

On between-individual and residual (co)variances in the study of animal personality: are you willing to take the “individual gambit”?

Jon E. Brommer

Received: 21 December 2012 / Revised: 14 March 2013 / Accepted: 14 March 2013 / Published online: 26 March 2013
© Springer-Verlag Berlin Heidelberg 2013

Summary A recent discussion in this journal (Dingemanse et al. Behav Ecol Sociobiol 66: 1543–1548, 2012; Garamszegi and Herczeg Behav Ecol Sociobiol 66:1651–1658, 2012) deals with a core issue in animal personality research: Can animal personality research quantify correlated behaviors on the between-individual level, or is this too demanding in terms of design and analysis of the data, and should behavioral ecologists therefore take the “individual gambit” and work on the phenotypic level only. Taking this gambit implies accepting that the between-individual correlation in behavioral traits (which is the correlation of interest) may be masked by a residual correlation of different magnitude or sign. Understanding (co)variances on different levels is the main thrust of quantitative genetics, and animal personality research can make good use of the plethora of ideas and analytical approaches developed in this field. I, here, outline reasons why the “individual gambit” may or may not work out and its relationship to the quantitative genetic “phenotypic gambit”. I especially emphasize the meaning of residuals and phenotypic plasticity which has not been fully appreciated in the debate thus far. I conclude that instead of a priori assuming that between-individual correlations are captured sufficiently well by the phenotypic correlation, animal personality researchers should set up

more ambitious data collection and analysis designs to critically test this conjectured equality.

Introduction

When a metric of behavior is repeatable, it is commonly considered a trait which quantifies an aspect of animal personality. Correlated behavioral traits, e.g., boldness and aggression, are considered a behavioral syndrome (Réale et al. 2007; Sih et al. 2004). Recently, an interesting discussion has started concerning how we should define and measure this “correlation” (Dingemanse et al. 2012; Garamszegi and Herczeg 2012). Personality research emphasizes the between-individual (co)variation of behavioral traits. Just like the between-individual, variance is only a part of the total phenotypic variance (typically 37 %, Bell et al. 2009); the between-individual covariance is only a part of the phenotypic covariance. Partitioning out the between-individual correlation from the phenotypic one is potentially important because a between-individual correlation need not be captured adequately by the phenotypic correlation (detailed in Dingemanse and Dochtermann 2013). Clearly, partitioning the covariance into these different levels requires repeated measures to be taken on individuals. This means that data collection and study design should be adjusted accordingly.

Garamszegi and Herczeg (2012) state that “Until the problems with separating between- and within-individual behavioral correlations are convincingly solved, we suggest focusing on the phenotypic correlations of individual-specific estimates of traits (or their ranks) together with the repeatability of behaviors”. These authors advocate a kind of mixed approach, where between-individual variance is the focus in a study of a single trait, but when multiple traits are studied; their phenotypic covariance is conjectured to be

Communicated by J. Lindström.

J. E. Brommer (✉)
Department of Biology, University Hill, University of Turku,
Turku 20014, Finland
e-mail: jon.brommer@utu.fi

J. E. Brommer
ARONIA Coastal Zone Research Team,
Novia University of Applied Sciences and Åbo Akademi
University, Raseborgsvägen 9,
10600 Ekenäs, Finland

sufficient in capturing correlated aspects of animal personality or behavioral syndromes. Their argument bases on the assertion that (1) partitioning the phenotypic correlation into a between-individual and residual correlation has a problematic basis, and (2) the sample size needed to statistically reliably do this is too large for behavioral ecologists to collect in reality. Although technical, the ongoing discussion on partitioning of covariances is interesting because it circles around a core question. On what level (phenotype, between-individual, genotype) should we aim to study animal personality? In this paper, I look at these issues from the perspective of quantitative genetics. I outline that the basis for partitioning (co) variances has a long and proven history in this field. I continue by pointing out some assumptions of the nature of residuals and their covariance, which appear to have been appreciated insufficiently, and which may have caused confusion in the ongoing debate.

