Ambiguous terminology in animal personality research: a self-report questionnaire and a systematic review

Alfredo Sánchez-Tójar[1*$], Maria Moiron[2,3$], Petri T. Niemelä[4,5$]

[$] All authors contributed equally to this work

[1]Department of Evolutionary Biology, Bielefeld University, Germany

[2]CEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France

[3]Institute of Avian Research, Wilhelmshaven, Germany

[4]Behavioral Ecology, Department of Biology, Ludwig-Maximillians University of Munich, Planegg-Martinsried, Germany

[5]Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland

[*]Author for correspondence: Alfredo Sánchez-Tójar (alfredo.tojar@gmail.com)

ORCIDs and email addresses:

Alfredo Sánchez-Tójar: 0000-0002-2886-0649; alfredo.tojar@gmail.com

Maria Moiron: 0000-0003-0991-1460; mariamoironc@gmail.com

Petri T. Niemelä: 0000-0002-7518-4057; petri.t.niemela@gmail.com

Abstract

Whether animal personality studies provide insights of broader evolutionary and ecological relevance to behavioural ecology is frequently questioned. One source of controversy is the vast, but often vague conceptual terminology used. From a statistical perspective, animal personality is defined as repeatable among-individual variance in behaviour; however, numerous conceptual definitions of animal personality exist. Here, we performed a 1) self-report questionnaire and 2) systematic literature review to quantify how researchers interpret

conceptual and statistical definitions commonly used in animal personality research. We also investigated whether results from the questionnaire agree with those of the literature review. Among the 430 self-reported researchers that participated in our questionnaire, we observed discrepancies in key questions such as the conceptual definition of animal personality or the interpretation of repeatability. Our literature review generally confirmed the global patterns revealed by the questionnaire. Overall, we identified common disagreements in animal personality research and discussed potential solutions. We advocate for the usage of statistically-oriented terminology because conceptual definitions can seemingly be interpreted at multiple levels of biological organization. We expect that adopting such statistically-oriented terminology will, at least partly, avoid the confusion generated by the label “animal personality”, and ultimately help to clarify and move the field forward.

Keywords: behavioural syndrome, coping style, behavioural consistency, intraclass correlation coefficient, phenotypic variation, individual differences

1. Introduction

Labile traits such as behaviour and physiology are phenotypic traits whose expression can change as the environment changes (Scheiner, 1993). Labile traits can vary at multiple hierarchical levels: i) individuals can differ in their average trait expression (called among-individual variation) while at the same time, ii) individuals can change their trait expression from one instance to another (called within-individual variation, phenotypic flexibility or reversible plasticity; Forsman, 2015). Together, these two levels contribute to the iii) phenotypic expression - also called total phenotypic variation - of a labile trait. Studying among-individual variation in the expression of labile traits has become increasingly popular during the last two decades, particularly in the field of behavioural ecology (e.g., Sih et al.

2004, Stamps 2007, Dingemanse et al., 2010, Réale et al., 2007). The existence of repeatable among-individual variation in behaviour (i.e. “animal personality”) has been hypothesized and, subsequently, shown to have important consequences in ecology (population dynamics, survival, predator-prey interactions; e.g. Duckworth and Badyaev, 2007; Moiron et al., 2020), and evolution (social evolution, speciation, constraints on adaptive evolution; e.g. Dingemanse and Araya-Ajoy, 2015; Wolf and Weissing, 2012).

The ecological and evolutionary causes and consequences of the expression of among-individual variation in behaviour are relatively well investigated from a theoretical perspective (e.g., Dall et al., 2004; Dingemanse and Wolf, 2010; Rands et al., 2003; Sih et al., 2015). However, despite the important theoretical advances in this field and a series of highly-cited methodological articles providing guidelines for empiricists on how to study and interpret among- and within-individual variation in behaviour (Cleasby et al., 2015; Dingemanse and Dochtermann, 2013; van de Pol and Wright, 2009; Westneat et al., 2015), animal personality researchers often fail to empirically and statistically test those theories at the among-individual level, which is the level of variation of interest in animal personality research (Moiron et al., 2020; Niemelä and Dingemanse, 2018a, 2018b; Royauté et al., 2018). The current mismatch between theory, and data and methods used to test the theory may have arisen from different reasons. Empiricists might make the (informed or uninformed*)* decision of not using the required data (e.g., repeated measurements) and statistical methods (e.g., univariate and multivariate mixed-effects models) that allow to answer questions related to among-individual differences. In such cases, empiricists are (knowingly or unknowingly) applying the “individual gambit”, i.e., assuming that the total phenotypic expression of a trait can be used as a proxy for the among-individual level expression (Brommer, 2013; Dingemanse et al., 2012; Niemelä and Dingemanse, 2018b). Another potential reason for the

mismatch between theory and its empirical testing might be that empiricists are likely forced to choose among a plurality of different conceptual definitions of animal personality from the literature, which might be hard to connect to a statistical definition of animal personality. For example, the statistical definition of animal personality refers unambiguously to the existence of significant among-individual variance or variance-standardized among-individual variance in behaviour, i.e., repeatability (Dingemanse and Dochtermann, 2013), but conceptual definitions of animal personality very often include terminology that refers to both among- and within-individual variance in behaviour (e.g. among-individual differences and limited plasticity in behavioural expression, respectively; Table 1 and S1). Inconsistent and vague use of the conceptual terminology has potential to generate ambiguous and/or erroneous interpretations across animal personality studies.

