ORIGINAL PAPER

A meta-analysis of correlated behaviours with implications for behavioural syndromes: mean effect size, publication bias, phylogenetic effects and the role of mediator variables

László Zsolt Garamszegi · Gábor Markó · Gábor Herczeg

Abstract In evolutionary and behavioural ecology, increasing attention is being paid to the fact that functionally distinct behaviours are often not independent from each other. Such phenomenon is labelled as behavioural syndrome and is usually demonstrated by phenotypic correlations between behaviours like activity, exploration, aggression and risk-taking across individuals in a population. However, published studies disagree on the strength, and even on the existence of such relationships. To make general inferences from this mixed evidence, we quantitatively reviewed the literature using modern meta-analytic approaches. Based on a large dataset, we investigated the overall relationship between behaviours that are expected to form a syndrome and tested which factors can mediate heterogeneities in study outcomes. The average strength of the phenotypic correlation between behaviours was weak; we found no effect of the phylogeny of species but did observe significant publication bias. However, even accounting for this bias, the mean effect size was positive and statistically different from zero ($r = 0.198$). Effect sizes

L. Z. Garamszegi (✉)
Department of Evolutionary Ecology, Estación Biológica de Doñana–CSIC, c/Americo Vespucio, s/n, 41092 Seville, Spain
e-mail: laszlo.garamszegi@ebd.csic.es

G. Markó · G. Herczeg
Behavioural Ecology Group, Department of Systematic Zoology and Ecology, Eötvös Loránd University, Pázmány Péter sétány 1/c, 1117 Budapest, Hungary

G. Markó
Department of Plant Pathology, Corvinus University of Budapest, Ménesi út 44, 1118 Budapest, Hungary

G. Herczeg
Ecological Genetics Research Unit, Department of Biosciences, University of Helsinki, P.O. Box 65, 00014 Helsinki, Finland

showed considerable heterogeneity within species, implying a role for population-specific adaptation to environmental factors and/or between-study differences in research design. There was a significant positive association between absolute effect size and repeatability of behaviours, suggesting that within-individual variation of behavioural traits can set up an upper limit for the strength of the detected phenotypic correlations. Moreover, spatial overlap between the contexts in which different behaviours were assayed increased the magnitude of the association. The small effect size for the focal relationship implies that a huge sample size would be required to demonstrate a correlation between behaviours with sufficient statistical power, which is fulfilled only in very few studies. This suggests that behavioural syndromes often remain undetected and unpublished. Collectively, our meta-analysis revealed a number of points that might be worth to consider in the future study of behavioural syndromes.

Introduction

Interest in the study of behavioural variation within and between populations has recently turned towards understanding individual consistency of behaviours, which can be addressed both within the same behavioural trait and between different behaviours (Dingemanse et al. 2010b; Dingemanse and Wolf 2010). Consistent behaviours are often interchangeably labelled as "personality" (Gosling 2001), "behavioural syndromes" (Sih et al. 2004a, b), "temperament" (Réale et al. 2007) or "coping style" (Koolhaas et al. 1999), albeit these phenomena may have different meaning. For example, in our understanding, animal personality deals with the consistencies of a *single behavioural trait* across time and context (where context is an environmental gradient) and thus calls for statistical issues about repeatability. On the other hand, we interpret behavioural syndrome as a phenomenon that concerns with the non-independence of *functionally distinct behaviours*, which necessitates the use of correlations between different traits. The current paper focuses on consistent behaviours from this latter sense, thus our main interest is to make implications for behavioural syndromes and not for animal personality.

Evolutionary and behavioural ecologists are particularly interested in behavioural syndrome structure. It is challenging to disentangle the ultimate and proximate mechanisms beyond the emergence of non-independence between such seemingly different behavioural traits as aggression towards conspecifics, risk-taking in the presence of predators and foraging activity in a novel environment (Sih et al. 2004a, b; Sih and Bell 2008). Such coupled behaviours set up important evolutionary constraints because they infer limited phenotypic plasticity: individuals cannot optimally choose their responses in each ecological or social situation that they are confronted with in their life (Wilson 1998). If behaviours are linked, bold individuals will tend to display aggressive responses towards conspecifics as well as towards predators, even if the adaptive values of these responses are opposing (i.e. being bold can be advantageous in one situation but can be costly in another). Accordingly, behavioural syndromes may drive life-history tradeoffs and ultimately determine fitness that can be achieved in different environments (Sih et al. 2004a, b; Sih and Bell 2008).

But why would different behaviours become non-independent from each other? On the proximate level, behavioural syndromes can be explained by the shared physiological or

genetic regulation of different behaviours that acts as a constraint on the independent evolution of behavioural phenotypes (Arnold 1992; Lande 1992). For example, the pleiotropic effect of hormones (Ketterson and Nolan 1999), genes (van Oers et al. 2005a) or neurotransmitters (Bond 2001) can represent such mechanisms. From an ultimate perspective, it is more compelling to understand why the regulation of functionally independent behaviours has not been uncoupled during phylogenetic time-scales. Different hypotheses have been put forward to explain the adaptive value of non-independent behaviours and highlighted those conditions in which the non-flexible combination of behaviours within individuals can be favoured by natural selection. These explanations include (1) the existence of inherently stable differences in individual states (e.g. long-term differences in morphology, physiology or social status) that determine behavioural responses; (2) the existence of labile state differences (e.g. condition, territory quality, experience) with positive feedback mechanisms that conserve consistent differences in time; and (3) models that do not assume state-dependence and variation in state, such as the co-evolution model of responsiveness and behavioural consistency, or models of social conventions or signalling mechanisms (Dall et al. 2004; Dingemanse and Wolf 2010; Wolf and Weissing 2010). Furthermore, a theoretical framework has been developed to account for the coexistence of different behavioural phenotypes within and between populations (Wolf and Weissing 2010; Sih et al. 2012). These evolutionary mechanisms that may maintain natural variation include frequency-dependent selection, spatio-temporal variation in the environment, non-equilibrium dynamics and externally induced differences in states.

Recent studies in evolutionary ecology have started to accumulate empirical evidence to support the evolutionary importance of consistent behaviours. Field and laboratory examples showed that consistent individual differences in behavioural traits such as exploration, aggression and risk-taking can confer fitness benefits and can be shaped by natural selection (Réale et al. 2007; Smith and Blumstein 2008), but can also be linked to sexual selection (Schuett et al. 2009). Others were successful in decoupling additive genetic and environmental components behind the between-individual variation of certain traits (Stirling et al. 2002; van Oers et al. 2005a). Some studies have discovered that genes and (stress) hormones can be involved in the proximate regulation of consistent behaviours (Fidler et al. 2007; Carere et al. 2010). At the interspecific level, comparative studies in birds identified the most important evolutionary determinants of risk-taking and exploration (Mettke-Hofmann et al. 2002; Blumstein 2006; Møller and Garamszegi 2012).

However, most of these facts focus on particular behaviours in isolation from each other. Therefore, our understanding of the patterns of and mechanisms behind behavioural consistency are particularly based on studies of animal personality, while our knowledge about the evolutionary significance of behavioural syndromes lags behind. Most studies of this rarely got over the stage of reporting the presence or absence of phenotypic correlations between behaviours across individuals. By examining different populations, a few of them were able to demonstrate temporal and/or spatial variations in the correlation structure (Bell et al. 2005; Dingemanse et al. 2007; Sinn et al. 2010), but ecological factors, or methodological differences that shape such variance in nature often remain unidentified (Dingemanse and Wolf 2010; Sih et al. 2012). Additionally, research on behavioural syndromes has largely centred on phenotypic correlations, which do not directly prove the existence of trade-offs. A quantitative genetic framework based on genetic correlations would be much better suited for addressing evolutionary mechanisms (Dochtermann and Roff 2010). Another problem with phenotypic correlations is that they are composed of the

sum of the between-individual correlations (i.e. mean trait values correlate with each other across individuals) and within-individual correlations (i.e. within-individual changes in flexible traits correlate across observations) that are mediated by variances in each level (Snijders and Bosker 1999). Although only the former has implications for behavioural syndromes in a sense that individual-specific behaviours are linked, the potential role of within-individual correlations remain undetermined. Most of the studies fail to separate these two components, and their relevance for behavioural syndromes thus remains little understood.

In fact, although the phenotypic association between behaviours is generally expected and often observed, it does not seem to be a universal rule. Papers investigating this pattern often report dissimilar results concerning the relationship between different traits, which currently makes generalizations difficult about the magnitude of behavioural correlations. For instance, in some populations of three-spined stickleback, *Gasterosteus aculeatus*, the correlations between exploration, activity, risk-taking and aggressiveness are different in predator-sympatric and predator-naïve populations (Bell 2005; Dingemanse et al. 2007), while in others, correlations were nearly absent irrespective of population background (Brydges et al. 2008). In birds, some researchers found correlations between lab and wild conditions (Herborn et al. 2010; van Overveld and Matthysen 2010), but others did not (Minderman et al. 2010). In the great tit, *Parus major*, the intensity of alarm call correlated positively with aggression while movement activity and aggression correlated negatively (Hollander et al. 2008). In the same species, a strong correlation between aggression and novel environment exploration was often (Verbeek et al. 1996; Dingemanse and de Goede 2004; Amy et al. 2010) but not always detected (Carere et al. 2005). In the collared flycatcher, *Ficedula albicollis*, the strength of certain behavioural correlations was different in magnitude in different samples (captured, non-captured) of individuals (Garamszegi et al. 2009b).

