Animal Behaviour 145 (2018) 99–105

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Commentary

On the usage of single measurements in behavioural ecology research on individual differences

Petri T. Niemelä*, Niels J. Dingemanse

Behavioural Ecology, Department of Biology, Ludwig-Maximilians University of Munich, Planegg-Martinsried, Germany

ARTICLE INFO

Article history:
Received 26 May 2018
Initial acceptance 9 July 2018
Final acceptance 7 September 2018

MS. number: 18-00345R

Keywords:
among-individual variation
animal behaviour
animal personality
correlation
individual gambit
repeatability

Since the early 21st century, behavioural ecologists have increasingly focused on the ecology and evolution of repeatable individual differences in behaviour (aka 'animal personality'). More and more studies have investigated correlations between behaviour(s) and other phenotypic traits, thereby seeking to understand among-individual variation in behaviour from either a proximate or an ultimate perspective. Statistically, such studies require the estimation of among-individual correlations, necessitating study designs where suites of labile traits are repeatedly assayed. Most published studies, by contrast, instead assume that among-individual correlations can be approximated using data of suites of phenotypic traits measured only once. Such studies take an 'individual gambit' as they assume that phenotypic correlations between two traits measured once match among-individual correlations. A literature survey shows that this assumption was made in 62% of empirical studies; this is a worrying trend as a mismatch between the research question and the study (or statistical) design can lead to biased conclusions about biological patterns. In this paper we use a visual approach to illustrate these concerns for a broad audience, thereby complementing previous papers using a jargon more suitable for the statistically oriented. We thereby seek to reiterate the notion that reliable answers to any scientific question require specific types of data. While this commentary underlines the importance of spelling out key assumptions in discussions of (suboptimal) data in scientific papers on individuality, our main message is that researchers should typically be able to avoid publishing studies with conclusions heavily hinging on unverifiable assumptions. This is because researchers normally have the freedom to decide a priori to focus their research towards questions for which appropriate data can arguably be collected. Doing so would greatly facilitate our ability to robustly address key questions of biological interest, such as the ecology and evolution of individuality.

The purpose of this paper will have been accomplished…if it prevents the future computation of meaningless correlations and stimulates the study of similar problems with the use of meaningful correlations between the properties of individuals. **W.S. Robinson (1950, p. 357)**

Behavioural ecologists are currently highly interested in studying the adaptive integration between behaviour and other labile traits, seeking to address questions related to the evolutionary emergence of individual differences in behaviour (i.e. 'personality'; Biro & Stamps, 2010; Careau, Thomas, Humphries, & Réale, 2008; Dammhahn, Dingemanse, Niemelä, & Réale, 2018; Niemelä & Dingemanse, 2018; Réale et al., 2010). Such research questions typically require repeated measures as this allows the statistical partitioning of correlations between traits into their among- and within-individual components. Only in this way can researchers acquire insight into the correlation between the average trait values among individuals (called behavioural syndromes when applied to two behaviours) versus how changes in traits are correlated within individuals across subsequent measurements (plasticity integration; Dingemanse & Dochtermann, 2013; Dingemanse, Dochtermann, & Nakagawa, 2012; Dochtermann & Dingemanse, 2013). For example, among-individual correlations are present when individuals that on average, over all repeated observations of their behavioural phenotype (e.g. activity), also express, on average, higher values of their physiology (e.g. metabolic rates). By contrast, within-individual correlations occur when changes in one labile trait (e.g. activity) from one moment to the next are correlated with changes in another (e.g. hormone levels) within the same individual over time. The common practice is that researchers typically

* Correspondence: P.T. Niemelä, Behavioural Ecology, Department of Biology, Ludwig-Maximilians University of Munich, Planegg-Martinsried, Germany.
E-mail address: petri.t.niemela@gmail.com (P. T. Niemelä).

https://doi.org/10.1016/j.anbehav.2018.09.012
0003-3472/

measure correlations between labile traits at the unpartitioned phenotypic level (Niemelä & Dingemanse, 2018), but interpret these as among-individual correlations, thereby taking an 'individual gambit' (Brommer, 2013; Dingemanse et al., 2012).

Studies taking this 'individual gambit' illustrate a well-known problem where there is a mismatch between the information embedded in particular data and how these data are interpreted (Goodman, 1953; Robinson, 1950). The specific concern with using unpartitioned phenotypic correlations to address questions regarding individual level associations has been raised before, particularly in papers making the argument statistically (Brommer, 2013; Dingemanse & Dochtermann, 2013; Dingemanse et al., 2012). Interestingly, there is no quantitative information on how widely this suboptimal approach is applied in studies seeking to study among-individual level patterns. The purpose of this paper is to (1) estimate how commonly behavioural studies apply the 'individual gambit', (2) clarify the problems related to taking the 'individual gambit' visually, and thereby complement previous statistical treatments, (3) critically evaluate various key assumptions made when taking the 'individual gambit' and (4) provide general guidelines on how to better the match between scientific questions and study designs in the study of behaviour.

HOW OFTEN IS THE 'INDIVIDUAL GAMBIT' APPLIED?