Foundation for partitioning covariances and correlations

Most behavioral ecologists would emphasize that animal personality concerns “consistency in behavior”, and emphasize the statistical concept of repeatability and thus of between-individual variance. Conceptually, this means that we hypothesize that there is a single intrinsic value for each individual for the behavioral trait under study. In the context of animal personality, the existence of an intrinsic value is for example implied when referring to individuals falling on a continuum (e.g., boldness to shy, fast to slow explorers, Réale et al. 2007). Although an individual has one intrinsic value, our phenotypic measure may not be equal to this value because of random deviations. Thus, phenotypic measures z of traits x and y taken at time t on a specific individual n can be considered

$$\begin{aligned} Z_{x,n,t} &= \mu_x + i_{x,n} + \varepsilon_{x,n,t} \\ Z_{y,n,t} &= \mu_y + i_{y,n} + \varepsilon_{y,n,t} \end{aligned} \quad (1)$$

Where, for convenience, the individual n 's intrinsic value i for each trait is considered as a deviation from the mean value μ . The distribution of i across individuals is assumed to be normal, with a mean of zero and a certain between-individual variance. Further, ε are residuals, i.e., random values drawn from a normal distribution with a certain (residual) variance and a mean of zero. Because of these residuals, phenotypic measures taken at different time steps will tend to differ from each other and from the individual's intrinsic value i . Traits that are expressed repeatedly by the same individual are termed labile traits, such as many behavioral and life-history traits (Lynch and Walsh 1998).

Dropping the subscripts, repeatability (r^2) can in general be denoted as

$$r^2 = \text{var}(i) / \text{var}(z) = \text{var}(i) / [\text{var}(i) + \text{var}(\varepsilon)], \quad (2)$$

where $\text{var}(i)$ stands for variance between individuals in their intrinsic value, and $\text{var}(z)$ and $\text{var}(\varepsilon)$ quantify the variances in the phenotypic and residuals, respectively (Falconer and MacKay 1996).

Starting from these premises, we can follow the logic of Searle (1961). Briefly, using the definition of a covariance and Eq. (1), the covariances between traits x and y measured on a set of individuals can be partitioned as

$$\begin{aligned} \text{cov}(z_x, z_y) &= \text{cov}(i_x, i_y) + \text{cov}(\varepsilon_x, \varepsilon_y) \\ &\quad + \text{cov}(i_x, \varepsilon_y) + \text{cov}(\varepsilon_x, i_y) \end{aligned} \quad (3a)$$

$$= \text{cov}(i_x, i_y) + \text{cov}(\varepsilon_x, \varepsilon_y) \quad (3b)$$

Thus, the phenotypic covariance $\text{cov}(z_x, z_y)$ is broken up into a sum of the covariance between individuals $\text{cov}(i_x, i_y)$ and between the residuals $\text{cov}(\varepsilon_x, \varepsilon_y)$. The covariance between the intrinsic values of trait x and the residuals of trait y (and vice versa) cancel out of Eq (3a) to form Eq. (3b) because the residuals values are random deviations (Eq. (1)). Scaling the covariances to a correlation produces

$$\begin{aligned} \frac{\text{cov}(z_x, z_y)}{\sqrt{\text{var}(z_x) \text{var}(z_y)}} &= \frac{\text{cov}(i_x, i_y)}{\sqrt{\text{var}(z_x) \text{var}(z_y)}} \\ &\quad + \frac{\text{cov}(\varepsilon_x, \varepsilon_y)}{\sqrt{\text{var}(z_x) \text{var}(z_y)}} \end{aligned} \quad (4a)$$

$$\begin{aligned} \frac{\text{cov}(z_x, z_y)}{\sqrt{\text{var}(z_x) \text{var}(z_y)}} &= \frac{\text{cov}(i_x, i_y)}{\sqrt{\text{var}(z_x) \text{var}(z_y)}} \\ &\quad \times \frac{\sqrt{\text{var}(i_x) \text{var}(i_y)}}{\sqrt{\text{var}(i_x) \text{var}(i_y)}} \\ &\quad + \frac{\text{cov}(\varepsilon_x, \varepsilon_y)}{\sqrt{\text{var}(z_x) \text{var}(z_y)}} \\ &\quad \times \frac{\sqrt{\text{var}(\varepsilon_x) \text{var}(\varepsilon_y)}}{\sqrt{\text{var}(\varepsilon_x) \text{var}(\varepsilon_y)}} \end{aligned} \quad (4b)$$