Despite that several great attempts at clarifying concepts and terminology in animal personality have been made (Carter et al., 2013; David and Dall, 2016; Dingemanse and Wright, 2020; Kaiser and Müller, 2021; Niemelä and Dingemanse, 2018b; Roche et al., 2016; Sih et al., 2020; Stamps and Groothuis, 2010), diversity in terminology among those studies is present. Moreover no study has quantified diversity in, and preferred, conceptual terminology across the literature. Our work focuses on quantifying the diversity of definitions and interpretations of terminology used in the field of animal personality, and how conceptual definitions match with statistical definitions. To do so, we combine a 1) systematic literature review and a 2) self-report questionnaire. Finally, without generating additional concepts, we make suggestions on how to unify the already-in-use terminology.

2. Methods

2.1. Literature review

We performed a systematic literature search across all years in Web of Science Core Collection (details in Figure S1) using the search string TS=("animal personalit*") on the 3rd of June 2019, and obtained 1,223 references. Our main interest was not to perform a comprehensive review of the entire field, but rather have a standardized subsample of the literature on this topic, i.e. “animal personality”, which is why we intentionally decided to avoid searching for related terms such as “behavioural syndrome” or “coping style” (e.g. Coppens et al., 2010; Dingemanse et al., 2012). The search was used to: (a) compile contact details (email addresses) from a broad sample of animal personality researchers to forward the questionnaire (full search used); (b) review the animal personality terminology of the most cited articles in the field (full search used); and (d) review the interpretation and use of the statistical term “repeatability” (i.e. the most often used metric to define the existence of “animal personality”) in the recent literature (records published since 2015 used; more below).

To review the interpretation and use of the term “repeatability” in the literature, we selected the ten most represented journals in our search (i.e. Animal Behaviour (149), Behavioral Ecology (94), Behavioral Ecology and Sociobiology (78), Ethology (53), Proceedings of the Royal Society B (53), PLoS ONE (50), Behaviour (41), Behavioural Processes (41), Applied Animal Behaviour Science (28), Current Zoology (26)) and downloaded all those full records (613 references) for subsequent screening. Titles and abstracts were screened using the software Rayyan (Ouzzani et al., 2016). The three authors performed the title-and-abstract screening using a decision tree (Figure S2) and 22% of the records (n = 135) were screened by two observers to increase the reproducibility and reliability of the process. Only 2 conflicting decisions (1%) were recorded, which were then discussed and resolved collectively. We then

randomly selected half of the studies published since 2015 (n = 154 references), and extracted our variables of interest related to the interpretation and use of repeatability (see below). We did not attempt to extract data from all the studies that passed the title-and-abstract screening because our goal was not to perform a fully comprehensive review, but to explore the levels of agreement between the results of the self-report questionnaire and the literature review. We used 2015 as a reference year based on the publication of Cleasby et al. (2015), where the differences between repeatability, consistency and predictability were extensively described, and a statistical framework to investigate consistency and predictability was provided. The three authors performed data extraction, and data from 25% of the records (n = 39) were independently extracted by two observers to increase the reproducibility and reliability of the process. Conflicting decisions between observers were discussed and resolved collectively. Subsequently, each observer revised all decisions in their data subset, and a single observer (AST) performed a full database double-check to ensure overall data quality and consistency throughout (all corrections are detailed in the provided data file). After excluding those articles that did not estimate repeatability, our final database for the systematic literature review included a total of 88 articles (PRISMA diagram provided in Figure S3; Moher et al., 2009).

From each of the articles included in our final database, we extracted information about the interpretation of repeatability estimates. Specifically, we noted: (i) whether repeatability was interpreted at the among- and/or within-individual level; (ii) whether repeatability was interpreted as individual consistency and/or individual predictability; (iii) whether repeatability estimates between different groups (i.e., different treatments or groups of animals) within the same study were compared to each other, and if so, (iv) whether that comparison was interpreted at the among- and/or within-individual level; and (v) whether

unstandardized variances were reported within the focal study. For the purpose of our review, we defined among-individual level interpretations as those referring to interpretations based on average behavioural differences among individuals, whereas within-individual level interpretations referred to interpretations based on behavioural flexibility (also known as reversible plasticity). We considered an interpretation as "both" whenever the repeatability interpretation referred to both among- and within-individual levels, including those articles interpreting repeatability as "consistent individual differences" (see all decisions in the provided data). We provide a sensitivity test to show that conclusions remain qualitatively the same when interpreting "consistent individual differences" as only among- rather than both within- and among-individual variance (Appendix S3). To extract the information of interest from each focal article, we focused on the sections: "Title", "Abstract", "Statistical analyses" (or alike), "Results" and "Discussion". We did not extract data from the "Introduction" or any "Methods" sections except the "Statistical analyses" because we were specifically interested in the authors' own interpretations of their results. The intraclass correlation coefficient (ICC) was considered as a synonym of repeatability, unless stated otherwise in the article (Cleasby et al., 2015; Nakagawa and Schielzeth, 2010; Wolak et al., 2012).

2.2. Self-report questionnaire

We used the 30 most cited articles - according to Web of Science - from the literature search described above to obtain a tentative list of conceptual definitions of "animal personality", which we used in our questionnaire (Table 1). One author (AST) screened the titles and abstracts of those 30 articles and excluded all empirical studies. Articles that passed the title-and-abstract screening (n = 26) were subsequently read in detail by two authors (AST, MM) to extract animal personality definitions. We compiled five definitions of animal personality that were deemed to represent the whole spectrum of conceptual definitions found in those 26

highly cited articles, and presented them in the self-report questionnaire (full list available in the provided data).