The occurrence of studies taking the 'individual gambit' was estimated using data from a recently published meta-analysis (Niemelä & Dingemanse, 2018). This meta-analysis contained published correlations between behaviours and other phenotypic traits (e.g. body condition, hormones, metabolism) often predicted to be integrated among individuals (Niemelä & Dingemanse, 2018). As part of our data collection, papers were categorized into those making individual level inferences about correlations between behaviours and other phenotypic traits based on designs that either (1) measured at least one of the two focal phenotypic traits once (i.e. studies lacking repeated measurements), (2) took forward single measurements of each phenotypic trait to the statistical analysis of correlations despite collecting repeated measures, or (3) collected repeated measures data from all involved traits (i.e. took the first critical step in estimating level-specific correlations). For search terms, see the Supplementary Material in Niemelä and Dingemanse (2018). The initial sample size was 1086 studies, which was reduced to 145 studies making individual level interpretations following data screening based on the PRISMA protocol (Niemelä & Dingemanse, 2018). Of these studies, 52% (75/145) measured at least one of the traits only once and 10% (15/145) based their statistical analysis on a single measurement per trait (despite having repeated measurements). Thus, the majority of studies (62%) applied the 'individual gambit' (Fig. 1). Our reading of these papers suggested that most of them drew conclusions on individuality without mentioning key, unverified, assumptions (see section Critical assumptions when applying the 'individual gambit').

The main problem in applying the 'individual gambit' is that the researcher is inherently blind to whether the phenotypic correlations match among-individual correlations. Depending on whether within- and among-individual correlations are the same (e.g. both positive or both negative) versus very different (e.g. one positive, the other negative), phenotypic correlations may or may not match among-individual correlations in either sign or magnitude (Figs. 2–4).

HOW DO EMPIRICISTS JUSTIFY THE APPLICATION OF THE 'INDIVIDUAL GAMBIT'?

Often empiricists justify the 'individual gambit' by stating that the focal behavioural trait is repeatable (either by using repeatability from previous studies or by estimating repeatability using a subsample of their data). Repeatability (R) is calculated by dividing the among-individual variance by the total phenotypic variance (Dingemanse & Dochtermann, 2013; Falconer & Mackay, 1996; Lessells & Boag, 1987; Lynch & Walsh, 1998). R provides information about the degree of individual differences in phenotypic mean in the population (Dingemanse & Dochtermann, 2013; Lessells & Boag, 1987), which implies that the trait might harbour nonzero additive genetic variance (Boake, 1989; Falconer & Mackay, 1996; Lynch & Walsh, 1998). Researchers justifying the usage of phenotypic correlations between two traits measured once based on

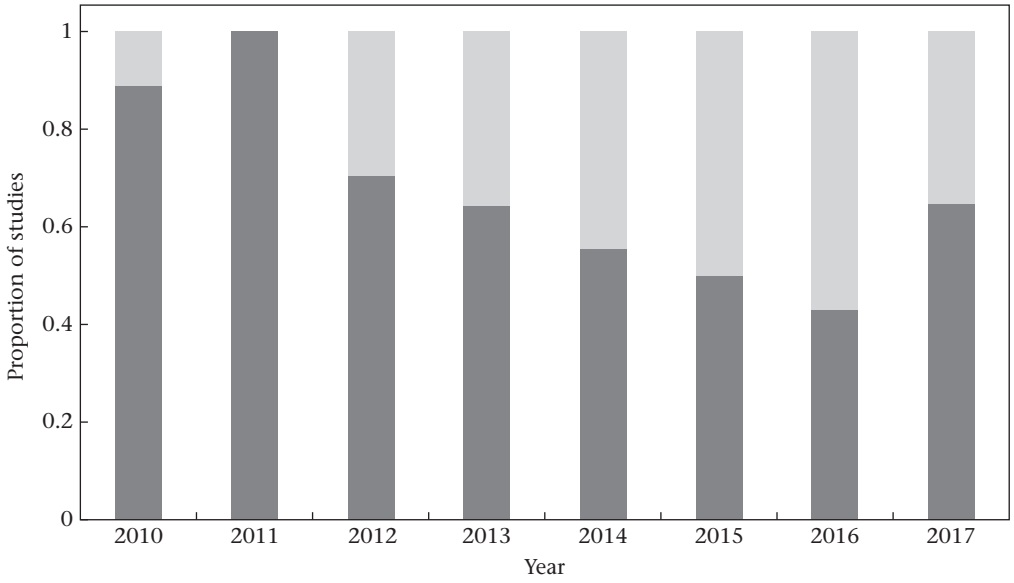

Figure 1. The proportion of papers (published between 2010 and 2017) quantifying among-individual correlations based on single values of at least one of these traits (dark grey). Light grey bars indicate the proportion of papers measuring each trait repeatedly. Sample sizes are 2010= 9 (studies), 2011=10, 2012=17, 2013=14, 2014=18, 2015=22, 2016=28 and 2017=17. An additional 10 studies spanned multiple years (1996–2009).