$$r_P(x, y) = r_I(x, y) \sqrt{r_x^2 r_y^2} + r_\varepsilon(x, y) \sqrt{(1 - r_x^2)(1 - r_y^2)}, \quad (4c)$$

where r_P denotes the phenotypic correlation, r_I the between-individual correlation, and r_ε the residual correlation. Eq. (4c) was also derived by Dingemans et al. (2012) in another fashion. The basis of this equation is established within the field of quantitative genetics, which hinges upon statistically partitioning (co)variances into hierarchically structured levels. The critical assumptions are that the effects on the different levels are additive (Eq. 1) and independent from each other (Eq. (3a): $\text{cov}(i_x, \varepsilon_y)$ and $\text{cov}(\varepsilon_x, i_y)$ are zero). Partitioning of (co)variances has been done in multivariate mixed models for decades and can readily be extended to include further levels when relevant.

Because the interest in life-history traits has been considerable in evolutionary ecology, there exists a strong framework for the analysis of labile traits (Roff 2007; Lynch and Walsh 1998). In particular, it is useful to widen the interpretation of the intrinsic value i . From the perspective of quantitative genetics, i can be assumed to be the sum of the breeding value (a) and the permanent environmental effect (pe). The breeding value is the expected value of the trait in case it was solely determined by the additive effects of the many loci which are assumed to determine its expression. The permanent environment effect captures nonheritable effects which are associated to the individual such as maternal or natal effects or effects related to the conditions an individual experienced during its lifetime. It also includes nonadditive genetic effects (dominance). Searle (1961) originally derived Eq. (4c) for a nonlabile trait, where the partitioning was between the breeding value of the trait and its residual (cf. Roff 1997), but its logic applies equally to between-individual differences. Assuming additive genetic and permanent environmental effects are additive and independent from each other, this additional level of (co)variance can be included, since $\text{cov}(i_x, i_y) = \text{cov}(a_x, a_y) + \text{cov}(pe_x, pe_y)$ to realize that

$$r_P(x, y) = r_A(x, y) \sqrt{h_x^2 h_y^2} + r_{PE}(x, y) \sqrt{p_x^2 p_y^2} + r_\varepsilon(x, y) \sqrt{(1 - r_x^2)(1 - r_y^2)}, \quad (5)$$

where h^2 and p^2 denote the proportion of phenotypic variance due to additive genetic effects (heritability) and permanent environmental effects, respectively. The sum of these two is the repeatability r^2 : That is, Eq. (5) is a further partitioning of Eq. (4c). The correlations r_A and r_{PE} are defined on the genetic and permanent environmental level, respectively. Because an individual's intrinsic trait value i is determined by both the additive genetic effect a and a permanent environmental effect pe , the repeatability is considered the upper estimate of heritability (Falconer and MacKay 1996). It is the upper estimate because an individual's trait expression will be partly due to permanent environmental effects and a trait's heritability will therefore be lower than the repeatability (Lynch and Walsh

1998), possibly substantially lower. The further partitioning of i into a breeding value and a permanent environmental effect is the focus of a special class of linear mixed models, the animal model (Lynch and Walsh 1998; Kruuk 2004; Wilson et al. 2010). The latter is a quantitative genetic model, whose key business is the estimation of additive genetic (co)variances, because these—together with selection—determine the evolutionary dynamics of the traits (Lynch and Walsh 1998). Conceptually, however, interpretation of i as a first-line estimate of the breeding value goes a long way, both in seeing the parallels of studying animal personality to the theoretical framework of quantitative genetics and in applying the analytical methods developed within this framework. It should be clear, however, that the statistical foundation for partitioning (co)variances into multiple levels is solidly established and not something new and unproven.