From these and other studies, two important questions arise. First, do behaviours generally show a non-zero correlation across phenotypes, and if so, in which direction and with what strength? Second, what accounts for the differences in study results? To be able to assess the evolutionary importance of behavioural correlations and to make stronger arguments for natural selection acting on consistent behaviours, it is crucial to answer to these questions. The meta-analytical approach is a useful tool to make a quantitative summary from the published literature that can serve as a statistical basis for theoretical generalizations (Arnquist and Wooster 1995; Lajeunesse and Forbes 2003; Adams 2008; Borenstein et al. 2009). In this approach, study results are converted to a common currency (effect size), which can be combined across studies by weighting them by their precision (sample size or confidence interval) and by correcting for potential biases due to phylogenetic inertia or the dominance of significant results in the published literature. Furthermore, heterogeneities in effect sizes across studies can be estimated, and one can seek biological predictors that affect the variance of effect sizes. Meta-analyses have been applied to explore the relationship between animal personality and fitness (Smith and Blumstein 2008) and to summarize what is known about the repeatability of behaviours (Bell et al. 2009), but no meta-analysis has been conducted on the phenotypic correlation of behaviours, which can have implications for behavioural syndromes.

In this paper, we use both conventional and phylogenetic meta-analyses and determine the general patterns of phenotypic correlations between different, individual-specific behaviours. We rely on a database of more than a hundred independent correlations from more than 50 vertebrate species, which provided considerable statistical power to calculate the average strength of behavioural correlations, assess publication bias, incorporate

phylogeny, and investigate how additional factors influence variation in effect sizes. Specifically, we test whether studies show an overall effect size that is consistent towards the positive or negative direction, and whether effect sizes are dependent on phylogeny and taxonomy. If behavioural syndromes are widespread across animals of different taxa, and correlations are more likely to fall in a certain (e.g. positive) direction than in the other direction, we generally predict that the overall effect size will be statistically distinguishable from zero. Given existing evidence for population differences in syndrome structure and the potential for sensitivity to various ecological factors, we also predict that effect sizes will be heterogeneous across different studies.

In addition, we assess the importance of some potentially relevant mediator variables. For example, we explore if differences in contextual overlap between the conditions in which the behaviours are measured (Sih et al. 2004b) can account for the heterogeneity of effect sizes. We predict that there is a stronger correlation between behaviours due to contextual overlap if they were measured in the same environment (i.e. same experimental room or territory) or in the same physiological state of the individual (i.e. short temporal interval between two behavioural measurement) than in cases where measurements are taken in different environments or physiological states. In a similar vein, we also predict that effect sizes from captivity would be higher than from natural conditions, because the more standardised environment circumstances of captivity would reduce the variations in environmental quality that affect the results (Martin and Réale 2008; Dingemanse et al. 2010b). We also test if effect sizes differ between major age or sex categories or between different seasons under the general prediction that selection would favour stronger or weaker correlations in any particular group, if individuals from a sex, age or seasonal category are subject to different selection regimes than individuals in the other categories, or if belonging to different groups involves different confounding factors that mediate the correlation between behaviours. Finally, we also investigate if publication bias affects our general picture about behavioural syndromes. Since the study of behavioural syndromes is flourishing in recent years, we predict that it might be more interesting (and easy) to publish new and positive results leading to a preponderance of larger effect sizes in the literature.

We note that recent recommendations call for the careful use of the pair-wise phenotypic correlations to make implications for the study behavioural syndromes. This caution is needed due to the non-independence of data in multiple comparisons, the increased type II error rate (Dingemanse et al. 2010a), the inability to decouple within- and between-individual variations if each individual is assayed only once (Dingemanse et al. 2010b) and to separate the additive effects of genetic and environmental correlations as well as between- and within-individual correlations (Roff 1996; Dochtermann 2011). Although behavioural syndromes may be better understood by using more complex statistical frameworks like multivariate mixed-effect models (Dingemanse et al. 2010b; Wilson et al. 2011b) or the combined use of structural equation modelling and information-theoretic approaches (Dochtermann and Jenkins 2007; Dingemanse et al. 2010a), such advanced methodologies are currently not widespread. Therefore, for the sake of generality, we base our quantitative review on the most commonly used approaches based on phenotypic correlations that allowed us to cover a large part of the related literature. However, at the level of analysis, we assess how within-individual variance in behaviour (repeatability) and issues about statistical power (type II errors) can affect the strength of between-individual correlations. Furthermore, at the level of interpretation of phenotypic correlations, we retain certain reservations given that the

Table 1 MCMC modelling of different random effect structures that were built in the meta-analysis models

Random terms	DIC	r	95 % CI (lower/upper)	pMCMC
–	141.9	0.249	0.207/0.286	<0.001
Study	18.4	0.265	0.217/0.313	<0.001
Species	52.11	0.274	0.211/0.336	<0.001
Species (phylogeny)	127.1	0.253	0.068/0.397	0.014
Study + species (phylogeny)	17.7	0.268	0.157/0.377	0.002

different components (e.g. within- and between-individual, genetic and environmental) of the association of traits cannot be distinguished.

Materials and methods

Due to space limitation, the full details of our methodology are presented on the Electronic Supplementary Material. Briefly, we relied on an intense literature search, in which—after establishing careful definitions for the main personality traits sensu Réale et al. (2007)—we gathered information on the strength of association between activity, aggression, exploration and risk-taking in a form of effect size correlation (r) (Cohen 1988). We sorted the available data at different levels (e.g. "study" or "species" levels) to allow hierarchical modelling. For each effect size entry, we assigned the applicable states of the moderator variables (species and higher taxonomic id, captivity, spatial and temporal overlap between the conditions of behavioural measurements, sex, age, season and the geometric mean of the repeatability of traits). We used mixed-effect meta-analytical modelling based on Bayesian quantitative genetic approach to correct for non-random sampling due to different sources as incorporated in the form of random effects (Hadfield and Nakagawa 2010). Our modelling revealed that focusing on the study level as a unit of analysis is reliable (Table 1). Accordingly, we continued our analysis to determine mean effect size, the associated 95 % confidence interval and the heterogeneity across studies by using a dataset that combined effect sizes within studies and so had independent entries with no overlap in the sample of individuals. We assessed the effect of different moderator variables by using them as a grouping variable in an ANOVA-like design or by conducting a meta-regression with a continuous predictor. To investigate if the repeatability of traits has an influence on the strength of the detected correlation between them, we also performed a meta-regression with the geometric mean of the traits' repeatability as a predictor and the absolute effect size as the response variable. We applied Begg's method to identify publication bias (Begg and Mazumdar 1994) and applied the Trimfill algorithm to impute missing data points as caused by publication bias (Duval and Tweedie 2000). After constructing a phylogenetic tree of species (Fig. 1), we conducted phylogenetic meta-analyses (Adams 2008; Lajeunesse 2009) on the species-specific effects sizes, and calculated cumulative effect sizes while accounting for phylogenetic inertia. We generally used random-effects models and performed all meta-analyses tests using Fisher's Z, a normalized score of r (Borenstein 2010). All effect sizes are reported as correlation coefficients (r) with 95 % confidence intervals. In ecological literature, $r \approx 0.1$ is generally considered a small effect, $r \approx 0.3$ a medium effect and $r \approx 0.5$ a strong effect (Cohen 1988; Møller and Jennions 2002). To allow the dynamic development of the underlying database of this study, we uploaded it to www.evolutionary-meta-analysis.net (see Garamszegi et al. 2012).

Results

Hierarchical modelling: within-study and within-species variation

Using the raw data with multiple entries per studies and species, we tested if the non-independence of data at different levels should be of concern. The random-effect modelling based on Bayesian approaches demonstrated that the most important confounding effect to consider was the within-study variance (Table 1). The model that only included "study" as a random factor and assumed no phylogenetic structure returned the lowest Deviance Information Criterion (DIC), implying the best model fit. This model estimated the main effect size for the general correlation between behaviours as 0.265 with a relatively narrow confidence interval. Although there was a relatively large difference between models with different hierarchical structures in terms of DIC, the difference in terms of the mean effect size estimate was minor (Table 1).

When we applied a simple ANOVA to compare the variation of effect sizes within and between studies, we also found that non-independent effect sizes within the same study tended to vary more similarly than independent effect sizes across studies ($F_{104,142} = 2.389$, $P < 0.001$).