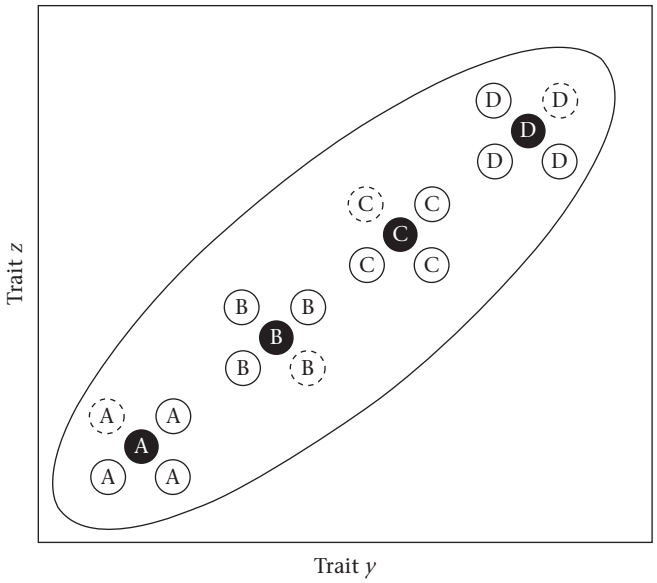

Figure 2. Illustration of a scenario where four individuals (letters A, B, C, D) are assayed five times (circles) for each of two traits (y, z) simultaneously, both of which are repeatable (R>0). Circles represent observations; black circles represent each individual's average trait value. The scenario shown here is one where within the same individual the two traits are not correlated. Among individuals, by contrast, the two traits are correlated. In fact, the among-individual correlation here is equal to 1, whereas the phenotypic correlation (ellipse) is clearly lower than 1. This demonstrates the scenario where a zero within-individual correlation causes an attenuated estimate of the among-individual correlation when based on the phenotypic correlation (Adolph & Hardin, 2007). If we had just collected one data point per individual (broken circle), the among-individual correlation among these four single values per individual would actually misrepresent the among-individual correlation.

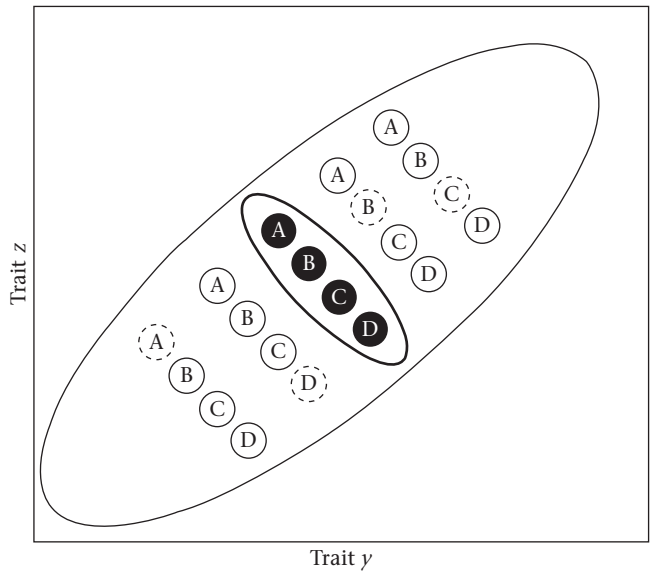

Figure 3. Illustration of a scenario where four individuals (letters A, B, C, D) are assayed five times (circles) for each of two traits (y, z) simultaneously, both of which are repeatable (R>0). Circles represent observations; black circles represent each individual's average trait value. The scenario shown here is one where within-individual changes in trait y covary positively with the within-individual changes in trait z, causing a strongly positive within-individual correlation. By contrast, the among-individual correlation is negative, as the mean values (black circles) are negatively correlated. Again, the phenotypic correlation (nonbolded ellipse) does not represent the among-individual correlation (bolded ellipse). If we had just collected one data point per individual (broken circles), the among-individual correlation among these four single values per individual would misrepresent the among-individual correlation. In this case the sign is opposite to the true value.

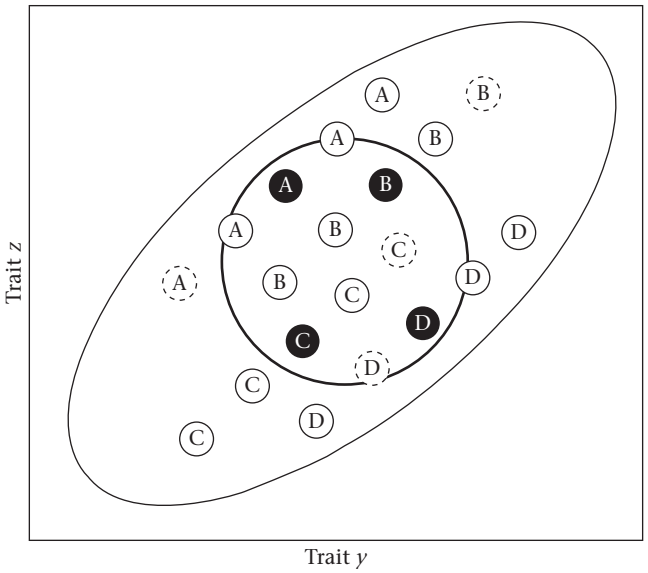

Figure 4. Illustration of a scenario where four individuals (letters A, B, C, D) are assayed five times (circles) for each of two traits (y, z) simultaneously, both of which are repeatable (R>0). Circles represent observations; black circles represent each individual's average trait value. The scenario shown here is one where within-individual changes in trait y covary positively with within-individual changes in trait z, causing a strongly positive within-individual correlation. By contrast, the among-individual correlation is zero because mean values are not correlated. Again, the phenotypic correlation (nonbolded ellipse) does not represent the among-individual correlation (bolded ellipse). If we had just collected one data point per individual (broken circles), the among-individual correlation among these four single values per individual would misrepresent the among-individual correlation. In this case the estimate would be biased upwards.

nonzero R inadvertently assume that phenotypic correlations qualitatively and quantitatively match among-individual correlations; this key assumption will often not be met (see section Critical assumptions when applying the 'individual gambit'). For example, Figs. 2–4 illustrate scenarios where two labile traits are both repeatable, yet the phenotypic correlation between two traits measured once misrepresents the among-individual correlation in all cases.

