

Published in final edited form as:

Anim Behav. 2009 April 1; 77(4): 771–783.

The repeatability of behaviour: a meta-analysis

Alison M. Bell*, Shala J. Hankison, and Kate L. Laskowski

School of Integrative Biology, University of Illinois, Urbana-Champaign

Abstract

There is increasing interest in individual differences in animal behaviour. Recent research now suggests that an individual's behaviour, once considered to be plastic, may be more predictable than previously thought. Here, we take advantage of the large number of studies that have estimated the repeatability of various behaviours to evaluate whether there is good evidence for consistent individual differences in behaviour and to answer some outstanding questions about possible factors that can influence repeatability. Specifically, we use meta-analysis to ask whether different types of behaviours were more repeatable than others, and if repeatability estimates depended on taxa, sex, age, field versus laboratory, the number of measures and the interval between measures. Some of the overall patterns that were revealed by this analysis were that repeatability estimates were higher in the field compared to the laboratory and repeatability was higher when the interval between observations was short. Mate preference behaviour was one of the best studied but least repeatable behaviours. Our findings prompt new insights into the relative flexibility of different types of behaviour and offer suggestions for the design and analysis of future research.

Keywords

behavioural syndrome; coping style; courtship; individual difference; mate preference; personality; temperament

Within the field of animal behaviour, there is growing interest in consistent individual differences in behaviour (Dall et al. 2004; Sih et al. 2004a, b; Dingemanse & Reale 2005; Bell 2007; Reale et al. 2007). Accumulating evidence from a wide variety of species suggests that some individuals are consistently more aggressive, more exploratory, or more bold than other individuals and that these consistent individual differences in behaviour are often heritable (Boake 1994; Stirling et al. 2002; Kolliker 2005; van Oers et al. 2005) and related to fitness (Dingemanse & Reale 2005; Smith & Blumstein 2008). However, to date, the published data have not been summarized in a way that allows us to assess the evidence for consistent individual differences in behaviour and to explain why the magnitude of individual differences is greater in some studies compared to others.

Many studies over the past several decades have already quantified consistent individual differences in behaviour by measuring the behaviour of individuals on more than one occasion. A variety of statistics have been used to estimate behavioural consistency such as the product moment correlation or the Spearman rank correlation, but the most widely used

© 2009 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved

* Correspondence: A. M. Bell, School of Integrative Biology, 505 S. Goodwin Ave., Urbana, IL 61801, U.S.A.
alisonmb@life.uiuc.edu (A.M. Bell).

Supplementary Material

Supplementary Material for this article may be found, in the online version, at doi:10.1016/j.anbehav.2008.12.022.

statistic is the intraclass correlation coefficient, which estimates repeatability (Hayes & Jenkins 1997). Repeatability is the fraction of behavioural variation that is due to differences between individuals. Formally, repeatability is $r = \frac{s_A^2}{s_A^2 + s^2}$ where s_A^2 is the variance among individuals and s^2 is the variance within individuals over time. Behaviours that show relatively low within-individual variance compared to high among-individual variance are more repeatable. In other words, when individuals behave consistently through time and when individuals behave differently from each other, then the behaviour is repeatable. In the past, most studies measured repeatability as a first step towards studying the genetic basis for a behaviour in order to set an upper bound to heritability (Boake 1989; Dohm 2002). A different rationale for estimating repeatability is to assess interobserver reliability and the internal consistency of an instrument, (Hoffmann 2000).

From a different perspective, the large collection of repeatability estimates provides an opportunity to evaluate whether there is good evidence for consistent individual differences in behaviour and to determine whether there are systematic factors that can explain variation in behavioural consistency. Therefore summarizing this literature (previously reviewed in part in: Boake 1989; Hayes & Jenkins 1997; Forstmeier & Birkhead 2004) will provide a strong foundation for moving the study of animal personality forward.