The “individual gambit” in syndrome research

Behavioral ecologists noted early on that Eq. (5) implies that the phenotypic correlation may poorly reflect the genetic correlation (Krebs and Davis 1978). The genetic correlation is “hidden” beneath several additional layers of other factors (i.e., pe and ε) affecting the phenotype. Each of these factors often has a large effect, and thus largely determines the resulting phenotypic correlation. Based chiefly on pragmatic reasons, the phenotypic correlation was argued to be a reasonable proxy for the genetic one (Cheverud 1988). Making this assumption has been termed the “phenotypic gambit” (Grafen 1984). From an evolutionary perspective, it is important to ask whether this gambit holds, since the genetic correlations between traits will determine how they respond to selection (Lynch and Walsh 1998). The ongoing discussion on the importance of distinguishing the between-individual correlation from the phenotypic correlation can be understood as a special (because one level higher) case of the phenotypic gambit. The bulk of research on behavioral syndromes to date implicitly assumes that the phenotypic correlation between behavioral traits is a reasonable proxy for the between-individual correlation across traits, and thereby makes what could be termed the “individual gambit”.

This “individual gambit” may hold: A standard assumption in statistics is that residuals are random and therefore uncorrelated, both across observations made for a single trait and, by extension, between traits. The expectation is, therefore, that the residual correlation is low ($r_\varepsilon \approx 0$). When repeatability and heritability are low, as for many life-history and behavioral traits, the phenotypic correlation will mostly reflect the residual correlation [Eq. (4)] (cf. Dingemans and Dochtermann 2013). As a consequence, if $r_\varepsilon \approx 0$, the phenotypic correlation r_P will reflect the sign,

but will underestimate the magnitude, of the between-individual correlation r_I . Hence, the null assumption is that the phenotypic correlation is biased towards zero in comparison to the between-individual correlation, such that finding a nonzero phenotypic correlation between behavioral traits would, under the null assumption of noncorrelated residuals, constitute evidence for a behavioral syndrome. Dochtermann (2011) reviewed the phenotypic gambit made in animal personality research and indeed found that the phenotypic correlation r_P is correlated to the genetic correlation r_A but underestimates its magnitude. An explicit investigation of the “individual gambit” is currently lacking, but the notion of $r_\varepsilon \approx 0$, and thus of the gambit holding, is consistent with the results obtained in the meta-analysis of Garamszegi et al. (2012), who find that the phenotypic correlation between behavioural traits correlates positively with the geometric mean repeatability of these traits.

The problem is, of course, that the null assumption of uncorrelated residuals may be violated. Because repeatability of behavioral traits is low, a positive or negative residual covariance effectively determines the sign and magnitude of the phenotypic correlation and thereby masks the between-individual correlation. Residual covariance is possible whenever traits have been measured on the same individuals. This covariance may occur because of correlated measurement error (a clear risk when doing more subjective behavioral assays). Another mechanism for residual covariance is that there is an unidentified variable which affects both traits. This scenario is especially plausible for researchers working in wild populations where the environmental conditions under which measures are taken are difficult to control. Several behavioral traits may, for example, be affected by an element of the habitat which is not recognized or measured on the individual level (e.g., territory quality or food supply experienced by an individual). If this factor varies over the time it takes to collect the repeated measures on the individuals and its effect is not modeled, then it may create covariation between traits which ends up in the residual covariance. Despite the statistical null assumption of uncorrelated residuals, there is a clear possibility for a nonzero residual correlation in a behavioral syndrome. This makes it risky to take the “individual gambit”. In particular, a negative correlation on the between-individual level (e.g., a trade-off) can be masked by a strong positive residual covariance driven by positive effects of the unmeasured variable on both behavioral traits. As a consequence, the phenotypic correlation of traits with low repeatability will mostly reflect this positive residual correlation. This scenario is well established in the study of life-history traits where trade-offs (e.g., between reproduction and survival) are expected (van Noordwijk and de Jong 1986; Price et al. 1988; Roff and Fairbairn 2007).