It is well known that repeatedly expressed traits can vary (and covary) both among- and within individuals (Boake, 1989; Brommer, 2013; Dingemanse & Dochtermann, 2013; Dingemanse et al., 2012; Lynch & Walsh, 1998). The repeatability (R) of behavioural and physiological traits is shown to be, on average, ca. 0.4 (Bell, Hankison, & Laskowski, 2009; Holtmann, Lagisz, & Nakagawa, 2017). This means that the residual (unexplained within-individual) variation accounts for ca. 60% (1-R) of the total variation in such labile traits. Thus, when using single (phenotypic) measurements, a logical conclusion would in fact be, owing to (1-R)>R, that unpartitioned phenotypic correlations mostly represent within-individual level patterns! This is obvious when looking at the equation relating a phenotypic correlation to its among- and within-individual components: when (1-R)>R, the residual within-individual correlation largely defines the phenotypic correlation, as pointed out repeatedly in statistically oriented papers on this subject (Brommer, 2013; Dingemanse & Dochtermann, 2013; Dingemanse et al., 2012; Downs & Dochtermann, 2014) (equation 1):

$$r_{p_y,p_z} = r_{i_y,i_z}\sqrt{R_y R_z} + r_{e_y,e_z}\sqrt{(1-R_y)(1-R_z)} \qquad (1)$$

where r_{p_y,p_z}, r_{i_y,i_z} and r_{e_y,e_z} are the phenotypic, among-individual and within-individual correlations between traits y (e.g. a

behavioural trait) and z (e.g. a physiological trait), respectively. R_y and R_z are the repeatabilities of the behaviour and physiology, respectively.

CRITICAL ASSUMPTIONS WHEN APPLYING THE 'INDIVIDUAL GAMBIT'

Empiricists adopting the 'individual gambit' implicitly (perhaps often unknowingly) make at least two, mutually nonexclusive, yet critical assumptions. If even one of these two assumptions is violated, taking the 'individual gambit' will lead to a biased interpretation, where phenotypic correlations do not match the among-individual level correlation of interest.

The first major assumption is that among- and within-individual correlations are identical. Indeed, when a focal behaviour and another phenotypic trait of interest correlate in the same way among- and within-individuals, phenotypic correlations (even among traits each measured once) match among-individual correlations (for mathematical arguments, see Dingemanse et al., 2012; Brommer, 2013; Dingemanse & Dochtermann, 2013) (equation 1). Unfortunately, our recent meta-analysis shows that among- and within-individual correlations between behaviours and hormone levels, metabolic rates, body mass and body size actually do differ within versus between individuals: among-individual correlations were tighter than within-individual correlations (Niemelä & Dingemanse, 2018). Phenotypic correlations, consequently, generally underestimate among-individual correlations (as illustrated in Fig. 2). This notion is, notably, firmly embedded in the ecological literature. For example, Adolph and Hardin (2007) implied that the residual within-individual correlation (r_{e_y,e_z}) may often be zero because measurement errors and patterns of within-individual plasticity should typically not be correlated across traits. In such cases (where $r_{e_y,e_z} = 0$), equation 1 can be simplified into equation 2:

$$r_{p_y,p_z} = r_{i_y,i_z}\sqrt{R_y R_z} \quad (2)$$

Rearranging this equation gives equation 3:

$$r_{i_y,i_z} = r_{p_y,p_z} / \sqrt{R_y R_z} \quad (3)$$

This demonstrates that phenotypic correlations overestimate among-individual correlations for such cases because geometric mean repeatabilities ($\sqrt{R_y R_z}$) are always lower than 1 for labile traits or traits measured with error (Adolph & Hardin, 2007). Recent research has shown that this assumption ($r_{e_y,e_z} = 0$) is extreme, and often invalid (absolute magnitude of $r_{e_y,e_z} > 0$; Niemelä & Dingemanse, 2018). Nevertheless, the general conclusion stands: unpartitioned phenotypic correlations, or phenotypic correlations between traits each measured just once, can greatly misestimate the among-individual correlation when correlation structures differ within versus between individuals (Figs. 2–4).

The second implicit assumption made when taking the 'individual gambit' is that the repeatability (i.e. R) of the focal traits equals 1. If $R = 1$, by definition, the phenotypic correlation equates the among-individual correlation and the 'individual gambit' is indeed applicable (for a mathematical argument, see equations 2 and 3). However, repeatabilities of behaviour (Bell et al., 2009) and physiology (Holtmann et al., 2017) are on average only ca. 0.4. Repeatabilities of even seemingly more 'fixed' traits (Gosler, 1987; Iserbyt, Eens, Baetens, Vermeulen, & Müller, 2017; Radford & Plessis, 2004) are, similarly, typically below 1 (i.e. due to measurement error), indicating that this major assumption rarely holds true.