Consequently, we found statistical evidence that the overlap in the sample of individuals within the same study can lead to the non-random variation at the within-study level, which should be avoided in the context of the current meta-analysis. On the other hand, phylogenetic inertia in the raw data is less important. Therefore, for the subsequent more complex analyses, we focused on the combined mean effect sizes at the study level that are free of the confounding effect of non-independence and contain useful biological information. These 105 effect sizes are listed in the Electronic Supplementary Material (Table S1).

General trends: correlation between behaviours and publication bias

Relying on the study-specific estimates of the associations between the main behavioural traits that are involved in behavioural syndromes, we found that the mean effect size for the phenotypic correlation between behaviours was 0.264 (95 % CI 0.210–0.316) and differed significantly from a null effect ($Z = 9.254$, $P < 0.0001$). Note that this estimate was almost the same as the one obtained from the raw effect size data by using "study" as a random factor (Table 1) implying that the information loss due to the averaging within studies is negligible. We found significant heterogeneity in effect sizes across studies ($Q = 329.544$, $df = 104$, $P < 0.0001$).

A funnel plot of precision for effect sizes and the associated test for publication bias suggested that null and/or negative results are significantly underrepresented in the published literature (Kendall's tau = 0.126, $P = 0.028$, one-tailed without continuity correction; Fig. 2). To correct for this bias, the trim and fill analysis of the random effects model imputed 15 missing negative studies and reduced the point estimate of r to 0.198, which was still significantly larger than zero, as shown by its confidence interval (95 % CI 0.138–0.257). Therefore, the true correlation between behaviours is likely to be of weak magnitude ($r < 0.25$) but it falls systematically in the positive direction.

Based on the estimated overall effect size of 0.198 and by applying power analysis, we calculated the adequate sample size that would be needed in particular studies when using null hypothesis testing framework to detect the relationship of interest in the form or phenotypic correlation with the power of 80 % (Cohen 1988). This calculation revealed
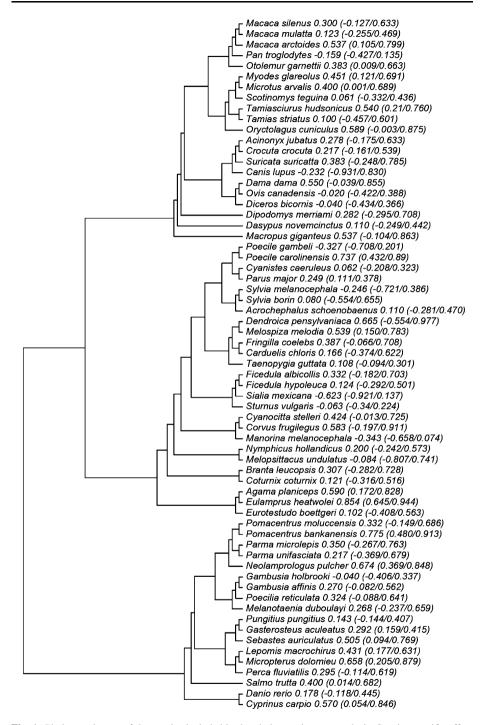
Fig. 1 Phylogenetic tree of the species included in the phylogenetic meta-analysis. Species-specific effect sizes for the correlation between behaviours and the associated 95 % confidence intervals (lower limit/upper limit) are shown next to the species' names

that at least 198 individuals would be required to demonstrate a statistical association between behaviours at a tolerable probability of committing type II error. With the exception of a single publication (Cote et al. 2010), all papers used a smaller sample size (mean ± SE = 44.93 ± 2.01) in tests of correlating behaviours across phenotypes. Among the 105 particular estimates derived, 40 were statistically significant. The remaining effect sizes would correspond to tests that could not reject the null hypothesis of $r = 0$, while in this non-significant set of effect sizes there are still 20 correlations that are actually larger than 0.264 and 26 that are larger than 0.198. Thus ca. 19–25 % of the studies might have misleadingly concluded that there was no phenotypic correlation between behaviours.

Sources of heterogeneity: moderator variables

We examined some biological factors that may have mediated heterogeneity among effect sizes. Among the categorical variables investigated, species, taxonomic class and spatial overlap partitioned effect sizes into groups that can be distinguished as statistically different, and there was a similar but only marginally significant tendency for sex categories (Table 2). The species-specific effects (without considering their phylogenetic relationships) are also evident from the Bayesian modelling of random factors applied to the raw effect size data (Table 1). Group-specific effect sizes for taxonomic classes, spatial overlap categories and sexes are shown in Fig. 3, while species-specific effects are given in Fig. 1. These group averages suggest that fishes and reptiles had effect sizes that are systematically larger than that of birds and mammals. Moreover, studies that perform the behavioural assays in the same environment obtain consistently stronger correlations than studies that score the traits in different circumstances. Finally, males tend to demonstrate stronger effects than females. Note that none of the covariates considered was fully responsible for the heterogeneity of effect sizes across studies, as in each case some within-group heterogeneity was left to be explained by unknown variables (Table 1).
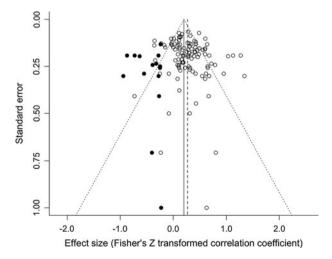

Fig. 2 Funnel plot of effect size (Fisher's Z) by standard error. The *sign* of the effect indicates whether the correlation is positive or negative. *Open circles* represent published studies, and *filled circles* represent missing imputed studies. *Dashed line* represents overall effect size without considering publication bias, *solid line* is for the corrected effect size. *Dotted lines* are for confidence intervals

Table 2 Sources of heterogeneities of effect sizes between and within categories of the tested moderator variables as defined in the "Materials and methods" section of the Electronic Supplementary Material

Factors	N^a	Between group heterogeneity[b]			Within group heterogeneity[c]		
		Q	df	P	Q	df	P
Species	105	115.03	64	<0.0001	83.462	40	<0.0001
Taxonomic class	105	13.597	3	0.004	293.06	101	<0.0001
Spatial overlap	91	4.057	1	0.044	292.09	89	<0.0001
Captivity	105	3.889	4	0.421	307.71	100	<0.0001
Age	96	1.231	2	0.540	288.93	93	<0.0001
Sex	79	5.147	2	0.076	267.50	76	<0.0001
Season	82	1.103	2	0.576	279.02	79	<0.0001

[a] Sample sizes vary because information on moderator variables was lacking in some source studies or because the combination of effect sizes resulted in a mixture of categories thus the moderator value could not be assigned; [b] by using random effects analysis; [c] by using fixed effect analysis
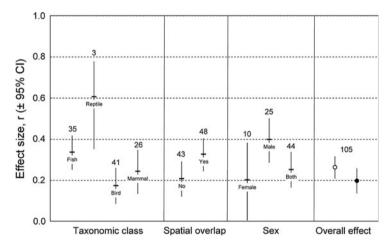

Fig. 3 Mean (*horizontal thick marks*) effect sizes (*r*) and their 95 % confidence intervals (*error bars*) for the general relationship between behaviours when tabulated separately for different taxonomic classes, spatial overlap categories and sexes. The rightmost panel shows effect sizes for the entire data without (*open circles*) and with (*filled circles*) considering and controlling for publication bias. *Numbers* indicate sample sizes in terms of the number of independent studies

To explore how the continuous measure of temporal overlap between behavioural assays affected the strength of the focal relationship, we conducted a meta-regression analysis. Using a mixed-effect regression with unrestricted maximum likelihood, we found a non-significant slope of -0.013 (95 % CI: -0.079–0.053, $Z = -0.399$, $P = 0.690$), showing that effect sizes are likely to be independent of the temporal context.

The meta-analysis of unsigned effect sizes that disregard the direction of the correlations also revealed only a moderate mean effect size ($r = 0.303$, 95 % CI: 0.260–0.346). This estimate was also loaded with publication bias (Kendall's tau = 0.188, $P = 0.002$, one-tailed without continuity correction), and a correction for such a bias returned a mean effect size of 0.257 (95 % CI 0.207–0.306). When we investigated the link between the repeatability of traits and the strength of the detected correlation, we observed a significantly positive slope in the corresponding meta-regression on absolute effect sizes
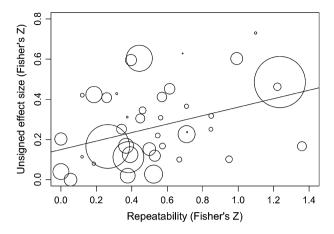
Fig. 4 The relationship between absolute effect size reflecting the strength of the behavioural correlations and the average repeatability of traits involved in the correlation. *Circles* are estimates from independent studies with a volume that is proportional to the corresponding sample size. *Line* is the regression line that is calculated by the meta-regression approach using unrestricted maximum likelihood

(slope $= 0.211$, 95 % CI: 0.033–0.389, $Z = 2.327$, $P = 0.019$, Fig. 4). We infer that this result is unlikely to be confounded by the non-normal distribution of absolute effect sizes, because the association with repeatability was also prevalent when we applied a non-parametric correlation on the raw data ($r_S = 0.334$, N $= 41$, $P = 0.033$). Consequently, we showed that it is more likely to find stronger phenotypic correlation when the repeatability of the traits is high.