Here, we perform a meta-analysis of the large number of published estimates of repeatability that are based on observations of a single behaviour measured on the same individuals on more than one occasion. Although closely allied with concepts of behavioural syndromes (Sih et al. 2004a), temperament (Reale et al. 2007), personality (Gosling 2001) and coping styles (Koolhaas et al. 1999), all of which generally refer to behavioural consistency through time and across situations, repeatability is more restrictive than these concepts because it ideally refers to consistency of a particular behaviour through time, not necessarily behavioural consistency across situations or contexts. However, in many cases, the specific environmental situations in which a behaviour is being measured are not known. As a result, repeatability estimates reflect both consistency through time and consistency across unmeasured situations (Martin & Reale 2008). Obviously, using a similar framework to assess the evidence for behavioural correlations across contexts is a promising subject for future meta-analyses.

In addition to assessing the claim that individual differences are common, we wish to know whether there are generalizations that can be made about the factors influencing repeatability. We perform an exploratory analysis to address the following questions.

Are Certain Types of Behaviour More Repeatable Than Others?

Studies have estimated the repeatability of behaviours ranging from mate preference to exploratory behaviour to parental behaviour. Therefore, we have an opportunity to ask whether certain types of behaviour are more repeatable than others. One prediction is that behaviours that are more sensitive to the environment (more plastic) are less repeatable. For example, we might assume that behaviours under morphological or physiological constraint should be relatively stable compared to behaviours influenced by energetic needs or the immediate social environment (Castellano et al. 2002; Smith & Hunter 2005). However, if all individuals respond to the environment in a similar way, the behaviour can still be repeatable despite this plasticity. Instead, repeatability estimates are especially affected by individual*environment interactions, or when individuals respond differently to the environment (Nussey et al. 2007; Martin & Reale 2008). Therefore comparing the repeatability of different types of behaviour has the potential to reveal new insights about the flexibility or canalization of different types of behaviour.

Are Certain Taxa More Repeatable Than Others?

Reviews of heritability estimates have found strong taxonomic differences (Mousseau & Roff 1987). Among vertebrates, for example, the heritability of morphological traits is significantly lower for ectotherms than it is for endotherms (Mousseau & Roff 1987), perhaps because ectotherms are more influenced by their environment. Here, we follow Mousseau & Roff's lead and test whether the same pattern applies to repeatability. We compared patterns of repeatability variation within four major phylogenetic groupings: invertebrates versus vertebrates and endothermic vertebrates versus ectothermic vertebrates. The invertebrate–vertebrate comparison allows us to evaluate the suggestion that the behaviour of taxa with less flexible nervous systems is less plastic, leading to higher repeatability estimates for invertebrates.

Does Repeatability Decrease with the Interval Between Observations?

From a genetic perspective, repeatability might decrease with the interval between measurements because the 'same' phenotypic trait may be influenced by different sets of genes at different ages. Therefore increasing the interval between measurements should decrease repeatability of the phenotypic traits because the two measures do not represent exactly the same trait at the genetic level.

Environmental effects might also cause repeatability to decrease with the interval between observations. For example, when the interval between observations of behaviour is short, it is likely that the animals are of similar state (hunger, size, age, condition, dominance, etc.) during both observations and are experiencing similar environments. For example, we might expect reproductive effort in birds to be more repeatable within broods rather than across seasons (Potti et al. 1999; Moreno et al. 2002). In contrast, when the interval between observations is long, there is more opportunity for developmental change; individuals are more likely to undergo dramatic change such as sexual maturity or a niche shift over a longer period of time. Indeed, consistency decreases with the interval between observations in humans (Roberts & DeVecchio 2000) and great tits, *Parus major* (Dingemanse et al. 2002). Published estimates of repeatability have used a wide variety of intervals between measurements; therefore, they provide an opportunity to test this intuitive suggestion.

Does Repeatability Increase with the Number of Observations Per Individual?