SINGLE MEASUREMENTS AND THE SCIENTIFIC METHOD

There are several reasons why researchers might measure traits only once despite being interested in questions requiring repeated measures data. One argument that is often used is that it would be too time consuming, expensive or seemingly impossible to measure the same set of traits multiple times on many individuals. This argument should, in our view, be taken with a grain of salt, because various laboratories do manage to collect vast amounts of repeated measures data on phenotypic traits that are 'hard' to assay (e.g. Boulton et al., 2015; Careau et al., 2015; Dosmann, Brooks, & Mateo, 2015; Ferrari et al., 2013; Gifford, Clay, & Careau, 2014; Iserbyt et al., 2017; Krams et al., 2017; Royauté, Greenlee, Baldwin, & Dochtermann, 2015). These laboratories may manage to collect the required data because they resolve financial or time allocation trade-offs in favour of investments in a few projects (with large sample sizes, e.g. repeated measurement of many traits) versus investments in many projects (with small sample sizes, e.g. single measurements), although we recognize that other explanations may be given too. In various cases, large data sets are collected because multiple principle investigators team up, such that many hands can jointly collect lots of data (e.g. Boulton et al., 2015; Nicolaus, Tinbergen, Ubels, Both, & Dingemanse, 2016). In other words, a priori planning of research strategies appears an important means to safeguard the collection of enough data. Examples of successful projects include even field studies with challenging longitudinal designs where each individual is only sampled once but every year of its life for multiple traits simultaneously (Duckworth & Badyaev, 2007; Quinn, Cole, Patrick, & Sheldon, 2011; Réale, Martin, Coltman, Poissant, & Festa-Bianchet, 2009).

Although the unbiased estimation of among-individual correlations strictly requires repeated measures data, the statistical methods required for data analyses are relatively flexible and can work with a large diversity of study designs. The multivariate mixed-effects models that have repeatedly been proposed as a heuristic tool to partition phenotypic correlations into among- and within-individual components (Brommer, 2013; Dingemanse & Dochtermann, 2013) are designed to handle unbalanced data sets, providing researchers with considerable flexibility. Briefly, the number of repeated measures may differ between individuals, the number of repeated measures may differ between traits, and the traits of interest do not need to be measured at the same time (e.g. Table 2 in Dingemanse & Dochtermann, 2013). This flexibility allows researchers to resolve financial and time allocation trade-offs in data collection. Having said that, we do of course recognize that many academics may suffer from a lack of resources (in terms of time and funding) to be in a position where such trade-offs can be meaningfully resolved. Importantly, various software packages exist that enable researchers to estimate accuracy, precision and statistical power associated with a set study design (Allegue et al., 2017; Araya-Ajoy, Mathot, & Dingemanse, 2015; van de Pol, 2012). For example, the R-package SQUID features a user-friendly online platform where students can investigate how power and bias in parameter estimates varies as a function of study design decisions (Allegue et al., 2017). Researchers could thereby determine the optimal study design given various constraints. For example, one could investigate whether detrimental effects associated with minimizing the number of repeated measures taken for traits difficult (or costly) to measure could be partly compensated for by taking more repeated measures for other traits, etc.

An important question, notably, is under which conditions researchers should pursue a particular scientific study. In our view, researchers should investigate the optimal study design required to address the scientific question of interest (e.g. using simulation, detailed above), and then decide whether this question can be

addressed robustly given various time allocation, financial or biological constraints (i.e. species-specific or traits-specific constraints making the acquisition of repeated measurements difficult or impossible; Fig. 5). If the answer is 'no', the best option would be to either alleviate the constraints mentioned above (e.g. by switching study systems, postponing the experiment until more resources or hands are available), or focus on other research questions where the answer is 'yes'. In other words, the decision on whether to collect, and how to use, the data should best be taken prior to data collection. Indeed, such recommendations are viewed as cornerstones of the scientific method (Creswell, 1994).

We fully appreciate that researchers nevertheless end up in situations where the intended research, whether in study design, data quality or quantity, did not proceed according to plan, resulting in suboptimal data sets (Fig. 5). In the context of studies requiring among-individual correlation estimates, this may apply to data sets where suites of correlated traits were measured only once. Such data sets have considerable value and are worth