The self-report questionnaire was anonymous, and consisted of three questions aimed at compiling information about the participant, and eight questions - and a "comments" section - aimed at understanding how participants interpreted and applied common terminology used in the field of animal personality (questionnaire available in Supplementary Information 2). We advertised the questionnaire on Twitter and ResearchGate on the 8th of January 2020, and shared the link with colleagues at multiple institutions. In addition, we compiled a list of 919 unique email addresses from the corresponding authors of all 1,223 references found by our search (see section 'Literature review'), and sent them an email on the 16th of January 2020 to encourage their participation in our questionnaire (email template available in Appendix S1). Participation in the questionnaire was possible until the 8th of February 2020 (i.e. one month after we first advertised it).

3. Results

3.1. Literature review

Our literature search identified 88 articles published in 9 journals (Animal Behaviour: 25, Behavioral Ecology: 15, Behavioral Ecology and Sociobiology: 10, Ethology: 9, Proceedings of the Royal Society B: 7, PLoS ONE: 1, Behaviour: 8, Behavioural Processes: 9, Current Zoology: 4) from 2015 to 2019 (2015: 14, 2016: 23, 2017: 25, 2018: 20, 2019: 6) that quantified repeatability estimates in behavioural traits and were therefore considered in our analyses.

After excluding articles that did not interpret repeatability (n = 2) or for which the interpretation was ambiguous (n = 2), we found that 70.2% of the articles (n = 59) interpreted

repeatability as reflecting both within- and among-individual level processes, whereas 25.0% (n = 21) and 4.8% (n = 4) interpreted repeatability only at the within-individual or only the among-individual level, respectively (see sensitivity test in Appendix S3). Additionally, 48.8% of the articles (n = 41) interpreted repeatability as a proxy of individual consistency and/or individual predictability. These two terms are, however, generally considered as metrics describing within-individual plasticity that require a different statistical approach than repeatability to be estimated (Cleasby et al., 2015).

Only 24 studies compared repeatability estimates between two or more treatments or groups of animals within the same study. We excluded those articles that did not provide the required data, i.e. those that did not interpret that comparison (n = 5) or for which the interpretation was ambiguous (n = 2). Among the remaining studies, 52.9% (n = 9) interpreted repeatability differences between groups as being caused by changes in both within- and among-individual level, whereas 35.3% (n = 6) and 11.8% (n = 2) interpreted repeatability differences between groups as being caused by changes only at the within-individual or only at the among-individual level, respectively. These results contrasted with the fact that comparing repeatability estimates between two or more groups is not biologically informative for the level of biological variation that is causing the difference in repeatability. To fully understand the biology underpinning differences in repeatability estimates would require that both within- and among-individual level variances are provided alongside the repeatability estimates (see Discussion, and Dochtermann and Royauté, 2019). Finally, our literature review showed that only 30.7% (n = 27) of articles calculating repeatability and 47.1% (n = 8) of articles interpreting the repeatability estimates between groups reported the unstandardized estimates for both with- and among-individual variances.

3.2. Self-report questionnaire

A total of 440 participants took part in our self-report questionnaire. Among them, 6.1% (n = 27) identified themselves as Master students, 28.6% (n = 126) as PhD researchers, 24.1% (n = 106) as early-career researchers (defined as having completed their doctorate degree within the past 5 years), 38.9% (n = 171) as senior researchers (defined as having completed their doctorate degree more than 5 years ago), and 0.02% (n = 10) did not identify as researchers and were therefore excluded from all subsequent analyses. From the remaining 430 participants, 57.9% (n = 249) reported to have been the leading, corresponding and/or senior author in at least one publication related to animal personality, whereas 14.7% (n = 63) reported to have been co-authors in such publications, and 25.8 % (n = 111) and 0.02 % (n = 7), respectively, reported to have not published on questions related to animal personality or being unsure about it. We present the results of the self-reported questionnaire based on all 430 participants who self-identified as researchers, but see Appendix S2 for the results based on only those researchers who reported to have (co)authored at least one publication related to animal personality (n = 312). Our questionnaire reached a wide geographic distribution in terms of self-reported country of affiliation (n = 38 countries, Figure 1), with USA (n = 73), Germany (n = 64), United Kingdom (n = 53), and Spain (n = 46) together accounting for 54.9% of all participations (Figure S4).

Figure 1. World map illustrating the self-reported affiliations of all 430 participants recorded in a questionnaire about animal personality terminology. Map generated using EviAtlas (Haddaway et al., 2019).

3.2.1. Personality definition and interpretation

Our questionnaire showed that the preferred definition of animal personality was "*Consistent between-individual differences in behaviour across time and/or contexts*" (43.5%, n = 187) with the other four suggested definitions being selected between 4.9% and 17.7% of the times (Table 1). Furthermore, the preferred interpretation of what animal personality represents biologically was "*Individual differences in average behavioural expression in a sample of individuals*" (46.7%, n = 201) closely followed by a combination of this interpretation and "*limited phenotypic plasticity in behavioural expression in a sample of individuals*" (38.6%, n = 166; Table 2).

Table 1. List of animal personality definitions provided and selected in a self-report questionnaire about animal personality terminology.

Animal personality definition	Number of times selected in the questionnaire (%)
Consistent between-individual differences in behaviour across time and/or contexts	187 (43.5%)
Within-individual and between-individual consistency in behaviours across time and/or ecological contexts	76 (17.7%)
Between-individual differences in behavioural tendencies across contexts and within-individual consistency over time	68 (15.8%)
Consistent between-individual differences in whole suites of correlated behaviours across time and/or contexts	66 (15.3%)
Variation among individuals in the intercept of their behavioural reaction norm	21 (4.9%)
Others (see Table S1)	12 (2.8%)

Table 2. List of biological interpretations of animal personality provided and selected in a self-report questionnaire about animal personality terminology.