Phylogenetic meta-analyses

When using Pagel's λ statistics (Freckleton et al. 2002), we found no significant phylogenetic signal in the distribution of effect sizes for the relationship between behaviours. The estimated λ value was close to zero ($\lambda = 0.144$) and not significantly different from zero (log-likelihood $= 1.035$, $P = 0.309$). When applying phylogenetic meta-analytic approach to the species averaged data we confirmed this finding of no phylogenetic inertia in the data. Traditional meta-analysis methods fit the data better than the phylogenetic meta-analysis (traditional random-effects meta-analysis: Akaike's Information Criterion, AIC $= 124.11$; phylogenetic random-effects meta-analysis: AIC $= 132.11$). Again, this finding is similar to what we obtained in the random-effect modelling from the raw effect size data (Table 1). Although the model fit the data less well, the results of the phylogenetic meta-analysis were consistent with the traditional meta-analysis, with an overall weak positive effect when phylogeny was taken into account (r $= 0.144$, 95 % CI 0.068–0.252). Given the low support for the phylogenetic meta-analysis model, we conclude that the detailed analyses of moderator variables and publication bias in the traditional meta-analysis provide statistically valid results.

Discussion

The existence of behavioural syndromes implies that different behaviours vary non-independently from each other (Sih et al. 2004a, b). So far, different studies have reported widely diverse estimates of the strength of phenotypic correlation between different

behaviours. Here, by using meta-analytical approaches and based on a large sample of species and papers published on the topic, we detected a generally positive phenotypic association between behavioural traits. However, the overall mean effect size (r) for a combined dataset was not particularly strong, even if the sign of the effects was disregarded. In fact, when accounting for publication bias towards positive results, the strength of the relationship is likely to be below 0.2, which can be considered a small effect. On the other hand, the magnitude of the association between two behaviours was dependent on their repeatability showing that low repeatability of traits can introduce a downward bias in the estimate of phenotypic correlation. We could also demonstrate that detected effect sizes are highly study- and species-specific, but there were also consistent differences between different vertebrate groups, as fishes and reptiles had larger effect sizes than birds or mammals. Furthermore, spatial overlap between behavioural assays increased the chance of finding stronger correlations. We also found a marginally significant tendency for sex differences in the strength of the behavioural correlations with males showing stronger correlations than females. These findings have practical and theoretical implications for the study of behavioural syndromes, which we detail below.

Before interpreting the results, we must acknowledge that phenotypic correlations are composed of two components along the following equation:

$$r_{P_y P_z} = r_{ind_y ind_z} \sqrt{\left(\frac{V_{ind_y}}{V_{ind_y} + V_{e_y}}\right)\left(\frac{V_{ind_z}}{V_{ind_z} + V_{e_z}}\right)} + r_{e_y e_z} \sqrt{\left(\frac{V_{e_y}}{V_{ind_y} + V_{e_y}}\right)\left(\frac{V_{e_z}}{V_{ind_z} + V_{e_z}}\right)} \quad (1)$$

where $r_{P_y P_z}$ is the phenotypic correlation, $r_{ind_y ind_z}$ is the between-individual correlation, $r_{e_y e_z}$ is the within-individual correlation between traits y and z; whereas V_{ind_y} and V_{ind_z} are the between-individual variances, and V_{e_y} and V_{e_z} are the within-individual variances of traits y and z, respectively (Snijders and Bosker 1999). Note that the weight associated with the between-individual correlation is the geometric mean of the repeatability of the two traits. In our study, we could estimate the mean effect size for $r_{P_y P_z}$ because this is what is usually presented in the source papers. Unfortunately, based on the available data it is impossible to separate $r_{ind_y ind_z}$ and $r_{e_y e_z}$, which can have different implications for behavioural syndromes. As is done in most of the studies relying on a single observational snapshot of behavioural correlations of traits that are assumed or observed to be repeatable within individuals, for simplicity, we will here assume that phenotypic correlations mostly reflect the correlation between mean trait values that are individual-specific attributes. In support of this assumption, we found that the phenotypic correlation increases with repeatability (Fig. 4). If between- and within-individual correlations both exist and stable across the data included in this meta-analysis, this finding implies that the between-individual correlation is stronger than the within-individual correlation (see Eq. 1: $r_{P_y P_z}$ scales positively with the geometric mean of the repeatability of traits if $r_{e_y e_z} \sim 0$, but $r_{P_y P_z}$ scales positively with the geometric mean of 1-repeatability of traits if $r_{ind_y ind_z} \sim 0$). However, we must remain cautious about generally equating phenotypic correlations with between-individual correlations. Given the lack of direct evidence, the role of within-individual correlation cannot fully be factored out, as in some particular cases it can be larger than zero (e.g. Bell and Stamps 2004).

The strength of the main effect

The key result that the general phenotypic correlation between behaviours was not particularly strong could signify that behaviours are subject to considerable within-individual

variations, and the generally low repeatability of traits sets up an upper limit for the maximal strength of the relationships. Low repeatability may involve the possibility that behaviours change flexibly within individuals due to biological reasons or that they all behave in the same way, but also that these traits can be scored only with substantial error (Nakagawa and Schielzeth 2010). Such possibilities can be separated by estimating the repeatability of traits over time and by testing for the inter-observer agreement in the assessment of the same behaviour. Studies calculating the consistency of the same behaviour over multiple sampling sessions tend to show modest repeatability. In a meta-analysis of the repeatability of behaviours Bell et al. (2009) showed that on average behaviours demonstrate repeatabilities of r ∼0.4, which was also the case in the current study (r = 0.477, 95 % CI 0.407–0.547, N = 41), and which is considerably smaller than that of physiological traits (e.g. Nespolo and Franco 2007 detected r ∼0.75 for metabolic rate). In addition, differences in measurement errors in the assessment of behaviours cannot be ruled out. For example, although there was a good concordance between the ratings of different observers of 44 behavioural traits in a study of the Cockatiels, *Nymphicus hollandicus*, no trait showed perfect agreement across all scorers (Fox and Millam 2010). In any case, our results show that it is important to consider the repeatability of traits when interpreting the phenotypic correlation between them.

The weak phenotypic correlation may also be explained by the possibility that the genetic correlation is strong, but the behaviours have low heritability and this undermines the strength of phenotypic correlations (Roff 1997; Stirling et al. 2002). Risk-taking and exploration have both been shown to have modest heritability (Dingemanse et al. 2002; Drent et al. 2003; van Oers et al. 2004b), and in such cases, phenotypic correlations often underestimate genetic correlations (Kruuk et al. 2008; Dochtermann and Roff 2010; Dochtermann 2011). However, based on the available data, we cannot exclude the scenario that generally weak phenotypic correlations emerge because the underlying genetic linkage between behaviours is also weak. Therefore, it is equally plausible that the constraints for the coupled evolution of traits are relaxed and this raises low between-individual correlation that is manifested in the strength of the phenotypic correlation. Accordingly, some individuals can behave differently in different ecological situations (for example, some will respond aggressively to a social challenge but will react in a risk-averse manner when faced with a predator). Such scenarios call for the syndrome deviation concept (Herczeg and Garamszegi 2012) that emphasizes the importance of individuals that apparently do not submit themselves to behavioural syndromes (i.e. their behaviours are not predictable from each other). As such, individual-specific consistencies in the performance of different behaviours might also be the subject of selection. Finally, it is also possible that the emergence of phenotypic correlations is fully environmentally induced and can result from individual experience, learning or physiological state.

The generally weak effect size may also result from the considerable variance in effect sizes among studies. A few studies find strong effects while most others report weak effects, and the large amount of studies detecting negative evidence (i.e. close to null effects) may balance positive findings. Moreover, different mechanisms may operate in different species or populations or different roles may apply to different behaviours, and such effects are superimposed on each other in the combined dataset we used in our analyses. For example, we showed systematic differences in effect size between species and larger taxa, but even after controlling for such factors, a considerable amount of heterogeneity remained, leaving possibilities open for adaptations at a population level (we note that the large variation between species might have been a bias stemming from the fact that most species were only represented by a single population in our dataset). For

example, effect sizes varied largely within the two most intensively studied model species. We could obtain 11 independent effect sizes for the behavioural correlations from five different populations of the great tit, which spanned from −0.223 to 0.572. In a similar vein, the 11 effect sizes ranged between 0.037 and 0.810 across eight populations of the three-spined stickleback. Local differences can be explained by local divergence in genetic or other constraints, or by environment-driven effects without genetic adaptations, but the role of random drift cannot be ruled out (Whitlock et al. 2002; Walsh and Blows 2009). Unfortunately, these possibilities are hard to tackle, because the use of phenotypic correlations based on wild-caught individuals or even F1 laboratory generations does not allow separating genetic versus environmental effects (Kruuk et al. 2008; Dochtermann and Roff 2010; Dochtermann 2011). Moreover, instead of true between-population differences, within-species variation can also be caused by differences in the methodologies between research groups studying different populations of the same species. Given that we detected a role for spatial overlap of assay conditions in mediating the detected relationship, we can infer that differences in the experimental setups may also have considerable importance driving heterogeneity in research finding. The weak phylogenetic signal in the data also suggests a large degree of variation in the strength of the relationship between different traits. The variation that can be caused by behaviour-specific roles and inconsistencies in the pair-wise correlation of traits are discussed in detail elsewhere (L. Z. Garamszegi, G. Markó, G. Herczeg unpublished manuscript). Finally, studies may vary with regard to importance of between- and within-individual correlations. If in some cases, they have different signs but are weighted similarly, their opposing effect will cause small phenotypic correlation to be measured (see Eq. 1).