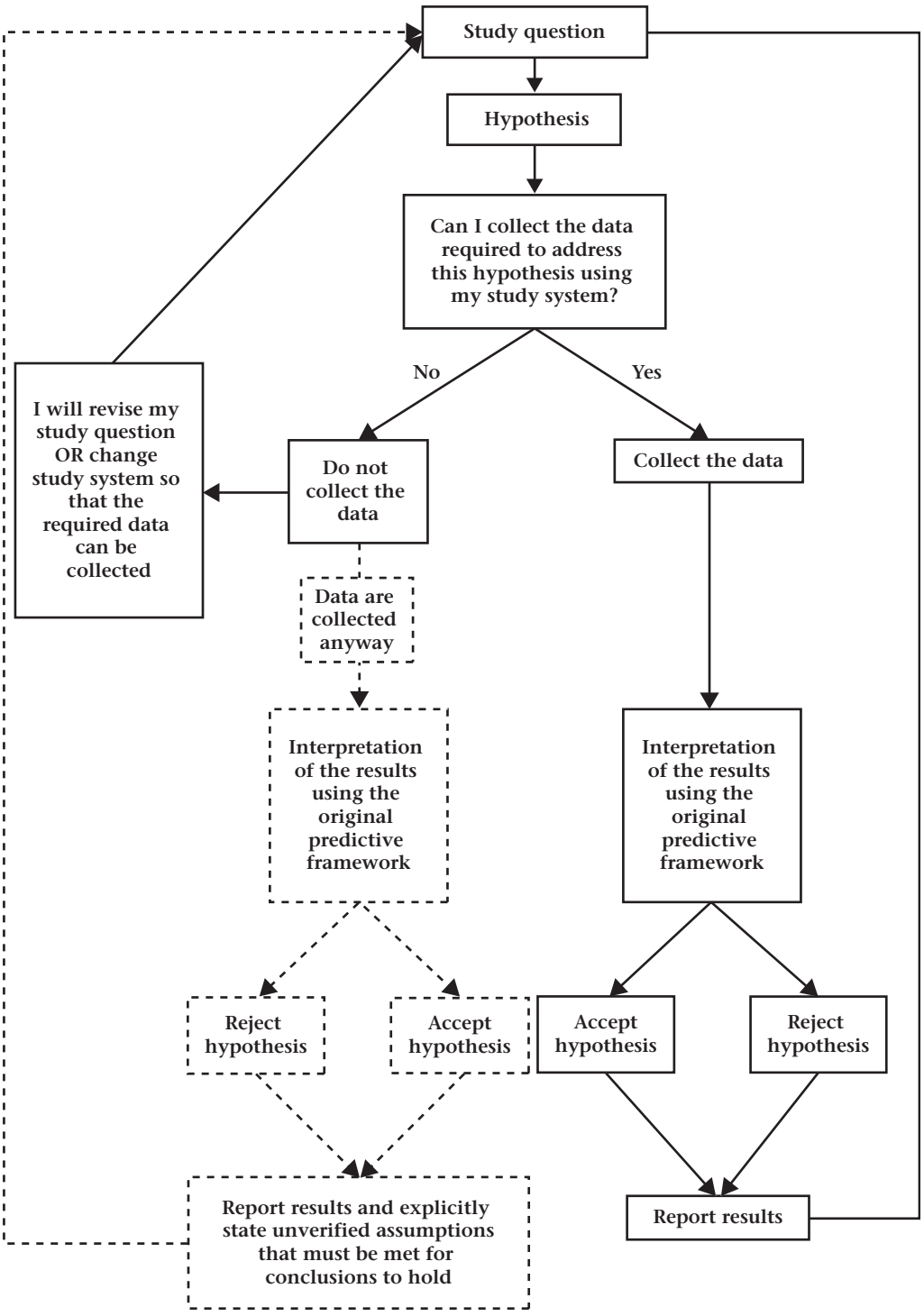

Figure 5. Schematic presentation of a decision tree. Dashed area indicates a path that we consider to be suboptimal.

publishing; however, researchers should then openly and transparently acknowledge that their conclusions hinge on major assumptions that may or may not hold (see section Critical assumptions when applying 'individual gambit'). In our view, such an approach will not take away from the quality of the paper; rather it will increase it! Importantly, it will help make readers aware that such issues may be overcome by using other study designs in the future.

STATISTICAL APPROACHES

The statistical knowledge of how to estimate and study level-specific correlations has existed for decades and has been primarily developed by quantitative geneticists (Falconer & Mackay, 1996; Lynch & Walsh, 1998). These methods have also been introduced several years ago to behavioural ecology and animal behaviour (Dingemanse & Dochtermann, 2013; Wilson et al., 2010). Despite their power, these methodologies are still not routinely used by many. For example, multivariate mixed-effects models, allowing the estimation of level-specific correlations between labile, repeatedly measured, traits (Brommer, 2013; Dingemanse & Dochtermann, 2013; Lynch & Walsh, 1998; Wilson et al., 2010) are still rarely used to estimate individual level correlations in personality research (Niemelä & Dingemanse, 2018). Importantly, the usage of repeated measurements in the statistical model does not automatically mean that empiricists would be estimating individual level associations between traits. In fact, despite repeated measures data used in the statistical model, many empiricists still use statistical tools that return unpartitioned phenotypic level parameter estimates when studying individual level questions (Niemelä & Dingemanse, 2018). We fully agree with Bolker et al. (2009, p. 127) that 'researchers should use statistical approaches that match their data'. However, equally importantly, researchers should collect data and use statistical methods that allow them to answer their study question.

