127**, 45-86.
48 618 GOSLING, S. D. & JOHN, O. P. (1999). Personality dimensions in nonhuman animals: A cross-
49 619 species review. *Current Directions in Psychological Science* **8**, 69-75.
50 620 GOSLING, S. D., KWAN, V. S. Y. & JOHN, O. P. (2003). A dog's got personality: A cross-species
51 621 comparative approach to personality judgments in dogs and humans. *Journal of*
52 622 *Personality and Social Psychology* **85**, 1161-1169.
53 623 HALL, C. S. (1934). Emotional behavior in the rat. I. Defecation and urination as measures of
54 624 individual differences in emotionality. *Journal of Comparative Psychology* **18**, 385-
55 625 403.

- 1
2
3 626 HODOS, W. & CAMPBELL, C. B. (1969). Scala Naturae - Why There Is No Theory in Comparative
4 627 Psychology. *Psychological Review* **76**, 337-&.
- 5 628 JACOBY, J. (1978). Consumer Research - State of Art Review. *Journal of Marketing* **42**, 87-96.
- 6 629 JOHN, O. P., ROBINS, R. P. & PERVIN, L. A. (2008). *Handbook of Personality: Theory & Research*.
7 630 The Guilford Press, New York, NY.
- 8 631 JONES, A. C. & GOSLING, S. D. (2005). Temperament and personality in dogs (*Canis familiaris*): A
9 632 review and evaluation of past research. *Applied Animal Behaviour Science* **95**, 1-53.
- 10 633 KAVANAU, J. L. (1967). Behavior of captive white-footed mice. *Science* **155**, 1623-39.
- 11 634 KING, J. E., WEISS, A. & FARMER, K. H. (2005). A chimpanzee (*Pan troglodytes*) analogue of
12 635 cross-national generalization of personality structure: Zoological parks and an
13 636 African sanctuary. *Journal of Personality*, **73**, 389-410.
- 14 637 KONEČNÁ, M., LHOTA, S., WEISS, A., URBANEK, T., ADAMOVA, T. & PLUHACEK, J. (2008). Personality
15 638 in Free-Ranging Hanuman Langur (*Semnopithecus entellus*) Males: Subjective Ratings
16 639 and Recorded Behavior. *Journal of Comparative Psychology* **122**, 379-389.
- 17 640 KOSKI, S. E. (2011). How to Measure Animal Personality and Why Does It Matter? Integrating
18 641 the Psychological and Biological Approaches to Animal Personality. In *From Genes to*
19 642 *Animal Behaviour: Social structures, personalities, communication by color*, vol. 0.
20 643 *Primate Monographs* (ed. M. Inoue-Murayama, S. Kawamura and A. Weiss), pp.
21 644 115-136. Springer Japan.
- 22 645 KURVERS, R. H. J. M., EIJKELINKAMP, B., VAN OERS, K., VAN LITH, B., VAN WIEREN, S. E., YDENBERG, R.
23 646 C. & PRINS, H. H. T. (2009). Personality differences explain leadership in barnacle
24 647 geese. *Animal Behaviour* **78**, 447-453.
- 25 648 KURVERS, R. H. J. M., PRINS, H. H. T., VAN WIEREN, S. E., VAN OERS, K., NOLET, B. A. & YDENBERG, R. C.
26 649 (2010). The effect of personality on social foraging: shy barnacle geese scrounge
27 650 more. *Proceedings of the Royal Society B-Biological Sciences* **277**, 601-608.
- 28 651 LAY, C. (1973). Some notes on the concept of cross-situational consistency. In *Personality at*
29 652 *the crossroads: Current issues in interactional psychology* (ed. D. Magnusson and N.
30 653 S. Endler), pp. 143-146. John Wiley & Sons, New York.
- 31 654 LESTER, D. (1968). Curiosity and the role of the role of the individual in the imposition of
32 655 change. *Psychol Rep* **22**, 109-12.
- 33 656 LOCKARD, R. B. (1971). Reflections on the fall of comparative psychology: is there a message
34 657 for us all? *American Psychologist* **26**, 168-179.
- 35 658 LOEHLIN, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural*
36 659 *analysis* (3rd ed.). Lawrence Erlbaum Associates, Mahwah, NJ.
- 37 660 LOPEZ, P., HAWLENA, D., POLO, V., AMO, L. & MARTIN, J. (2005). Sources of individual shy-bold
38 661 variations in antipredator behaviour of male Iberian rock lizards. *Animal Behaviour*
39 662 **69**, 1-9.
- 40 663 MATHER, J. A. & ANDERSON, R. C. (1993). Personalities of Octopuses (*Octopus rubescens*).
41 664 *Journal of Comparative Psychology* **107**, 336-340.
- 42 665 MCCRAE, R. R. (1982). Consensual Validation of Personality Traits - Evidence from Self-
43 666 Reports and Ratings. *Journal of Personality and Social Psychology* **43**, 293-303.
- 44 667 METTKE-HOFMANN, C., EBERT, C., SCHMIDT, T., STEIGER, S. & STIEB, S. (2005). Personality traits in
45 668 resident and migratory warbler species. *Behaviour* **142**, 1357-1375.
- 46 669 MILLER, K. A., GARNER, J. P. & MENCH, J. A. (2005). The test-retest reliability of four behavioural
47 670 tests of fearfulness for quail: a critical evaluation. *Applied Animal Behaviour Science*
48 671 **92**, 113-127.
- 49 672 MISCHEL, W. (1968). *Personality and assessment*. John Wiley and Sons, New York, NY.
- 50 673 MISCHEL, W. (1973a). The interaction of person and situation. In *Personality at the*
51 674 *crossroads: Current issues in interactional psychology* (ed. D. Magnusson and N. S.
52 675 Endler), pp. 333-352. John Wiley & Sons, New York.