Interpreting the residual correlation

Correlated residuals flag a potential violation of the standard statistical assumption that the residual ε is a random deviation from i . It is important to verify that a residual correlation does not arise from the fact that, for example, individuals with a high intrinsic value i for trait x and y tend to have positive residual values, and individuals with a low i tend to have negative residuals. Careful inspection of the residuals using standard procedures (e.g., the covariance between the residuals to the fitted values) is warranted. It is also worthwhile to critically consider the possibility of correlated measurement error and whether additional factors can be included in the model or in the measuring protocol in order to reduce the residual correlation. Most importantly, however, is the realization that a residual is not a property of the individual. Dingemanse et al. (2012) and Garamszegi and Herczeg (2012) use the term “within-individual” instead of residual, which can be erroneously thought to reflect some adjustments in behavior the individual makes. This view is problematic because it belies the assumption of random residuals underlying the model used to generate fixed effect estimates of interest and quantification of repeatability. The arguments and examples provided by Dingemanse et al. (2012) and Dingemanse and Dochtermann (2013) are based on an apparent belief that there is considerable biology in the residual correlation, which leads these researchers to expect it to deviate from zero and thereby determine the phenotypic correlation between behavioral traits. Again, this is not an assumption born from the underlying statistics and, as far as I am aware, also not solidly founded in available estimates. It may be true, but we do not know this for certain yet. A second particular example of interpreting residuals as having a biological meaning stems from Garamszegi and Herczeg (2012) who consider residuals as plasticity. Plasticity means that individuals change their trait value as a function of an environmental covariate; e.g., they become more aggressive when the ambient temperature is warmer. Statistically, however, such an effect is no longer a residual effect because it hypothesizes a specific factor affecting the individual's intrinsic value i (in the case of plasticity on the level of the individual rather than the genetic level, Nussey et al. 2007). Thus, plasticity implies an expansion of Eq. (1) where the potential modification of the environmental value E is described on the intrinsic value itself. For example, the phenotypic value z of individual n measured in different environments (or contexts) E can be modeled as

$$z_{n,E} = \mu + E + f_i(x, E) + \varepsilon_{n,E}, \quad (6)$$

where μ denotes the fixed-effect mean, and E the environmental value fitted as a fixed effect, and ε the residual error, which is here allowed to be specific for each value for E (heterogeneous residuals). Random-regression function $f_i(x, E)$

describes an orthogonal polynomial of order x on the between-individual level (Henderson 1982; Kirkpatrick and Heckman 1989). These are random effects modeling the difference to the fixed effect mean specific to each environmental value. For example, a first-order polynomial of f_i would represent the function $i_{0,n} + i_{1,n} \times E$, where the variances between individuals in i_0 (elevation) and $i_{1,n}$ (slope) and the covariance between slope and elevation are estimated. This approach is outlined in more detail by Nussey et al. (2007). The important point here is that plasticity (captured by $f_i(x, E)$ in Eq. (6)) is defined separate from the residuals. If a researcher believes that part of the residuals could be due to a specific environmental value E , then this hypothesis can be explored by comparing the fit of the model described by Eq. (6) to that of Eq. (1). This difference is not trivial because the presence of plasticity alters the definition of repeatability. If there is variation in plasticity across individuals, the between-individual variances and the between-individual covariances depend on where these (co) variances are evaluated along the environmental gradient (Meyer 1998). This means that repeatability of the behavioral trait is no longer defined by Eq. (1), but instead becomes a function of the environmental context (E) under which it is measured, and the contextual information on E therefore needs to be included when calculating and presenting the repeatability. Depending on the pattern of variation in plasticity across the individuals, a behavior may be repeatable only under certain environmental conditions, and the correlation in the behavior across environmental contexts may be low. Worked examples on how a random regression approach can be used to study variation in plasticity of behavioral traits are provided, e.g., by Kontiainen et al. (2009) and Klueen and Brommer (2013), and an overview is provided by Dingemanse and Dochtermann (2013). In general, several so-called function-valued approaches can be used to model plasticity (Stinchcombe et al. 2012), or the trait measured can be considered as specific to each environment (character-state approach, Lynch and Walsh 1998).

Garamszegi and Herczeg (2012) are particularly concerned that measures of two traits taken at a different point in time make it impossible to define the residual covariance and argue that this is especially problematic in animal personality research because it is almost impossible to measure two behavioral traits at the same time. Again, this concern partly stems from the confusion of interpreting the residuals as a variable related to some undefined covariate (which then changes between measures of the first and consecutive behavioral traits). If we expect this to be the case, we should of course model this covariate. For example, Eq. (6) can be expanded to a bivariate function-valued approach, where both traits x and y and their covariance are considered as a function of E . Details of this approach are not relevant for the present argument, but see Husby et al. (2010) and Dingemanse and Dochtermann (2013) for worked examples. If we do not know the factor