CONCLUSION

Recent work (including our own) demonstrates that behavioural ecologists generally do not seek to answer questions concerning among-individual variation by using repeated measures data (Niemelä & Dingemanse, 2018; Royauté, Berdal, Garrison, & Dochtermann, 2018). Even though this application of the 'individual gambit' may sometimes produce accurate results (Brommer & Class, 2017), critical assumptions are typically neither discussed nor empirically shown to hold true. Some have proposed that the general match between phenotypic and individual correlations is sufficient validation. However, species and populations are known to differ in correlation structures (Brommer & Class, 2017; Niemelä & Dingemanse, 2018). This means that a focal study cannot rely on these major assumptions to hold true. Another problem is that any validation might hold only within specific environments. Indeed, environmental treatments are known to affect correlation structures, and may do so differently across different levels of variation (e.g. Royauté & Dochtermann, 2017). This implies that validations in one type of environment may not necessarily apply generally. The collection of the appropriate repeated measures data for each study focusing on individual level questions will help to alleviate such concerns. Future studies should perhaps make a greater effort to reduce the potential biases in the behavioural literature either (1) by adopting study designs and statistical tools that allow estimation of among-individual parameters or, in suboptimal cases, (2) by making an effort to fully acknowledge that the 'individual gambit' is applied and that some critical assumptions are made (see above). The present state of the literature shows that such a shift in scientific practice is both needed and beneficial for the field as a whole.

Acknowledgments

P.T.N. was supported by Deutsche Forschungsgemeinschaft (DFG; grant number NI 1539/2-1) and N.J.D was supported by the Ludwig-Maximilians University. We declare no competing interests. We thank Nicholas DiRienzo, Ralf Kurvers, Maria Moiron, Alfredo Sánchez-Tójar, Alastair Wilson and two anonymous referees for their constructive comments on the manuscript.

References

Adolph, S. C., & Hardin, J. S. (2007). Estimating phenotypic correlations: Correcting for bias due to intraindividual variability. *Functional Ecology, 21*, 178–184. https://doi.org/10.1111/j.1365-2435.2006.01209.x.

Allegue, H., Araya-Ajoy, Y. G., Dingemanse, N. J., Dochtermann, N. A., Garamszegi, L. Z., Nakagawa, S., et al. (2017). Statistical quantification of individual differences (SQuID): An educational and statistical tool for understanding multilevel phenotypic data in linear mixed models. *Methods in Ecology and Evolution, 8*, 257–267. https://doi.org/10.1111/2041-210X.12659.

Araya-Ajoy, Y. G., Mathot, K. J., & Dingemanse, N. J. (2015). An approach to estimate short-term, long-term and reaction norm repeatability. *Methods in Ecology and Evolution, 6*, 1462–1473. https://doi.org/10.1111/2041-210X.12430.

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: A meta-analysis. *Animal Behaviour, 77*(4), 771–783. https://doi.org/10.1016/j.anbehav.2008.12.022.

Biro, P. A., & Stamps, J. A. (2010). Do consistent individual differences in metabolic rate promote consistent individual differences in behavior? *Trends in Ecology & Evolution, 25*(11), 653–659. https://doi.org/10.1016/j.tree.2010.08.003.

Boake, C. R. B. (1989). Repeatability: Its role in evolutionary studies of mating behavior. *Evolutionary Ecology, 3*, 173–182. https://doi.org/10.1007/BF02270919.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., et al. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution, 24*, 127–135. https://doi.org/10.1016/j.tree.2008.10.008.

Boulton, K., Couto, E., Grimmer, A. J., Earley, R. L., Canario, A. V. M., Wilson, A. J., et al. (2015). How integrated are behavioral and endocrine stress response traits? A repeated measures approach to testing the stress-coping style model. *Ecology and Evolution, 5*(3), 618–633. https://doi.org/10.1002/ece3.1395.

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"? *Behavioral Ecology and Sociobiology, 67*(6), 1027–1032. https://doi.org/10.1007/s00265-013-1527-4.

Brommer, J. E., & Class, B. (2017). Phenotypic correlations capture between-individual correlations underlying behavioral syndromes. *Behavioral Ecology and Sociobiology, 71*, 50. https://doi.org/10.1007/s00265-017-2278-4.

Careau, V., Montiglio, P. O., Garant, D., Pelletier, F., Speakman, J. R., Humphries, M. M., et al. (2015). Energy expenditure and personality in wild chipmunks. *Behavioral Ecology and Sociobiology, 69*(4), 653–661. https://doi.org/10.1007/s00265-015-1876-2.

Careau, V., Thomas, D., Humphries, M. M., & Réale, D. (2008). Energy metabolism and animal personality. *Oikos, 117*(5), 641–653. https://doi.org/10.1111/j.0030-1299.2008.16513.x.

Creswell, J. W. (1994). *Research design: qualitative & quantitative approaches*. London, UK: SAGE Publications.

Dammhahn, M., Dingemanse, N. J., Niemelä, P. T., & Réale, D. (2018). Pace-of-life syndromes: A framework for the adaptive integration of behaviour, physiology and life history. *Behavioral Ecology and Sociobiology, 72*, 62. https://doi.org/10.1007/s00265-018-2473-y.

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology, 82*(1), 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. *Behavioral Ecology and Sociobiology, 66*, 1543–1548. https://doi.org/10.1007/s00265-012-1416-2.

Dochtermann, N. A., & Dingemanse, N. J. (2013). Behavioral syndromes as evolutionary constraints. *Behavioral Ecology, 24*(4), 806–811. https://doi.org/10.1093/beheco/art002.

Dosmann, A. J., Brooks, K. C., & Mateo, J. M. (2015). Within-individual correlations reveal link between a behavioral syndrome, condition, and cortisol in free-ranging Belding's ground squirrels. *Ethology, 121*(2), 125–134. https://doi.org/10.1111/eth.12320.