- 1
2
3 676 MISCHEL, W. (1973b). Toward a Cognitive Social Learning Reconceptualization of Personality.
4 677 *Psychological Review* **80**, 252-283.
- 5 678 MISCHEL, W. (1999). Implications of person-situation interaction: Getting over the field's
6 679 borderline personality disorder. *European Journal of Personality* **13**, 455-461.
- 7 680 MISCHEL, W., SHODA, Y. & MENDOZA-DENTON, R. (2002). Situation-behavior profiles as a locus of
8 681 consistency in personality. *Current Directions in Psychological Science* **11**, 50-54.
- 9 682 MISSLIN, R. & CIGRANG, M. (1986). Does Neophobia Necessarily Imply Fear or Anxiety.
10 683 *Behavioural Processes* **12**, 45-50.
- 11 684 NETTLE, D. & PENKE, L. (2010). Personality: bridging the literatures from human psychology
12 685 and behavioural ecology. *Philosophical Transactions of the Royal Society B-Biological*
13 686 *Sciences* **365**, 4043-4050.
- 14 687 NEVO, B. (1985). Face Validity Revisited. *Journal of Educational Measurement* **22**, 287-293.
- 15 688 NUNNALLY, J. C. (1978). *Psychometric Theory*. McGraw-Hill Book Company, New York.
- 16 689 NUSSEY, D. H., WILSON, A. J. & BROMMER, J. E. (2007). The evolutionary ecology of individual
17 690 phenotypic plasticity in wild populations. *Journal of Evolutionary Biology* **20**, 831-
18 691 844.
- 19 692 PEDERSON, A. K., KING, J. E. & LANDAU, V. I. (2005). Chimpanzee (*Pan troglodytes*) personality
20 693 predicts behavior. *Journal of Research in Personality* **39**, 534-549.
- 21 694 PETER, J. P. (1979). Reliability - Review of Psychometric Basics and Recent Marketing
22 695 Practices. *Journal of Marketing Research* **16**, 6-17.
- 23 696 PRUITT, J. N., RIECHERT, S. E. & JONES, T. C. (2008). Behavioural syndromes and their fitness
24 697 consequences in a socially polymorphic spider, *Anelosimus studiosus*. *Animal*
25 698 *Behaviour* **76**, 871-879.
- 26 699 RÉALE, D., DINGEMANSE, N. J., KAZEM, A. J. N. & WRIGHT, J. (2010). Evolutionary and ecological
27 700 approaches to the study of personality. *Philosophical Transactions of the Royal*
28 701 *Society B-Biological Sciences* **365**, 3937-3946.
- 29 702 RÉALE, D., READER, S. M., SOL, D., MCDUGALL, P. T. & DINGEMANSE, N. J. (2007). Integrating
30 703 animal temperament within ecology and evolution. *Biological Reviews* **82**, 291-318.
- 31 704 SHORT, K. H. & PETREN, K. (2008). Boldness underlies foraging success of invasive
32 705 *Lepidodactylus lugubris* geckos in the human landscape. *Animal Behaviour* **76**, 429-
33 706 437.
- 34 707 SIH, A., BELL, A. & JOHNSON, J. C. (2004a). Behavioral syndromes: an ecological and
35 708 evolutionary overview. *Trends in Ecology & Evolution* **19**, 372-378.
- 36 709 SIH, A., BELL, A. M., JOHNSON, J. C. & ZIEMBA, R. E. (2004b). Behavioral syndromes: An
37 710 integrative overview. *Quarterly Review of Biology* **79**, 241-277.
- 38 711 SIH, A., KATS, L. B. & MAURER, E. F. (2003). Behavioural correlations across situations and the
39 712 evolution of antipredator behaviour in a sunfish-salamander system. *Animal*
40 713 *Behaviour* **65**, 29-44.
- 41 714 SIH, A. & WATTERS, J. V. (2005). The mix matters: behavioural types and group dynamics in
42 715 water striders. *Behaviour* **142**, 1417-1431.
- 43 716 STAMPS, J. & GROOTHUIS, T. G. G. (2010). The development of animal personality: relevance,
44 717 concepts and perspectives. *Biological Reviews* **85**, 301-325.
- 45 718 TETT, R. P. & GUTERMAN, H. A. (2000). Situation trait relevance, trait expression, and cross-
46 719 situational consistency: Testing a principle of trait activation. *Journal of Research in*
47 720 *Personality* **34**, 397-423.
- 48 721 TOMS, C. N., ECHEVARRIA, D. J. & JOUANDOT, D. J. (2010). A methodological review of
49 722 personality-related studies in fish: Focus on the shy-bold axis of behaviour.
50 723 *International Journal of Comparative Psychology* **23**, 1-25.
- 51 724 UHER, J. (2008). Comparative personality research: Methodological approaches. *European*
52 725 *Journal of Personality* **22**, 427-455.