which may vary between the measures of the various traits, then clearly the covariance this potentially creates ends up in the residual covariance, as this term concerns covariance in unexplained differences between the measured value z and the intrinsic value i of two or more traits. I think of this line of thought not as a hindrance, but as an illustration for why researchers need to partition out the residual covariance. Because the measurement protocol may produce correlated residuals (which also include correlated measurement error), potentially correlated residuals form a clear nuisance parameter which we need to partition out of the phenotypic covariance. The between-individual correlation can be estimated across temporally displaced behavioral assays, which is viewed as a major obstacle by Garamszegi and Herczeg (2012), provided researchers commit themselves to proper design and analysis. For example, Wilson et al. (2011) assayed the same individuals in two different behavioral assays but did so repeatedly, which allows to separate the individual-level correlation r_I for the two behaviors (the behavioral syndrome) from the phenotypic correlation r_P between these behaviors, by separating the intrinsic value i for each behavior from the residual noise surrounding it.

Conclusion

Behavioral ecologists agree that a between-individual correlation defines a behavioral syndrome. The core of the debate concerns differences in willingness to make the “individual gambit” of a priori assuming that the between-individual correlation is captured adequately by the phenotypic correlation between traits. In essence, this boils down to how strongly we believe the statistical null assumption of uncorrelated residuals holds for two or more behavioral traits. Dingemanse et al. (2012) view this assumption as problematic and stress that researchers must aim to estimate the between-individual correlation explicitly by adjusting their sampling design, study protocol, and analysis accordingly. Garamszegi and Herczeg (2012), on the other hand, are more willing to make the “individual gambit” and infer the existence of a syndrome based on the phenotypic correlation only. I believe the important aspect is that researchers are aware of the assumptions and restrictions of the “individual gambit” and properly identify these. A phenotypic correlation presents weaker evidence for a syndrome than a between-individual correlation because it necessitates the assumption of uncorrelated residuals. Using multivariate mixed models, we can overcome this assumption and study correlations between behavioral traits on various levels, where especially the genetic level is important when we are interested in how natural selection creates and maintains syndromes. This is because only genes are transmitted down to further generations and covariances on all other levels,

especially on the residual level, are transient. Even if the genetic level is not the prime motivation of the research, behavioral ecologists can still put the general framework of quantitative genetics to good use. In addition to quantifying the between-individual correlation, it can accommodate interactions with the environment (plasticity), indirect effects between interacting individuals (social interactions), cross-sex genetic correlations, and trade-offs; all of these are of interest to personality research and factors which shape the potential of a species' evolutionary capacity (Moore et al. 1997; Bijma et al. 2007; Roff and Fairbairn 2007; Sprenger et al. 2012). The goal, even if challenging, should be to place behavioral ecology within the existing evolutionary framework (Dochtermann and Roff 2010). To get at these deeper levels, we need to start partitioning the phenotypic (co)variances whenever possible. Doing so will clarify whether we can trust phenotypic correlations to indeed describe the between-individual correlation sufficiently well. I, thus, hope animal personality researchers will increasingly take up the quantitative genetic toolbox to test the “individual gambit”.

Acknowledgments I thank Professor Niels J. Dingemanse for the stimulating discussion during and around his acting as an opponent for my PhD student Edward Klunen in December 2012 in Helsinki. Niels J. Dingemanse, Gabor Herczeg, and Alistair J. Wilson provided constructive reviews which improved this paper.