Biological interpretation of animal personality	Number of times selected in the questionnaire (%)
a) Limited phenotypic plasticity in behavioural expression in a sample of individuals	27 (6.3%)
b) Individual differences in average behavioural expression in a sample of individuals	201 (46.7%)
c) Both a) and b) are correct	166 (38.6%)
d) I do not know the answer	18 (4.2%)
e) Others (see Table S2)	18 (4.2%)

3.2.2. Repeatability definition and interpretation

Most participants interpreted estimates of repeatability as the "*(Relative) amount of individual differences in average trait expression in a sample of individuals*" (43.3%, n = 186;

Table 3). However, 38.1% (n = 164) of participants also considered the "*(Relative) amount of phenotypic plasticity in trait expression in a sample of individuals*" either as the single interpretation of behavioural repeatability (16.5%, n = 71) or in combination with the former interpretation (21.6%, n = 93; Table 3). Additionally, 69.1% of the participants (n = 297) reported that repeatability provides an estimate of individual consistency or individual predictability, while 19.5% (n = 84) considered that it does not. 11.4% (n = 49) reported not to know the answer to this question.

Furthermore, when asked about the biological interpretation of one group of individuals (Group A) expressing higher repeatability than other group of individuals (Group B), all suggested choices obtained a rather similar share of preference (range = 17.7% to 33.3%, with 5.8% reporting not to know the answer; Table 4). The preferred choice was nevertheless that no interpretation can be drawn about whether the two groups differ in relative plasticity or in the relative amount of individual differences in average trait expression (33.3%, n = 143; Table 4).

Table 3. List of repeatability interpretations provided and selected in a self-report questionnaire about animal personality terminology.

Biological interpretation of repeatability	Number of times selected in the questionnaire (%)
a) (Relative) amount of phenotypic plasticity in trait expression in a sample of individuals	71 (16.5%)
b) (Relative) amount of individual differences in average trait expression in a sample of individuals	186 (43.3%)
c) Both a) and b) are correct	93 (21.6%)
d) I do not know the answer	39 (9.1%)
e) Others (see Table S3)	41 (9.5%)

Table 4. List of interpretations provided and selected for a comparison of repeatability estimates between two groups in a self-report questionnaire about animal personality terminology. Specifically, the question from the questionnaire was: "*What biological interpretation can one make when one only knows that a group of individuals "A" expresses higher repeatability than a group of individuals "B"?*"

Biological interpretation of comparing repeatability estimates of two groups of animals	**Number of times selected in the questionnaire (%)**
a) Group "A" expresses less plasticity in their trait expression than group "B"	103 (24%)
b) Individuals differ more from each other in their average trait expression in group "A" than in group "B"	83 (19.3%)
c) Both interpretations a) and b) can be made	76 (17.7%)
d) Neither interpretations a) nor b) can be made	143 (33.3%)
e) I do not know the answer	25 (5.8%)

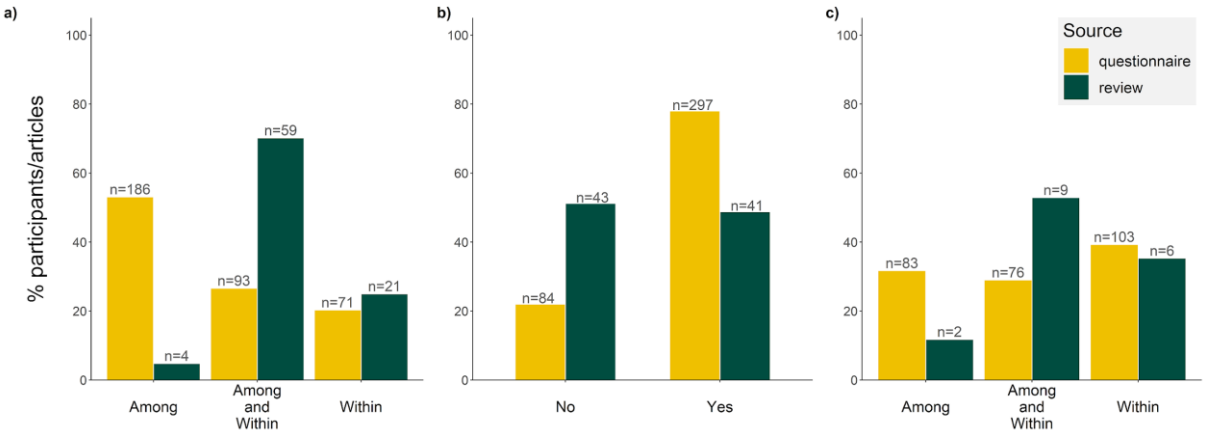

Figure 2. Comparison between the results of a literature review and a self-report questionnaire aimed at understanding how repeatability was interpreted by researchers. Panel a) shows the percentage of participants and articles that interpreted repeatability at each of the three levels of variation. Panel b) shows the percentage of participants and articles

that interpreted (or not) repeatability as individual consistency and/or individual predictability. Panel c) shows the percentage of participants and articles that interpreted differences in repeatability estimates between groups at each of the three levels of variation. Labels “Among” and “Within” refer to the among- and within-individual levels of variation, respectively. Contrary to how results were presented in the tables above, the percentages in this figure were calculated after having excluded NA’s and the questionnaire’s categories “Others” and “I do not know the answer” to facilitate the comparison between the results of the literature review and the questionnaire.