Several studies have measured the same individuals more than two times to calculate repeatability. On one hand, increasing the number of measurements per individual can decrease the measurement error associated with each observation, and therefore might increase repeatability (Hoffmann 2000). On the other hand, when individuals are measured repeatedly, they might habituate to the behavioural assay and become less responsive, or alternatively, become sensitized (Martin & Reale 2008). It is of practical importance to evaluate the relationship between the number of measures and repeatability for the design and analysis of future experiments (i.e. if there is much to be gained by measuring individuals more than twice; Adolph & Hardin 2007).

Does Repeatability Vary Among Age Groups?

In humans, behavioural consistency increases with maturity (Roberts & DeVecchio 2000); older people behave more consistently than younger ones, perhaps because the cumulative experience of the environment leads to increasing consistency with age. Other mechanisms that could cause repeatability to increase with age are the process of consolidated identity or reputation (Roberts & DeVecchio 2000; Dall et al. 2004). Because some studies have estimated the repeatability of behaviours in juveniles while others have measured adults, here, we have an opportunity to ask whether the same trend applies to nonhuman animals.

Do Repeatability Estimates Differ Between the Field and the Laboratory?

Presumably, environmental variance is greater in the field compared to the stable conditions in the laboratory. To the extent that a changing environment is associated with behavioural plasticity, we might expect repeatability to be lower in the field, as has been found for estimates of heritability in *Drosophila* (Hoffmann 2000).

Do Males and Females Differ in Repeatability?

Two lines of thought in the literature suggest that males might be more repeatable than females. First, the older literature on the persistence of aggression (e.g. Andrew 1972; Wingfield 1994) suggests that testosterone can cause males to be more predictable than females. Second, honest indicator models of sexual selection predict that the behaviours indicated by a sexually selected trait are predictable because females use the trait as a reliable cue for how her mate will behave, for example, as a father (Kokko 1998; Garamszegi et al. 2006a). Here, we assess whether there are sex differences in repeatability generally.

Measuring the behaviour of individually marked animals on several occasions is laborious and therefore it is not surprising that few studies have systematically attempted to compare repeatability across ages (Bakker 1986; Masters et al. 1995; Battley 2006; Missoweit et al. 2007), sexes (Dingemanse et al. 2002; Schwagmeyer & Mock 2003; Nakagawa et al. 2007), locations (Howard & Young 1998; Kolluru 1999), intervals (Allen 1998) or treatments (Spencer & Thompson 2003; Magellan & Magurran 2007). Indeed, simultaneously addressing all of the questions listed above is beyond the scope of any project. However, meta-analysis allows us to address these questions using the growing body of literature concerning repeatability. By drawing on the available data, we can test our hypotheses regarding repeatability as it relates to behaviour, taxa, developmental stage, sex, and so on, to gain insight into how and why repeatability varies. This broad, exploratory analysis is also useful for stimulating new hypotheses and identifying particularly unexplored research directions.

METHODS

We compiled the data set by searching for published estimates of repeatability using the Web of Science search engine with combinations of the following topic terms in July 2008: repeatability, behav*, repeatab*, intraclass correlation coefficient, mate choice, preference, migration, predator. We also searched the reference list of each paper to identify studies that were missed in the initial search.

We used the following criteria when compiling the data set. (1) Studies had to measure the repeatability of individual behaviour, as opposed to the repeatability of the behaviour of a group, pair, chorus, colony, etc. (2) Studies on domesticated animals or animals in a zoo setting were excluded. (3) To facilitate comparisons across studies, the study needed to estimate repeatability as the intraclass correlation coefficient. This criterion excluded studies that measured binary behaviours (e.g. Preziosi & Fairbairn 1996), or that estimated repeatability using Pearson or Spearman correlation coefficients. (4) Studies that measured the repeatability of physiological (e.g. metabolic rate, hormone titre), performance-related (e.g. sprint speed, etc.), morphological (e.g. sperm characters such as the number of sperm per ejaculate or feather length), or life history (e.g. timing of breeding, laying date, clutch size, hatch date, arrival date, growth rate) traits were excluded. The final list of studies is given in Table 1 and the entire data set is given in Supplementary Table S1.