References

- Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behavior: a meta-analysis. *Anim Behav* 77:771–783
- Bijma P, Muir WM, van Arendonk JAM (2007) Multilevel selection 1: quantitative genetics of inheritance and response to selection. *Genetics* 175:277–288
- Cheverud JM (1988) A comparison of genetic and phenotypic correlations. *Evolution* 42:958–968
- Dingemanse NJ, Dochtermann NA (2013) Quantifying individual variation in behavior: mixed-effect modeling approaches. *J Anim Ecol* 82:39–54
- Dingemanse NJ, Dochtermann NA, Nakagawa S (2012) Defining behavioural syndromes and the role of “syndrome deviation” in understanding their evolution. *Behav Ecol Sociobiol* 66:1543–1548
- Dochtermann NA (2011) Testing Cheverud's conjecture for behavioral correlations and behavioral syndromes. *Evolution* 65:1814–20
- Dochtermann NA, Roff DA (2010) Applying a quantitative genetics framework to behavioural syndrome research. *Philos T Roy Soc B* 365:4013–4020
- Falconer DS, MacKay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Harlow
- Garamszegi LZ, Herczeg G (2012) Behavioral syndromes, syndrome deviation and the within- and between-individual components of phenotypic correlations: when reality does not meet statistics. *Behav Ecol Sociobiol* 66:1651–1658
- Garamszegi LZ, Markó G, Herczeg G (2012) A meta-analysis of correlated behaviours with implications for behavioural syndromes: mean effect size, publication bias, phylogenetic effects, and the role of mediator variables. *Evol Ecol* 26:1213–1235
- Grafen A (1984) Natural selection, kin selection and group selection in Krebs JR and Davies NB (edn) *Behavioural ecology: an evolutionary approach*. Blackwell Scientific, Oxford
- Henderson CR (1982) Analysis of covariance in the mixed model - higher-level, non-homogeneous, and random regressions. *Biometrics* 38:623–640
- Husby A, Nussey DH, Visser ME, Wilson AJ, Sheldon BC, Kruuk LEB (2010) Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit (*Parus major*) populations. *Evolution* 64:2221–2237
- Kirkpatrick M, Heckman N (1989) A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *J Math Biol* 27:429–450
- Klunen E, Brommer JE (2013) Context-specific repeatability of personality traits in a wild bird: a reaction-norm perspective. *Beh Ecol* Published Online. doi:10.1093/beheco/ars221
- Kontiaainen P, Pietiäinen H, Huttunen K, Karell P, Kolunen H, Brommer JE (2009) Aggressive Ural owl mothers recruit more offspring. *Behav Ecol* 20:789–796
- Krebs JR, Davies NB (1978) *Behavioural ecology: an evolutionary approach*. Blackwell Science, Oxford
- Kruuk LEB (2004) Estimating genetic parameters using the “animal model”. *Philos T Roy Soc B* 359:873–890
- Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer, Sunderland
- Meyer K (1998) Estimating covariance functions for longitudinal data using a random regression model. *Genet Sel Evol* 30:221–240
- Moore AJ, Brodie ED, Wolf JB (1997) Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. *Evolution* 51:1352–1362
- Nussey D, Wilson AJ, Brommer JE (2007) The evolutionary ecology of individual phenotypic plasticity in wild populations. *J Evol Biol* 20:831–844
- Price T, Kirkpatrick M, Arnold SJ (1988) Directional selection and the evolution of breeding date in birds. *Science* 240:798–799
- Réale D, Reader SM, Sol D, McDougall P, Dingemanse NJ (2007) Integrating temperament in ecology and evolutionary biology. *Biol Rev* 82:291–318
- Roff DA (1997) *Evolutionary quantitative genetics*. Chapman and Hall, New York
- Roff DA, Fairbairn DJ (2007) The evolution of trade-offs: where are we? *J Evol Biol* 20:433–447
- Searle SR (1961) Phenotypic, genetic, and environmental correlations. *Biometrics* 17:474–480
- Sih A, Bell AM, Johnson JC, Ziemba RE (2004) Behavioral syndromes: an integrative overview. *Q Rev Biol* 79:241–277
- Sprenger D, Dingemanse NJ, Dochtermann NA, Theobald J, Walker SPW (2012) Aggressive females become aggressive males in a sex-changing reef fish. *Ecol Lett* 15:986–992
- Stinchcombe JR, Function-valued Traits WorkingGroup, Kirkpatrick M (2012) Genetics and evolution of function-valued traits: understanding environmentally responsive phenotypes. *Trends Ecol Evol* 27:637–647
- van Noordwijk AJ, de Jong G (1986) Acquisition and allocation of resources: their influence on variation in life history tactics. *Am Nat* 128:137–142
- Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LEB, Nussey DH (2010) An ecologist's guide to the animal model. *J Anim Ecol* 79:13–26
- Wilson AJ, de Boer M, Arnott G, Grimmer A (2011) Integrating animal personality research and animal context theory: aggressiveness in the green swordtail *Xiphophorus helleri*. *PLoS One* 6: e28024