4. Discussion

The magnitude of individual differences in the average behavioural expression — or any other labile phenotypic trait — is captured by the among-individual variance component in a statistical (mixed-effects) model. From a statistical point of view, the concept of “animal personality” leaves no room for disagreement: animal personality is captured by a statistically significant among-individual variance component in a repeated measured data (Dingemanse and Dochtermann, 2013). However, the use of the statistical definition of animal personality is not ubiquitous among researchers, and numerous conceptual definitions of the biological meaning of animal personality are available in the literature. By combining a literature review and a self-report questionnaire, we observed discrepancies among researchers in how concepts commonly used in the field of animal personality are defined and interpreted. Particularly, our study showed that statistical and conceptual definitions did not always agree, and therefore that either conceptual definitions are sometimes incorrect or that the statistical approaches used to test the conceptual definitions are sometimes inappropriate.

Questionnaire and comparison with the literature review

While the self-report questionnaire showed that "*Consistent between-individual differences in behaviour across time and/or contexts*" was the preferred conceptual definition of animal personality, the questionnaire also identified a lack of consensus on the preferred biological interpretation of animal personality. About half of the participants (47%) interpreted animal personality as "*Individual differences in average behavioural expression in a sample of individuals*", and an additional 39% considered that interpretation to be correct but only when also considering "*Limited phenotypic plasticity in behavioural expression*" (Table 2). These results evidence that, although most participants interpreted animal personality at least partially at the among-individual level (85%), about half of them (45%) interpreted it also at the within-individual level (i.e., 45% interpreted animal personality also as phenotypic flexibility or reversible plasticity), highlighting the lack of general agreement among the questionnaire participants on what animal personality means biologically.

Participants did not agree on the biological interpretation of repeatability either. The questionnaire showed that repeatability was most often interpreted at the among-individual level either by itself (43%) or in combination with the within-individual level (22%). In relative terms, these results are similar to those about the biological interpretation of animal personality, although repeatability was interpreted at the within-individual level only relatively more often (17%, compared to 6% for the definition of animal personality; Tables 2 and 3). Interestingly, we also found discrepancies between the literature review and the questionnaire. While evidence from the literature review and the questionnaire showed similar percentages in terms of interpreting repeatability at the within-individual level, there were substantial differences at the other levels (i.e., at among-individual level and both levels combined; Figure 2). Hence, participants reported an interpretation of repeatability that was

not reflected in the published literature. This disagreement is in line with one of the main premises of our study: the use of conceptual, often vague, terminology might obscure the interpretation of the biological phenomena. Finally, we interpret the discrepancy between the literature review and the questionnaire as true differences, however we cannot discard that the observed differences are, in part, due to different subsets of researchers. While plausible, we deem this option less likely because of the large coverage of our questionnaire and the high percentage of researchers that reported to have worked in animal personality (73%).

Another important insight from our questionnaire is that it showed differences among participants regarding how to interpret the comparison of repeatability estimates between groups (Table 4). Although the preferred (and correct) answer was that no biologically-meaningful interpretation can be made about whether the two groups differ in amount of plasticity or in amount of individual differences in average trait expression (33%), the other options still obtained substantial support (range = 18% to 24%; Table 4). These differences among researchers were similar to those observed in the published literature. Of the 24 studies that compared repeatability estimates between groups, 17 interpreted the difference in repeatability estimates, with 53% interpreting repeatability differences between groups as being caused by changes in both within- and among-individual variation, whereas 35% and 12% interpreted repeatability differences between groups as being caused by changes only at the within-individual or at the among-individual variation, respectively. Overall, comparing and interpreting differences in repeatability estimates between groups is problematic because repeatability estimates can differ due to change in variance at the among-individual level, within-individual level or both (Dochtermann et al., 2015; Dochtermann and Royauté, 2019; Hansen et al., 2011; Houle, 1992; Niemelä and Dingemanse, 2018b; Wilson, 2018). For example, if repeatability is higher in one group than the other, it cannot be stated that among-

individual variation, i.e., expression of "animal personality", would be higher or within-individual variation, i.e. reversible plasticity, lower in that group. The same logic stands for making biological interpretations between groups *not* differing in their repeatability estimates. Our results from the literature review and the questionnaire showed that this is a common pitfall among empiricists working in animal personality research.

To determine the biological mechanism underpinning differences in repeatability between groups, researchers would need to know both the within- and among-individual variance components from which repeatability was calculated (more in Dochtermann and Royauté, 2019). However, comparing variance components across groups makes sense only within studies since researchers would need to have measured exactly the same traits for each group in a comparable manner so that the comparisons of variances are meaningful. This requirement leads to the very reason why repeatability is estimated in the first place, i.e. to allow the comparison of standardized variance components between data sets, with the cost of losing ability to make accurate biological interpretations (see the discussion above). One straightforward solution to avoid this problem would be to apply a mean standardization (e.g. coefficient of variation) rather than a variance standardization (e.g. repeatability) to variance components (Dochtermann and Royauté, 2019; Hill and Mulder, 2010; Holtmann et al., 2017; Houle, 1992). Mean-standardizing variance components would allow to estimate variance per unit of a trait value for each variance component (i.e., among and within-individual variance) separately, partially solving the abovementioned problem.