The small effect size we detected also has practical implications, which should be considered for the interpretations of detected correlations as well (Nakagawa and Cuthill 2007). Our survey indicates that because of limited sample sizes, 19–25 % of studies found statistically non-significant associations between behavioural traits, which cover magnitudes that mights have biological significance (i.e. higher than the mean effect size we found here, but see also Møller and Jennions 2002 for a more general discussion). Therefore, based on a null hypothesis testing approach, a large proportion of published studies may have misleadingly concluded that there is no phenotypic correlation between behaviours when there is actually a modest relationship between them that only appears statistically non-significant because of the limited power involved in the available sample (see also Dingemanse et al. 2010a). However, such modest relationships might well have biological relevance, which must be reported and appropriately discussed. Given that we detected evidence for a publication bias, we assume that the problem is even more pronounced when considering the unpublished fraction of studies with weak, null or negative effects. Hence, we advise refraining from making strong inferences for the behavioural syndrome concept from small sample sizes by focusing merely on the significance of correlations, and from arriving at conclusions as being "all or nothing" effects (Garamszegi et al. 2009a). Based on such statistical considerations and the potential relevance of syndrome deviation (Herczeg and Garamszegi 2012), we recommend presenting and discussing behavioural correlations based on the magnitude of the effect along a continuous scale, while the precision of the estimate can be emphasized in the associated confidence intervals and not in the P values. Given the potential importance of within-individual variations, such presentations of effect sizes should also be accompanied by the reports of repeatability.

The direction of the main effect

The overall mean of the effects size, as well as the boundaries of the 95 % confidence intervals, were positive. In fact, 195 of the 250 correlations entered in our raw database were positive, while in theory, behavioural correlations are as likely to be negative if evolutionary explanations apply, as selection can favour the coupled evolution of two traits in any direction if such configuration is adaptive (Bell 2005). Negative correlations may also be beneficial in certain environmental conditions. From the mechanistic perspective, physiological constraints can also act antagonistically and in turn can drive negative correlations, if a the same hormone/gene has opposing effect on different behaviours (e.g. while testosterone usually correlates positively with aggression as reviewed by Soma 2006; it is negatively associated with exporation in van Oers et al. 2011). From these aspects, the preponderance of positive effect sizes is surprising. A potential proximate explanation of this pattern is the presence of temperament-like drives behind behaviours (Buss et al. 1987; Clark and Wilson 1999; Gosling 2001). Accordingly, behavioural consistency both within and between behaviours can be caused by innate characteristics of individual that affect multiple aspects of behavioural phenotypes during life. That latent factor can be manifested in the form of systematic shy or bold behaviours in different ecological situations, and thus mediate positive correlations. Evolutionarily, the existence of temperament would call for mechanisms based on stable states and state-dependent behaviours (Wolf and Weissing 2010). However, we would be careful with the general extension of such explanations, because we focused on behavioural traits that are particularly studied in the "personality" literature (e.g. activity, aggression, exploration, risk-taking), and it might be that our sample is not representative to behavioural correlations in a more general context when not only "personality" traits (sensu Réale et al. 2007) are considered (e.g. a wide range of spatial and social behaviours, courtship behaviour etc.).

The role of mediator variables

Besides these general patterns regarding the statistics of mean effect size, we also tested a series of factors that might have an influence on study outcome. As an important confounder, we investigated whether the contextual overlap between the situations in which the behaviours were assayed had such an influence. If particular behaviours were scored across temporal or spatial conditions that largely coincide with each other, one would expect stronger relationship between traits due to the contextual similarity of the experimental circumstances. Moreover, effect size from captivity experiments should be stronger than in natural conditions, because the former setup offers a more standardised environment than the latter setup, and differences in environmental quality thus do not affect the results (Martin and Réale 2008; Dingemanse et al. 2010b). The predictions in association with the spatial overlap between assay conditions received support in our meta-analysis: we found higher effects sizes when behaviours were measured in spatially overlapping conditions than for non-overlapping conditions. This further implies that the strength of several behavioural correlations is shifted towards the weaker magnitudes in natural conditions, because individuals live in a broad range of different contexts and perform behaviours in different spatial compartments. Somewhat surprisingly, we did not get a similar effect of temporal overlap and captivity: the time interval between measurements and whether the traits were recorded in captive or wild conditions did not affect the strength of the correlation between behaviours.

We also detected sex-specific effects, as females apparently display weaker correlations than males. The marginally significant difference between sexes appears two-fold, which requires some explanations. We suggest that perhaps sex-specific behaviours during reproduction may cause some differences in effect sizes. During reproduction, males usually perform more risky behaviours or bear more conspicuous coloration, which might attract the interest of predators or parasite vectors putting them under a higher selection pressure due to predation or parasitism (Møller and Nielsen 2006; Møller et al. 2006). These sex-specific selection regimes can lead to the higher adaptive value of fitting to behavioural correlations in males than in females. Alternatively, the repeatability of behaviours (i.e. within-individual consistency) may also vary in a sex-specific manner, and the stronger phenotypic correlation between behaviours found in males may in fact be due to the higher intra-individual consistency of traits (Jenkins 2011). Accordingly, most studies that tested for within-individual consistency in a sex-specific manner found that male behaviour is more repeatable than female behaviour (Schuett et al. 2009). At any rate, this result suggest that pooling the sexes is not necessarily a good idea, and we indeed recommend the study of sex-specific behavioural syndromes given that such comparisons within population might provide important insights on how behavioural syndromes emerge/evolve.

Conclusion

Taken together, in our synthetic analysis, we found overall support for a general phenotypic association between behaviours. We demonstrated that the average strength of the relationship between traits is relatively small, but such modest correlations can have important theoretical and practical implications. The majority of the correlations were positive, suggesting that the same individual-specific state can be cause behaviours to vary in the same direction. However, the magnitude of the detected effect depends on the repeatability of traits, thus the interpretations of phenotypic correlations might be different when the within- and between-individual variances are considered. Furthermore, we were able to reveal that the association between traits varies consistently both within and across species, and even across higher taxonomical categories. We also found that contextual overlap and sex both have an effect on the strength of the relationship between behavioural traits, indicating the complexity behind correlating behaviours and the challenge in their detection. In the future, more fine-tuned analyses of multiple predictors and various behavioural traits are required to identify the confounding factors affecting the strength and direction of phenotypic correlations that remained unidentified in the present study. Moreover, research is required to help disentange the within- and between-individual as well as the genetic and environmental components of correlations and their implications for behavioural syndromes.

References (main text and Electronic Supplementary Material)

Adams DC (2008) Phylogenetic meta-analysis. Evolution 62:567–572

Adriaenssens B, Johnsson JI (2011) Shy trout grow faster: exploring links between personality and fitness-related traits in the wild. Behav Ecol 22:135–143

Amy M, Sprau P, de Goede P, Naguib M (2010) Effects of personality on territory defence in communication networks: a playback experiment with radio-tagged great tits. Proc R Soc B Biol Sci 277:3685–3692

Ancona KA, Loughry WJ (2009) Time budgets of wild nine-banded armadillos. Southeast Nat 8:587–598

Anestis SF (2005) Behavioral style, dominance rank, and urinary cortisol in young chimpanzees (*Pan troglodytes*). Behaviour 142:1245–1268

Arnold SJ (1992) Constraints on phenotypic evolution. Am Nat 140:S85–S107

Arnold KE, Owens IPF, Goldizen AW (2005) Division of labour within cooperatively breeding groups. Behaviour 142:1577–1590

Arnquist G, Wooster D (1995) Meta-analysis: synthesizing research findings in ecology and evolution. Trends Ecol Evol 10:236–240

Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. Biometrics 50:1088–1101

Bell AM (2005) Behavioural differences between individuals and two populations of stickleback (*Gasterosteus aculeatus*). J Evol Biol 18:464–473

Bell AM, Sih A (2007) Exposure to predation generates personality in threespined sticklebacks (*Gasterosteus aculeatus*). Ecol Lett 10:828–834

Bell AM, Stamps JA (2004) Development of behavioural differences between individuals and populations of sticklebacks, *Gasterosteus aculeatus*. Anim Behav 68:1339–1348

Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behaviour: a meta-analysis. Anim Behav 77:771–783

Bell AM, Henderson L, Huntingford FA (2010) Behavioral and respiratory responses to stressors in multiple populations of three-spined sticklebacks that differ in predation pressure. J Comp Physiol B 180:211–220

Bergvall UA, Schapers A, Kjellander P, Weiss A (2011) Personality and foraging decisions in fallow deer, *Dama dama*. Anim Behav 81:101–112

Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A (2007) The delayed rise of present-day mammals. Nature 446:507–512

Biro PA, Beckmann C, Stamps JA (2010) Small within-day increases in temperature affects boldness and alters personality in coral reef fish. Proc R Soc B Biol Sci 277:71–77

Blumstein DT (2006) Developing an evolutionary ecology of fear: how life history and natural history traits affect disturbance tolerance in birds. Anim Behav 71:389–399

Bond AJ (2001) Neurotransmitters, temperament and social functioning. Eur Neuropsychopharmacol 11:261–274

Boogert NJ, Reader SM, Laland KN (2006) The relation between social rank, neophobia and individual learning in starlings. Anim Behav 72:1229–1239

Boon AK, Reale D, Boutin S (2008) Personality, habitat use, and their consequences for survival in North American red squirrels *Tamiasciurus hudsonicus*. Oikos 117:1321–1328

Borenstein M (2010) Comprehensive meta analysis-a computer program. Version 2.2.055. Biostat Inc., Englewood

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Introduction to meta-analysis. Wiley, West Sussex

Brydges NM, Colegrave N, Heathcote RJP, Braithwaite VA (2008) Habitat stability and predation pressure affect temperament behaviours in populations of three-spined sticklebacks. J Anim Ecol 77:229–235

Buss AH, Chess S, Goldsmith HH, Hinde RA, McCall RB, Plomin R, Rothbart MK, Thomas A (1987) What is temperament: four approaches. Child Dev 58:505–529

Callicrate TE, Siewerdt F, Koutsos E, Estevez I (2011) Personality traits and the effects of DHA supplementation in the budgerigar (*Melopsittacus undulatus*). Appl Anim Behav Sci 130:124–134

Carere C, Drent PJ, Privitera L, Koolhaas JM, Groothuis TGG (2005) Personalities in great tits, *Parus major*: stability and consistency. Anim Behav 70:795–805

Carere C, Caramaschi D, Fawcett TW (2010) Covariation between personalities and individual differences in coping with stress: converging evidence and hypotheses. Curr Zool 56:728–740

Carlstead K, Mellen J, Kleiman DG (1999) Black rhinoceros (*Diceros bicornis*) in US zoos: I. Individual behavior profiles and their relationship to breeding success. Zoo Biol 18:17–34

Carter AJ, Pays O, Goldizen AW (2009) Individual variation in the relationship between vigilance and group size in eastern grey kangaroos. Behav Ecol Sociobiol 64:237–245

Carter AJ, Goldizen AW, Tromp SA (2010) Agamas exhibit behavioral syndromes: bolder males bask and feed more but may suffer higher predation. Behav Ecol 21:655–661

Clark LA, Wilson D (1999) Temperament: a new paradigm for trait psychology. In: Pervin LA, John OP (eds) Handbook of personality: theory and research. The Guilford Press, New York, pp 399–423

Cohen J (1988) Statistical power analysis for the behavioural sciences, 2nd edn. Lawrence Erlbaum Associates, Hillsdale

Cohen J (1992) A power primer. Psychol Bull 112:155–159

Cole EF, Cram DL, Quinn JL (2011) Individual variation in spontaneous problem-solving performance among wild great tits. Anim Behav 81:491–498

Colléter M, Brown C (2011) Personality traits predict hierarchy rank in male rainbowfish social groups. Anim Behav 81:1231–1237

Cote J, Fogarty S, Weinersmith K, Brodin T, Sih A (2010) Personality traits and dispersal tendency in the invasive mosquitofish (*Gambusia affinis*). Proc R Soc B Biol Sci 277:1571–1579

Crino OL, Larkin I, Phelps SM (2010) Stress coping styles and singing behavior in the short-tailed singing mouse (*Scotinomys teguina*). Horm Behav 58:334–340

Dall SRX, Houston AI, McNamara JM (2004) The behavioural ecology of personality: consistent individual differences from an adaptive perspective. Ecol Lett 7:734–739

Davis KE (2008) Reweaving the tapestry: a supertree of birds. PhD dissertation, University of Glasgow, Glasgow

DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Controlled Clin Trials 7:177–188

Dingemanse NJ, de Goede P (2004) The relation between dominance and exploratory behavior is context-dependent in wild great tits. Behav Ecol 15:1023–1030

Dingemanse NJ, Wolf M (2010) A review of recent models for adaptive personality differences. Philos Trans R Soc Lond B Biol Sci 365:3947–3958

Dingemanse NJ, Both C, Drent PJ, Oers KV, Noordwijk AJV (2002) Repeatability and heritability of exploratory behaviour in great tits from the wild. Anim Behav 64:929–938

Dingemanse NJ, Wright J, Kazem AJN, Thomas DK, Hickling R, Dawnay N (2007) Behavioural syndromes differ predictably between 12 populations of three-spined stickleback. J Anim Ecol 76:1128–1138

Dingemanse NJ, Dochtermann NA, Wright J (2010a) A method for exploring the structure of behavioural syndromes to allow formal comparison within and between datasets. Anim Behav 79:439–450

Dingemanse NJ, Kazem AJN, Reale D, Wright J (2010b) Behavioural reaction norms: animal personality meets individual plasticity. Trends Ecol Evol 25:81–89

Dochtermann NA (2011) Testing Cheverud's conjecture for behavioral correlations and behavioral syndromes. Evolution 65:1814–1820

Dochtermann NA, Jenkins SH (2007) Behavioural syndromes in Merriam's kangaroo rats (*Dipodomys merriami*): a test of competing hypotheses. Proc R Soc B Biol Sci 274:2343–2349

Dochtermann NA, Roff DA (2010) Applying a quantitative genetics framework to behavioural syndrome research. Philos Trans R Soc Lond B Biol Sci 365:4013–4020

Drent PJ, van Oers K, van Noordwijk AJ (2003) Realized heritability of personalities in the great tit (*Parus major*). Proc R Soc B Biol Sci 270:45–51

Duckworth RA (2006) Behavioral correlations across breeding contexts provide a mechanism for a cost of aggression. Behav Ecol 17:1011–1019

Dunn JC, Cole EF, Quinn JL (2011) Personality and parasites: sex-dependent associations between avian malaria infection and multiple behavioural traits. Behav Ecol Sociobiol 65:1459–1471

Duval S, Tweedie R (2000) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56:455–463

Dzieweczynski TL, Crovo JA (2011) Shyness and boldness differences across contexts in juvenile three-spined stickleback *Gasterosteus aculeatus* from an anadromous population. J Fish Biol 79:776–788

English S (2008) Individual variation in cooperative behaviour in meerkats. PhD thesis, University of Cambridge, Cambridge

Eriksson AC, Booth DJ, Biro PA (2010) 'Personality' in two species of temperate damselfish. Mar Ecol Prog Ser 420:273–276

Evans J, Boudreau K, Hyman J (2010) Behavioural syndromes in urban and rural populations of song sparrows. Ethology 116:588–595

Feenders G, Klaus K, Bateson M (2011) Fear and exploration in European starlings (*Sturnus vulgaris*): a comparison of hand-reared and wild-caught birds. PLoS One 6:e19074

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125:1–15

Fidler AE, van Oers K, Drent PJ, Kuhn S, Mueller JC, Kempenaers B (2007) Drd4 gene polymorphisms are associated with personality variation in a passerine bird. Proc R Soc B Biol Sci 274:1685–1691

Fox RA, Millam JR (2010) The use of ratings and direct behavioural observation to measure temperament traits in cockatiels (*Nymphicus hollandicus*). Ethology 116:59–75

Fox RA, Ladage LD, Roth TC, Pravosudov VV (2009) Behavioural profile predicts dominance status in mountain chickadees, *Poecile gambeli*. Anim Behav 77:1441–1448

Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetic analysis and comparative data: a test and review of evidence. Am Nat 160:712–726

Gabriel PO, Black JM (2010) Behavioural syndromes in Steller's jays: the role of time frames in the assessment of behavioural traits. Anim Behav 80:689–697

Garamszegi LZ, Eens M, Török J (2008) Birds reveal their personality when singing. PLoS One 3:e2647

Garamszegi LZ, Calhim S, Dochtermann N, Hegyi G, Hurd PL, Jørgensen C, Kutsukake N, Lajeunesse MJ, Pollard KA, Schielzeth H, Symonds MRE, Nakagawa S (2009a) Changing philosophies and tools for statistical inferences in behavioral ecology. Behav Ecol 20:1363–1375