- 1
2
3 726 UHER, J. (2011). Individual behavioral phenotypes: An integrative meta-theoretical
4 727 framework. Why “behavioral syndromes” are not analogs of “personality”.
5 728 *Developmental Psychobiology* **53**, 521-548.
- 6 729 UHER, J. & ASENDORPF, J. B. (2008). Personality assessment in the Great Apes: Comparing
7 730 ecologically valid behavior measures, behavior ratings, and adjective ratings. *Journal*
8 731 *of Research in Personality* **42**, 821-838.
- 9 732 UHER, J. Personality in nonhuman primates. What can we learn from human personality
10 733 psychology? *Personality and temperament in nonhuman primates* (eds. A. Weiss, J.
11 734 E. King, & L. Murray), pp. 41-76. Springer, New York, NY.
- 12 735 Personality and Temperament in Nonhuman Primates. *Developments in Primatology:*
13 736 *Progress and Prospects* (ed. L. Barrett), pp. 41-76. Springer New York.
- 14 737 VAN DE POL, M. V. & WRIGHT, J. (2009). A simple method for distinguishing within- versus
15 738 between-subject effects using mixed models. *Animal Behaviour* **77**, 753-758.
- 16 739 VAZIRE, S., GOSLING, S. D., DICKEY, A. S. & SCHAPIRO, S. J. (2007). Measuring personality in
17 740 nonhuman animals. In *Handbook of Research Methods in Personality Psychology* (ed.
18 741 R. W. Robins, R. C. Fraley and R. F. Krueger), pp. 190-206. Guilford, New York.
- 19 742 WALSH, R. N. & CUMMINS, R. A. (1976). The Open-Field Test - A Critical Review. *Psychological*
20 743 *Bulletin* **83**, 482-504.
- 21 744 WEBSTER, M. M., WARD, A. J. W. & HART, P. J. B. (2009). Individual boldness affects interspecific
22 745 interactions in sticklebacks. *Behavioral Ecology and Sociobiology* **63**, 511-520.
- 23 746 WEISS, A. & ADAMS, M. J. (In press). Differential behavioral ecology. In *Animal personalities:*
24 747 *Behavior, physiology and evolution*, (ed. C. Carere and D. Maestriperi). University of
25 748 Chicago Press, Chicago, IL.
- 26 749 WEISS, A., ADAMS, M. J., WIDDIG, A. & GERALD, M. S. (2011). Rhesus Macaques (*Macaca*
27 750 *mulatta*) as Living Fossils of Hominoid Personality and Subjective Well-Being. *Journal*
28 751 *of Comparative Psychology* **125**, 72-83.
- 29 752 WHITTINGHAM, M. J., BUTLER, S. J., QUINN, J. L. & CRESSWELL, W. (2004). The effect of limited
30 753 visibility on vigilance behaviour and speed of predator detection: implications for
31 754 the conservation of granivorous passerines. *Oikos* **106**, 377-385.
- 32 755 WINTER, D. G., JOHN, O. P., STEWART, A. J., KLOHNEN, E. C. & DUNCAN, L. E. (1998). Traits and
33 756 motives: Toward an integration of two traditions in personality research.
34 757 *Psychological Review* **105**, 230-250.
- 35 758 WOLF, M. & WEISSING, F. J. (2010). An explanatory framework for adaptive personality
36 759 differences. *Philosophical Transactions of the Royal Society B-Biological Sciences*
37 760 **365**, 3959-3968.
- 38
39
40
41 761
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

762 **Table 1**
763 Glossary of terms.

764

765 **Table 2**

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

768

769 **Table 3**

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

773

774 **Table 4**

775 When assessing the validity of a test, the validity should be checked using multiple other tests
776 (multi-trait, multi-test approach; see text). 'Convergent' represents a correlation between the
777 traits, 'discriminant' represents no correlation. The bottom diagonal demonstrates the
778 predicted correlations between tests if the test (e.g. novel object test) measured, in this
779 example, exploration. The top diagonal represents the predicted correlations between tests if
780 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 (Personality) trait	Between-individual differences in behaviour that persist through time. 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

782

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)

783

784

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)

785

786

787

Table 4

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

788

789

For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60