Finally, our questionnaire revealed that the vast majority of researchers (78%) considered repeatability to be a metric of individual consistency and/or individual predictability. This finding largely agreed with our literature review, which showed that repeatability was interpreted as individual consistency and/or predictability in approximately

half of the studies. If we combine these results with those of the biological meaning of repeatability in our questionnaire, it suggests that researchers generally consider individual differences in the average behavioural expression (i.e. animal personality) as "consistency". However, consistency is a very widespread term in the field of animal personality and used to refer to several biological phenomena such as individual differences in average trait expression and within-individual plasticity within- and among-environments. Cleasby et al. (2015) considered consistency (and predictability) as a within-individual level metric, requiring more sophisticated statistical tools (i.e. double-hierarchical mixed effects models) to estimate it compared to the simple repeatability estimate. Nevertheless, as explained above, repeatability is the proportion of total phenotypic variance explained by the among-individual variance component, and as such, both among and within-individual variance affects repeatability estimate. Thus, regardless of whether consistency is defined as among- or within-individual variation, it affects repeatability estimate. The problem of using the term consistency is that its exact meaning remains ambiguous unless explicitly stated in the study. Thus, we encourage researchers to clearly define consistency and predictability whenever they use it in their future work.

Potential reasons for diversity of interpretations and what can be done to unify the field?

First, we need to acknowledge that animal personality research is divided into researchers that apply a "behavioural ecologist" framework (i.e. among-individual variation in a behavioural trait) and researchers that apply a "personality psychologist" framework (i.e. correlation between multiple traits: termed as "behavioural syndrome" in behavioural ecology (Dingemanse et al., 2010; Dingemanse and Wright, 2020; Gosling, 2001; Réale et al., 2007; Sih et al., 2004). This dichotomy generates basal differences in how the concept of animal personality is being studied and discussed, and what statistical tools are required (e.g.

univariate versus multivariate mixed effects models). Another potential reason for the large diversity of terminology is that there is no consensus in the concept of animal personality among the most-cited conceptual articles (see extracted terminology in the provided data). Some of these conceptual articles, which were pioneers and became important cornerstones for the entire field, described animal personality as both high among-individual variation and low within-individual variation. Indeed, "limited plasticity" was often mentioned as a key component of animal personality (see extracted terminology in the provided data), even though it is arguable whether intrinsically limited plasticity is required for the existence of among-individual differences in average trait expression. Individuals might differ in their average trait expression even when all of them can express the whole spectrum of behavioural variation.

Importantly, we do not aim to yet create another definition of animal personality, but to advocate for the usage of a common definition, which ultimately would help to unify and clarify the field. In the light of our findings, we strongly encourage researchers to avoid the use of vague conceptual definitions that can be interpreted at multiple levels of biological organization. Our suggestion is to use more statistically-oriented terminology that directly express which level of variation any given study focuses on: among- or within-individual level variation. This would leave less room for misinterpretations of the used terminology. We expect that adopting such a strategy will, at least partly, avoid the confusion generated by the buzzword "animal personality".

Data requirements in animal personality research

Generally, our questionnaire evidenced that researchers are quite aware of the need of recording multiple observations (e.g. behavioural measurements) per individual to study

among-individual variance in behaviour. However, there was still a relatively large portion of researchers who were not in agreement with this requirement and/or who were not applying it. Indeed, 31% of researchers who self-reported having published in the field of animal personality also reported having used single measurements of behaviour(s) per individual (i.e. non-repeated measurements per individual). Furthermore, 23% of researchers answered that although repeated measurements per individual are necessary to study among-individual differences, single behavioural measurements per individual can be used as long as the focal behaviour is repeatable. However, this argument does not hold since single measurements can be used to study among-individual level phenomena only when repeatability of focal behaviour is 1; in such a case, behaviour is a fixed trait (Niemelä and Dingemanse 2018). While there are exemptions for the use of single behavioural observations per individual such as in quantitative genetics studies including pedigree data, most animal personality studies do not have data to provide estimates at the additive genetic level, and hence, require the use of repeated measures. Moreover, when using a quantitative genetics approach with a combination of single measurements to study, for example, whether animal personality is associated with other traits, one makes the critical assumption that permanent environment effects are negligible or in the same direction as additive genetic effects. The need for repeated measures per individual has been previously, and repeatedly, explained in detail (Brommer, 2013; Dingemanse et al., 2012; Niemelä and Dingemanse, 2018b). However, our questionnaire showed that a non-negligible percentage of researchers are still unaware of (or not fulfilling) this requirement. We hope our work encourages the researchers to pay further attention to the data requirements for studies of animal personality (more in Dingemanse and Dochtermann, 2013).

Limitations of this study

A major perceived limitation of our study may be related to the set of choices given in the questionnaire. In the comments section of our questionnaire, 3.7% (16 out of 430) of participants were critical about the wording of one or more questions (see provided data). While we believe this is a general concern in any self-report questionnaire, it might have been especially problematic in the context of our study since the field of animal personality is known for using multiple definitions, interpretations and approaches to address the same research topic, as pointed out in this work. With this issue in mind, we had designed the questions using a combination of statistical terms adapted to their biological meaning, and actual definitions present in the published literature. Furthermore, prior to publishing the questionnaire, we asked several experts in the field of animal personality to provide suggestions on how to improve our questionnaire. Despite our efforts in designing the most accommodating questionnaire, we cannot disregard that some of the differences in answers among participants might have been in part generated by participants understanding our questions differently.

Conclusions

If total phenotypic level estimates can be used to study among-individual differences in behaviour, then every single animal behaviour study ever published could be labelled as "animal personality" research. Most researchers working in animal personality, however, do aim to understand among-individual level processes (rather than total phenotypic level processes), but often do not test them at the appropriate level of variation (Moiron et al., 2020; Niemelä and Dingemanse, 2018b, 2018a) and/or communicate their results in an unambiguous way (our results). We strongly encourage animal personality empiricists to

clearly express which level of variation their focal work focuses on, and accordingly, collect suitable data and use suitable statistical frameworks (see Dingemanse and Dochtermann 2013 as a great educational tool). We hope that our study will raise awareness about the ambiguous terminology used in animal personality research, helping the entire field to move towards a more unified research framework.