Garamszegi LZ, Eens M, Török J (2009b) Behavioural syndromes and trappability in free-living collared flycatchers, *Ficedula albicollis*. Anim Behav 77:803–812

Garamszegi LZ, Nunn CL, McCabe CM (2012) Informatics approaches to develop dynamic meta-analyses. Evol Ecol. doi:10.1007/s10682-012-9592-0

Gosling SD (1998) Personality dimensions in spotted hyenas (*Crocuta crocuta*). J Comp Psychol 112:107–118

Gosling SD (2001) From mice to men: what can we learn about personality from animal research? Psychol Bull 127:45–86

Greenberg R (1984) Neophobia in the foraging-site selection of a neotropical migrant bird: an experimental study. Proc Natl Acad Sci USA 81:3778–3780

Groothuis TGG, Carere C (2005) Avian personalities: characterization and epigenesis. Neurosci Biobehav Rev 29:137–150

Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J Evol Biol 23:494–508

Harvey EM, Freeberg TM (2007) Behavioral consistency in a changed social context in Carolina chickadees. J Gen Psychol 134:229–245

Herborn KA, Macleod R, Miles WTS, Schofield ANB, Alexander L, Arnold KE (2010) Personality in captivity reflects personality in the wild. Anim Behav 79:835–843

Herborn KA, Coffey J, Larcombe SD, Alexander L, Arnold KE (2011) Oxidative profile varies with personality in European greenfinches. J Exp Biol 214:1732–1739

Herczeg G, Garamszegi LZ (2012) Individual deviation from behavioural correlations: a simple approach to study the evolution of behavioural syndromes. Behav Ecol Sociobiol 66:161–169

Herczeg G, Gonda A, Merilä J (2009) Predation mediated population divergence in complex behaviour of nine-spined stickleback (*Pungitius pungitius*). J Evol Biol 22:544–552

Highfill L, Hanbury D, Kristiansen R, Kuczaj S, Watson S (2010) Rating vs. coding in animal personality research. Zoo Biol 29:509–516

Hollander FA, Van Overveld T, Tokka I, Matthysen E (2008) Personality and nest defence in the great tit (*Parus major*). Ethology 114:405–412

Huntingford FA (1976) The relationship between anti-predator behaviour and aggression among conspecifics in the three-spined stickleback, *Gasterosteus aculeatus*. Anim Behav 24:245–260

Huntingford FA, Andrew G, Mackenzie S, Morera D, Coyle SM, Pilarczyk M, Kadri S (2010) Coping strategies in a strongly schooling fish, the common carp *Cyprinus carpio*. J Fish Biol 76:1576–1591

Jenkins SH (2011) Sexdifferences in repeatability of food-hoarding behaviour of kangaroorats. Anim Behav 81:1155–1162

Jennions MD, Møller AP (2002) Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method. Biol Rev 77:211–222

Ketterson ED, Nolan VJ (1999) Adaptation, exaptation, and constraint: a hormonal perspective. Am Nat 154:S4–S25

Kingsolver JG, Diamond SE, Siepielski AM, Carlson SM (2012) Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. Evol Ecol. doi:10.1007/s10682-012-9563-5

Koolhaas JM, Korte SM, De Boer SF, Van Der Vegt BJ, Van Reenen CG, Hopster H, De Jong IC, Ruis MAW, Blokhuis HJ (1999) Coping styles in animals: current status in behavior and stress-physiology. Neurosci Biobehav Rev 23:925–935

Korpela K, Sundell J, Ylonen H (2011) Does personality in small rodents vary depending on population density? Oecologia 165:67–77

Kruuk LEB, Slate J, Wilson AJ (2008) New answers for old questions: the evolutionary quantitative genetics of wild animal populations. Annu Rev Ecol Evol Syst 39:525–548

Kurvers R, Eijkelenkamp B, van Oers K, van Lith B, van Wieren SE, Ydenberg RC, Prins HHT (2009) Personality differences explain leadership in barnacle geese. Anim Behav 78:447–453

Lajeunesse MJ (2009) Meta-analysis and the comparative phylogenetic method. Am Nat 174:369–381

Lajeunesse MJ (2011) phyloMeta: a program for phylogenetic comparative analyses with meta-analysis. Bioinformatics 27:2603–2604

Lajeunesse MJ, Forbes MR (2003) Variable reporting and quantitative reviews: a comparison of three meta-analytical techniques. Ecol Lett 6:448–454

Lande R (1992) Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. Evolution 46:381–389

Lantová P, Šíchová K, Sedláček F, Lanta V (2011) Determining behavioural syndromes in voles—the effects of social environment. Ethology 117:124–132

Lee JSF, Bereijikian BA (2008) Stability of behavioral syndromes but plasticity in individual behavior: consequences for rockfish stock enhancement. Environ Biol Fish 82:179–186

Lessells CM, Boag PT (1987) Unrepeatable repeatabilities: a common mistake. Auk 104:116–121

Li CH, Lu GQ, Orti G (2008) Optimal data partitioning and a test case for ray-finned fishes (*Actinopterygii*) based on ten nuclear loci. Syst Biol 57:519–539

Light RJ, Pillemer DB (1984) Summing up: the science of reviewing research. Harvard University Press, Cambridge

Light RJ, Singer JD, Willett JB (1994) The visual presentation and interpretation of meta-analyses. In: Cooper H, Hedges LV (eds) The handbook of research synthesis. Russell Sage Foundation, New York, pp 439–445

MacDonald K (1983) Stability of individual-differences in behavior in a litter of wolf cubs (*Canis lupus*). J Comp Psychol 97:99–106

Maddison DR, Schulz KS, Maddison WP (2007) The tree of life web project. Zootaxa 1668:19–40

Mafli A, Wakamatsu K, Roulin A (2011) Melanin-based coloration predicts aggressiveness and boldness in captive eastern Hermann's tortoises. Anim Behav 81:859–863

Magnhagen C (2007) Social influence on the correlation between behaviours in young-of-the-year perch. Behav Ecol Sociobiol 61:525–531

Mainwaring MC, Beal JL, Hartley IR (2011) Zebra finches are bolder in an asocial, rather than social, context. Behav Process 87:171–175

Marchetti C, Zehtindjiev P (2009) Migratory orientation of sedge warblers (*Acrocephalus schoenobaenus*) in relation to eating and exploratory behaviour. Behav Process 82:293–300

Martin JGA, Réale D (2008) Temperament, risk assessment and habituation to novelty in eastern chipmunks, *Tamias striatus*. Anim Behav 75:309–318

Martins TLF, Roberts ML, Giblin I, Huxham R, Evans MR (2007) Speed of exploration and risk-taking behavior are linked to corticosterone titres in zebra finches. Horm Behav 52:445–453

Mettke-Hofmann C, Winkler H, Leisler B (2002) The significance of ecological factors for exploration and neophobia in parrots. Ethology 108:249–272

Mettke-Hofmann C, Ebert C, Schmidt T, Steiger S, Stieb S (2005) Personality traits in resident and migratory warbler species. Behaviour 142:1357–1375

Miller KA, Mench JA (2005) The differential effects of four types of environmental enrichment on the activity budgets, fearfulness, and social proximity preference of Japanese quail. Appl Anim Behav Sci 95:169–187

Miller KA, Garner JP, Mench JA (2005) The test-retest reliability of four behavioural tests of fearfulness for quail: a critical evaluation. Appl Anim Behav Sci 92:113–127

Minderman J, Reid JM, Hughes M, Denny MJH, Hogg S, Evans PGH, Whittingham MJ (2010) Novel environment exploration and home range size in starlings *Sturnus vulgaris*. Behav Ecol 21:1321–1329

Møller AP, Garamszegi LZ (2012) Between individual variation in risk taking behavior and its life history consequences. Behav Ecol 23:843–853

Møller AP, Jennions MD (2001) Testing and adjusting for publication bias. Trends Ecol Evol 16:580–586

Møller AP, Jennions MD (2002) How much variance can be explained by ecologists and evolutionary biologists. Oecologia 132:492–500

Møller AP, Nielsen JT (2006) Prey vulnerability in relation to sexual coloration of prey. Behav Ecol Sociobiol 60:227–233

Møller AP, Nielsen JT, Garamszegi LZ (2006) Song post exposure, song features and predation risk. Behav Ecol 17:155–163

Moretz JA, Martins EP, Robison BD (2007) Behavioral syndromes and the evolution of correlated behavior in zebrafish. Behav Ecol 18:556–562

Naguib M, Kazek A, Schaper SV, van Oers K, Visser ME (2010) Singing activity reveals personality traits in great tits. Ethology 116:763–769

Nakagawa S, Cuthill IC (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev 82:591–605

Nakagawa S, Schielzeth H (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biol Rev 85:935–956

Nespolo RF, Franco M (2007) Whole-animal metabolic rate is a repeatable trait: a meta-analysis. J Exp Biol 210:2000–2005

Nilsson ALK, Nilsson JA, Alerstam T, Backman J (2010) Migratory and resident blue tits *Cyanistes caeruleus* differ in their reaction to a novel object. Naturwissenschaften 97:981–985