Downs, C. J., & Dochtermann, N. A. (2014). Testing hypotheses in ecoimmunology using mixed models: Disentangling hierarchical correlations. *Integrative and Comparative Biology, 54*, 407–418. https://doi.org/10.1093/icb/icu035.

Duckworth, R. A., & Badyaev, A. V. (2007). Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird. *Proceedings of the*

National Academy of Sciences, 104, 15017–15022. https://doi.org/10.1073/pnas.0706174104.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. Longman. Introduction to quantitative genetics https://doi.org/10.1016/j.bbabio.2008.04.029.

Ferrari, C., Pasquaretta, C., Carere, C., Cavallone, E., von Hardenberg, A., & Réale, D. (2013). Testing for the presence of coping styles in a wild mammal. *Animal Behaviour, 85*(6), 1385–1396. https://doi.org/10.1016/j.anbehav.2013.03.030.

Gifford, M. E., Clay, T. A., & Careau, V. (2014). Individual (Co)variation in standard metabolic rate, feeding rate, and exploratory behavior in wild-caught semi-aquatic salamanders. *Physiological and Biochemical Zoology, 87*(3), 384–396. https://doi.org/10.1086/675974.

Goodman, L. A. (1953). Ecological regressions and behavior of individuals. *American Sociological Review, 18*, 663. https://doi.org/10.2307/2088121.

Gosler, A. G. (1987). Pattern and process in the bill morphology of the Great Tit Parus major. *Ibis, 129*, 451–476. https://doi.org/10.1111/j.1474-919X.1987.tb08234.x.

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*(3), 685–696. https://doi.org/10.1111/1365-2435.12779.

Iserbyt, A., Eens, M., Baetens, W., Vermeulen, A., & Müller, W. (2017). Within- and between-individual (co)variance partitioning reveals limited pleiotropic effects of testosterone on immune function, sexual signaling, and parental investment. *Behavioral Ecology and Sociobiology, 71*(4). https://doi.org/10.1007/s00265-017-2308-2.

Krams, I. A., Niemelä, P. T., Trakimas, G., Krams, R., Burghardt, G. M., Krama, T., et al. (2017). Metabolic rate associates with, but does not generate covariation between, behaviours in western stutter-trilling crickets, Gryllus integer. *Proceedings of the Royal Society B: Biological Sciences, 284*. https://doi.org/10.1098/rspb.2016.2481.

Lessells, C. M., & Boag, P. T. (1987). Unrepeatable repeatabilities: A common mistake. *Auk, 104*, 116–121. https://doi.org/10.2307/4087240.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer. Genetics and analysis of quantitative traits https://doi.org/10.1086/318209.

Nicolaus, M., Tinbergen, J. M., Ubels, R., Both, C., & Dingemanse, N. J. (2016). Density fluctuations represent a key process maintaining personality variation in a wild passerine bird. *Ecology Letters, 19*(4), 478–486. https://doi.org/10.1111/ele.12584.

Niemelä, P. T., & Dingemanse, N. J. (2018). 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.

van de Pol, M. (2012). Quantifying individual variation in reaction norms: How study design affects the accuracy, precision and power of random regression models. *Methods in Ecology and Evolution, 3*, 268–280. https://doi.org/10.1111/j.2041-210X.2011.00160.x.

Quinn, J. L., Cole, E. F., Patrick, S. C., & Sheldon, B. C. (2011). Scale and state dependence of the relationship between personality and dispersal in a great tit population. *Journal of Animal Ecology, 80*, 918–928. https://doi.org/10.1111/j.1365-2656.2011.01835.x.

Radford, A. N., & Plessis, M. A. du (2004). Extreme sexual dimorphism in green woodhoopoe (Phoeniculus purpureus) bill length: A case of sexual selection? *Auk, 121*, 178–183. https://doi.org/10.2307/4090067.

Réale, D., Garant, D., Humphries, M. M., Bergeron, P., Careau, V., & Montiglio, P. O. (2010). Personality and the emergence of the pace-of-life syndrome concept at the population level. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*, 4051–4063. https://doi.org/10.1098/rstb.2010.0208.

Réale, D., Martin, J., Coltman, D. W., Poissant, J., & Festa-Bianchet, M. (2009). Male personality, life-history strategies and reproductive success in a promiscuous mammal. *Journal of Evolutionary Biology, 22*, 1599–1607. https://doi.org/10.1111/j.1420-9101.2009.01781.x.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357.

Royauté, R., Berdal, M. A., Garrison, C. R., & Dochtermann, N. A. (2018). Paceless life? A meta-analysis of the pace-of-life syndrome hypothesis. *Behavioral Ecology and Sociobiology, 72*, 64.

Royauté, R., & Dochtermann, N. A. (2017). When the mean no longer matters: Developmental diet affects behavioral variation but not population averages in the house cricket (Acheta domesticus). *Behavioral Ecology, 28*(1), 337–345. https://doi.org/10.1093/beheco/arw164.

Royauté, R., Greenlee, K., Baldwin, M., & Dochtermann, N. A. (2015). Behaviour, metabolism and size: Phenotypic modularity or integration in Acheta domesticus? *Animal Behaviour, 110*, 163–169. https://doi.org/10.1016/j.anbehav.2015.09.027.

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., et al. (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology, 79*, 13–26. https://doi.org/10.1111/j.1365-2656.2009.01639.x.