Data and code availability section

Data and code to reproduce the results are available from: https://github.com/ASanchez-Tojar/animal_personality_terminology. Upon eventual acceptance of this manuscript, data and code will be hosted at Zenodo and provided with a doi.

Author contribution

AST, MM and PN contributed to conceptualization, methodology, investigation, writing - both original draft preparation and review and editing, project administration, and funding acquisition. AST also contributed to software, formal analysis, data curation and visualization. MM also contributed to validation. All authors contributed equally to this work.

Acknowledgments

We are grateful to Niels Dingemanse, Raphaël Royauté, Stephen Salazar and Alastair J. Wilson for their comments on a previous version of the questionnaire, and to everybody that took part in the questionnaire and/or helped to disseminate it. We are also grateful to Alastair J. Wilson for insightful discussions about some of the concepts presented in our study, and to

Niels Dingemanse, Marie I. Kaiser and Holger Schielzeth for their comments on a previous version of this manuscript.

Funding

AST was funded by the German Research Foundation (DFG: Deutsche Forschungsgemeinschaft) as part of the SFB TRR 212 (NC3)—Project numbers 316099922 and 396782608. MM was funded by a Marie Curie Individual Fellowship (PLASTIC TERN, Grant Agreement Number 793550). PTN was funded by the German Research Foundation (DFG; NI 1539/2-1).

Competing interests

Authors declare no competing interests.

References

Brommer, J.E., 2013. On between-individual and residual (co)variances in the study of animal personality: are you willing to take the "individual gambit"? Behav Ecol Sociobiol 67, 1027–1032. https://doi.org/10.1007/s00265-013-1527-4

Carter, A.J., Feeney, W.E., Marshall, H.H., Cowlishaw, G., Heinsohn, R., 2013. Animal personality: what are behavioural ecologists measuring? Biological Reviews 88, 465–475. https://doi.org/10.1111/brv.12007

Cleasby, I.R., Nakagawa, S., Schielzeth, H., 2015. Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. Methods in Ecology and Evolution 6, 27–37. https://doi.org/10.1111/2041-210X.12281

Coppens, C.M., de Boer, S.F., Koolhaas, J.M., 2010. Coping styles and behavioural flexibility: towards underlying mechanisms. Philosophical Transactions of the Royal Society B: Biological Sciences 365, 4021–4028. https://doi.org/10.1098/rstb.2010.0217

Dall, S.R.X., Houston, A.I., McNamara, J.M., 2004. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. Ecology Letters 7, 734–739. https://doi.org/10.1111/j.1461-0248.2004.00618.x
David, M., Dall, S.R.X., 2016. Unravelling the Philosophies Underlying 'Animal Personality' Studies: A Brief Re-Appraisal of the Field. Ethology 122, 1–9. https://doi.org/10.1111/eth.12445
Dingemanse, N.J., Araya-Ajoy, Y.G., 2015. Interacting personalities: behavioural ecology meets quantitative genetics. Trends in Ecology & Evolution 30, 88–97. https://doi.org/10.1016/j.tree.2014.12.002
Dingemanse, N.J., Dochtermann, N.A., 2013. Quantifying individual variation in behaviour: mixed-effect modelling approaches. Journal of Animal Ecology 82, 39–54. https://doi.org/10.1111/1365-2656.12013
Dingemanse, N.J., Dochtermann, N.A., Nakagawa, S., 2012. Defining behavioural syndromes and the role of 'syndrome deviation' in understanding their evolution. Behav Ecol Sociobiol 66, 1543–1548. https://doi.org/10.1007/s00265-012-1416-2
Dingemanse, N.J., Kazem, A.J.N., Réale, D., Wright, J., 2010. Behavioural reaction norms: animal personality meets individual plasticity. Trends in Ecology & Evolution 25, 81–89. https://doi.org/10.1016/j.tree.2009.07.013
Dingemanse, N.J., Réale, D., 2005. Natural selection and animal personality. Behaviour 142, 1159–1184. https://doi.org/10.1163/156853905774539445
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. https://doi.org/10.1098/rstb.2010.0221
Dingemanse, N.J., Wright, J., 2020. Criteria for acceptable studies of animal personality and behavioural syndromes. Ethology 126, 865–869. https://doi.org/10.1111/eth.13082
Dochtermann, N.A., Royauté, R., 2019. The mean matters: going beyond repeatability to interpret behavioural variation. Animal Behaviour 153, 147–150. https://doi.org/10.1016/j.anbehav.2019.05.012
Dochtermann, N.A., Schwab, T., Sih, A., 2015. The contribution of additive genetic variation to personality variation: heritability of personality. Proceedings of the Royal Society B: Biological Sciences 282, 20142201. https://doi.org/10.1098/rspb.2014.2201
Duckworth, R.A., Badyaev, A.V., 2007. Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird. PNAS 104, 15017–15022. https://doi.org/10.1073/pnas.0706174104
Forsman, A., 2015. Rethinking phenotypic plasticity and its consequences for individuals, populations and species. Heredity 115, 276–284. https://doi.org/10.1038/hdy.2014.92
Gosling, S.D., 2001. From mice to men: What can we learn about personality from animal research? - PsycNET. Psychological Bulletin 127, 45–86. https://doi.org/10.1037/0033-2909.127.1.45
Haddaway, N.R., Feierman, A., Grainger, M.J., Gray, C.T., Tanriver-Ayder, E., Dhaubanjar, S., Westgate, M.J., 2019. EviAtlas: a tool for visualising evidence synthesis databases. Environmental Evidence 8, 22. https://doi.org/10.1186/s13750-019-0167-1
Hansen, T.F., Pélabon, C., Houle, D., 2011. Heritability is not Evolvability. Evol Biol 38, 258. https://doi.org/10.1007/s11692-011-9127-6
Hill, W.G., Mulder, H.A., 2010. Genetic analysis of environmental variation. Genetics Research 92, 381–395. https://doi.org/10.1017/S0016672310000546