Pagel M (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. P Proc R Soc B Biol Sci 255:37–45

Pagel M (1997) Inferring evolutionary processes from phylogenies. Zool Scr 26:331–348

Palmer AR (1999) Detecting publication bias in meta-analyses: a case study of fluctuating asymmetry and sexual selection. Am Nat 154:220–233

Patterson LD, Schulte-Hostedde AI (2011) Behavioural correlates of parasitism and reproductive success in male eastern chipmunks, *Tamias striatus*. Anim Behav 81:1129–1137

Quinn JL, Cresswell W (2005) Personality, anti-predation behaviour and behavioural plasticity in the chaffinch *Fringilla coelebs*. Behaviour 142:1377–1402

Réale D, Gallant BY, Leblanc M, Festa-Bianchet M (2000) Consistency of temperament in bighorn ewes and correlates with behaviour and life history. Anim Behav 60:589–597

Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ (2007) Integrating animal temperament within ecology and evolution. Biol Rev 82:291–318

Rödel HG, Starkloff A, Bautista A, Friedrich AC, von Holst D (2008) Infanticide and maternal offspring defence in European rabbits under natural breeding conditions. Ethology 114:22–31

Roff DA (1996) The evolution of genetic correlations: an analysis of patterns. Evolution 50:1392–1403

Roff DA (1997) Evolutionary quantitative genetics. Chapman and Hall, New York

Rouff JH, Sussman RW, Strube MJ (2005) Personality traits in captive lion-tailed macaques (*Macaca silenus*). Am J Primatol 67:177–198

Ruuskanen S, Laaksonen T (2010) Yolk hormones have sex-specific long-term effects on behavior in the pied flycatcher (*Ficedula hypoleuca*). Horm Behav 57:119–127

Santillan-Doherty AM, Munoz-Delgado J, Arenas R, Marquez A, Cortes J (2006) Reliability of a method to measure novelty-seeking in nonhuman primates. Am J Primatol 68:1098–1113

Scheid C, Noe R (2010) The performance of rooks in a cooperative task depends on their temperament. Anim Cogn 13:545–553

Schielzeth H, Bolund E, Kempenaers B, Forstmeier W (2011) Quantitative genetics and fitness consequences of neophilia in zebra finches. Behav Ecol 22:126–134

Schuett W, Tregenza T, Dall SRX (2009) Sexual selection and animal personality. Biol Rev 85:217–246

Schuett W, Dall SRX, Royle NJ (2011) Pairs of zebra finches with similar 'personalities' make better parents. Anim Behav 81:609–618

Schürch R, Heg D (2010) Life history and behavioral type in the highly social cichlid *Neolamprologus pulcher*. Behav Ecol 21:588–598

Sih A, Bell AM (2008) Insights for behavioral ecology from behavioral syndromes. Adv Study Behav 38:227–281

Sih A, Bell A, Johnson JC (2004a) Behavioral syndromes: an ecological and evolutionary overview. Trends Ecol Evol 19:372–378

Sih A, Bell AM, Johnson JC, Ziemba RE (2004b) Behavioral syndromes: an integrative overview. Q Rev Biol 79:241–277

Sih A, Cote J, Evans M, Fogarty S, Pruitt J (2012) Ecological implications of behavioural syndromes. Ecol Lett 15:278–289

Sinn DL, Moltschaniwskyj NA, Wapstra E, Dall SRX (2010) Are behavioral syndromes invariant? Spatiotemporal variation in shy/bold behavior in squid. Behav Ecol Sociobiol 64:693–702

Smith BR, Blumstein DT (2008) Fitness consequences of personality: a meta-analysis. Behav Ecol 19:448–455

Smith BR, Blumstein DT (2010) Behavioral types as predictors of survival in Trinidadian guppies (*Poecilia reticulata*). Behav Ecol 21:919–926

Smith KL, Miner JG, Wiegmann DD, Newman SP (2009) Individual differences in exploratory and anti-predator behaviour in juvenile smallmouth bass (*Micropterus dolomieu*). Behaviour 146:283–294

Snijders TAB, Bosker RJ (1999) Multilevel analysis: an introduction to basic and advanced multilevel modelling. Sage, London

Sokal RR, Rohlf FJ (1995) Biometry, 3rd edn. W. H. Freeman & Co., New York

Soma KK (2006) Testosterone and aggression: Berthold, birds and beyond. J Neuroendocrinol 18:543–551

Stapley J, Keogh JS (2004) Exploratory and antipredator behaviours differ between territorial and nonter-ritorial male lizards. Anim Behav 68:841–846

Stirling DG, Reale D, Roff DA (2002) Selection, structure and the heritability of behaviour. J Evol Biol 15:277–289

van Oers K, de Jong G, Drent PJ, van Noordwijk AJ (2004a) A genetic analysis of avian personality traits: correlated, response to artificial selection. Behav Genet 34:611–619

van Oers K, Drent PJ, de Goede P, van Noordwijk AJ (2004b) Realized heritability and repeatability of risk-taking behaviour in relation to avian personalities. Proc R Soc Lond Ser B Biol Sci 271:65–73

van Oers K, de Jong G, van Noordwijk AJ, Kempenaers B, Drent PJ (2005a) Contribution of genetics to the study of animal personalities: a review of case studies. Behaviour 142:1185–1206

van Oers K, Klunder M, Drent PJ (2005b) Context dependence of personalities: risk-taking behavior in a social and a nonsocial situation. Behav Ecol 16:716–723

van Oers K, Buchanan KL, Thomas TE, Drent PJ (2011) Correlated response to selection of testosterone levels and immunocompetence in lines selected for avian personality. Anim Behav 81:1055–1061

van Overveld T, Matthysen E (2010) Personality predicts spatial responses to food manipulations in free-ranging great tits (*Parus major*). Biol Lett 6:187–190

Vazire S, Gosling SD, Dickey AS, Schapiro SJ (2007) Measuring personality in nonhuman animals. In: Robins RW, Fraley RC, Krueger RF (eds) Handbook of research methods in personality psychology. The Guilford Press, New York, pp 190–206

Verbeek MEM, Drent PJ, Wiepkema PR (1994) Consistent individual differences in early exploratory behavior of male great tits. Anim Behav 48:1113–1121

Verbeek MEM, Boon A, Drent PJ (1996) Exploration, aggressive behavior and dominance in pair-wise confrontations of juvenile male great tits. Behaviour 133:945–963

Walker DA (2003) JMASM9: converting Kendall's tau for correlational or meta-analytic analyses. J Mod Appl Stat Meth 2:525–530

Walsh B, Blows MW (2009) Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. Annu Rev Ecol Evol Syst 40:41–59

Watson SL, Ward JP (1996) Temperament and problem solving in the Small-eared bushbaby (*Otolemur garnettii*). J Comp Psychol 110:377–385

Watts HE, Blankenship LM, Dawes SE, Holekamp KE (2010) Responses of spotted hyenas to lions reflect individual differences in behavior. Ethology 116:1199–1209

Webster MM, Ward AJW, Hart PJB (2007) Boldness is influenced by social context in threespine stick-lebacks (*Gasterosteus aculeatus*). Behaviour 144:351–371

Weiss A, Adams MJ, Widdig A, Gerald MS (2011) Rhesus macaques (*Macaca mulatta*) as living fossils of hominoid personality and subjective well-being. J Comp Psychol 125:72–83

Whitlock MC, Phillips PC, Fowler K (2002) Persistence of changes in the genetic covariance matrix after a bottleneck. Evolution 56:1968–1975

Wielebnowski NC (1999) Behavioral differences as predictors of breeding status in captive cheetahs. Zoo Biol 18:335–349

Wilson DS (1998) Adaptive individual differences within single populations. Philos Trans R Soc Lond B Biol Sci 353:199–205

Wilson ADM, Godin J-GJ (2009) Boldness and behavioral syndromes in the bluegill sunfish, *Lepomis macrochirus*. Behav Ecol 20:231–237

Wilson ADM, Godin J-GJ, Ward AJW (2010) Boldness and reproductive fitness correlates in the Eastern mosquitofish, *Gambusia holbrooki*. Ethology 116:96–104

Wilson ADM, Binder TR, McGrath KP, Cooke SJ, Godin J-GJ (2011a) Capture technique and fish personality: angling targets timid bluegill sunfish, *Lepomis macrochirus*. Can J Fish Aquat Sci 68:749–757

Wilson AJ, de Boer M, Arnott G, Grimmer A (2011b) Integrating animal personality research and animal context theory: aggressiveness in the green swordtail *Xiphophorus helleri*. PLoS One 6:e28024

Wolak ME, Fairbairn DJ, Paulsen YR (2012) Guidelines for estimating repeatability. Methods Ecol Evol 3:129–137

Wolf M, Weissing FJ (2010) An explanatory framework for adaptive personality differences. Philos Trans R Soc Lond B Biol Sci 365:3959–3968