Holtmann, B., Lagisz, M., Nakagawa, S., 2017. Metabolic rates, and not hormone levels, are a likely mediator of between-individual differences in behaviour: a meta-analysis. Functional Ecology 31, 685–696. https://doi.org/10.1111/1365-2435.12779

Houle, D., 1992. Comparing evolvability and variability of quantitative traits. Genetics 130, 195–204.

Kaiser, M.I., Müller, C., 2021. What is an animal personality? Biology & Philosophy 36, 1. https://doi.org/10.1007/s10539-020-09776-w

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group, T.P., 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Medicine 6, e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moiron, M., Laskowski, K.L., Niemelä, P.T., 2020. Individual differences in behaviour explain variation in survival: a meta-analysis. Ecology Letters 23, 399–408. https://doi.org/10.1111/ele.13438

Nakagawa, S., Schielzeth, H., 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biological Reviews 85, 935–956. https://doi.org/10.1111/j.1469-185X.2010.00141.x

Niemelä, P.T., Dingemanse, N.J., 2018a. Meta-analysis reveals weak associations between intrinsic state and personality. Proceedings of the Royal Society B: Biological Sciences 285, 20172823. https://doi.org/10.1098/rspb.2017.2823

Niemelä, P.T., Dingemanse, N.J., 2018b. On the usage of single measurements in behavioural ecology research on individual differences. Animal Behaviour 145, 99–105. https://doi.org/10.1016/j.anbehav.2018.09.012

Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan—a web and mobile app for systematic reviews. Systematic Reviews 5, 210. https://doi.org/10.1186/s13643-016-0384-4

Rands, S.A., Cowlishaw, G., Pettifor, R.A., Rowcliffe, J.M., Johnstone, R.A., 2003. Spontaneous emergence of leaders and followers in foraging pairs. Nature 423, 432–434. https://doi.org/10.1038/nature01630

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. https://doi.org/10.1111/j.1469-185X.2007.00010.x

Roche, D.G., Careau, V., Binning, S.A., 2016. Demystifying animal 'personality' (or not): why individual variation matters to experimental biologists. Journal of Experimental Biology 219, 3832–3843. https://doi.org/10.1242/jeb.146712

Royauté, R., Berdal, M.A., Garrison, C.R., Dochtermann, N.A., 2018. Paceless life? A meta-analysis of the pace-of-life syndrome hypothesis. Behav Ecol Sociobiol 72, 64. https://doi.org/10.1007/s00265-018-2472-z

Scheiner, S.M., 1993. Genetics and Evolution of Phenotypic Plasticity. Annual Review of Ecology and Systematics 24, 35–68. https://doi.org/10.1146/annurev.es.24.110193.000343

Schuett, W., Tregenza, T., Dall, S.R.X., 2010. Sexual selection and animal personality. Biological Reviews 85, 217–246. https://doi.org/10.1111/j.1469-185X.2009.00101.x

Sih, A., Bell, A.M., Johnson, J.C., Ziemba, R.E., 2004. Behavioral Syndromes: An Integrative Overview. The Quarterly Review of Biology 79, 241–277. https://doi.org/10.1086/422893

Sih, A., Mathot, K.J., Moirón, M., Montiglio, P.-O., Wolf, M., Dingemanse, N.J., 2015. Animal personality and state–behaviour feedbacks: a review and guide for empiricists. Trends in Ecology & Evolution 30, 50–60. https://doi.org/10.1016/j.tree.2014.11.004

Sih, A., Munson, A., Pollack, L., 2020. Animal Personalities, in: The Wiley Encyclopedia of Personality and Individual Differences. John Wiley & Sons, Ltd, pp. 117–122. https://doi.org/10.1002/9781119547143.ch21

Stamps, J., Groothuis, T.G.G., 2010. The development of animal personality: relevance, concepts and perspectives. Biological Reviews 85, 301–325. https://doi.org/10.1111/j.1469-185X.2009.00103.x

van de Pol, M., Wright, J., 2009. A simple method for distinguishing within- versus between-subject effects using mixed models. Animal Behaviour 77, 753–758. https://doi.org/10.1016/j.anbehav.2008.11.006

Westneat, D.F., Wright, J., Dingemanse, N.J., 2015. The biology hidden inside residual within-individual phenotypic variation. Biological Reviews 90, 729–743. https://doi.org/10.1111/brv.12131

Wilson, A.J., 2018. How should we interpret estimates of individual repeatability? Evolution Letters 2, 4–8. https://doi.org/10.1002/evl3.40

Wolak, M.E., Fairbairn, D.J., Paulsen, Y.R., 2012. Guidelines for estimating repeatability. Methods in Ecology and Evolution 3, 129–137. https://doi.org/10.1111/j.2041-210X.2011.00125.x

Wolf, M., Weissing, F.J., 2012. Animal personalities: consequences for ecology and evolution. Trends in Ecology & Evolution 27, 452–461. https://doi.org/10.1016/j.tree.2012.05.001