We used the following grouping variables to characterize each repeatability estimate: developmental stage (adult, juvenile or both), sex (male, female or both), whether the study was conducted in the field or laboratory, if the interval between observations was greater or less than 1 year, the minimum number of times that individuals were measured in the study, the functional class of behaviour and taxonomic group. Following Mousseau & Roff (1987), we looked for differences between vertebrates versus invertebrates and ectothermic versus endothermic vertebrates. We categorized behaviours into the following functional classes: courtship, mate preference, activity, affiliation, aggression, antipredator (including risk taking), exploratory behaviour, foraging, habitat selection and territoriality (including nest site selection and thermal preference), migration, mating (any behaviours performed during mating and including extrapair copulations), parental care and other.

In an important paper, Lessells & Boag (1987) pointed out that MS_a (the mean square among individuals) depends on n_0 , the coefficient representing the number of observations per individual. When the number of observations per individuals is unequal, n is greater than n_0 . Estimates that do not correct for different numbers of observations per individuals systematically underestimate repeatability; the difference between n and n_0 increases with increasing spread in the number of measures per individual. Therefore, we compared repeatability estimates that either did or did not correct for different numbers of measures per individual, as suggested by Lessells & Boag (1987).

An advantage of meta-analytic techniques is that it scales the weight given to the results of each study based on its power and precision. This is done through the conversion on the original test statistic (here, repeatability) to an effect size. The effect size of each repeatability estimate was calculated in MetaWin 2.1 (Rosenberg et al. 2000). The average effect size was computed as a weighted mean, whereby the weights were equal to the inverse variance of each study's effect estimator. Larger studies and studies with less random variation were given greater weight than smaller studies. Analysis of effect sizes rather than raw repeatability estimates is preferable because more weight should be given to more powerful studies. Therefore, all subsequent analyses were performed on estimates of effect size, rather than the raw repeatability score.

To understand the causes of variation in repeatability estimates, we used fixed effects categorical or continuous models in MetaWin. For comparisons between groups of studies, we report Q_b , the between-groups homogeneity. This statistic is analogous to the between-groups component of variance in conventional analysis of variance, and it is χ^2 distributed with n groups minus one degree of freedom. We also report effect sizes and their 95% confidence intervals as CL1 effect size CL2.

Limitations of the data set and statistical options available for meta-analysis precluded us from formally testing statistical interactions between the grouping variables. We explored patterns in the data set by analysing subsets of the data according to different levels of the factor of interest. For example, after testing for a difference in effect size between males and females using all the data, we then performed the same analysis when field studies were excluded. We repeated the analysis when laboratory studies were excluded, and so forth. We infer that patterns that were common to several subsets of the total data set are robust and do not depend on other grouping variables (see Table 2). If the effect of a grouping variable was significant for one level of a different grouping variable but not for the other level, then we infer that there might be an interaction between the two grouping variables. We also pay particular attention to effect sizes because when a subset of data was eliminated from the analysis, our power to detect a significant effect was reduced. Therefore, in addition to asking whether comparisons are statistically significant for certain subsets of the data, we also report whether effect sizes changed. We view this exploratory analysis as a mechanism

for generating hypotheses and to suggest promising areas for future study. We ranked the P values in each column in Table 2 and used the sequential Bonferroni procedure to account for multiple comparisons (Rice 1989).

Many papers reported more than one repeatability estimate, introducing the possibility of pseudoreplication if multiple estimates from the same study are nonindependent of each other. For example, studies of calling behaviour in frogs often measure more than one attribute of a male's call on multiple occasions, including amplitude, duration, frequency, and so forth. If the attributes are correlated with each other (e.g. fundamental frequency is positively correlated with dominant frequency; Bee & Gerhardt 2001), then repeatability estimates for the different attributes are not independent. There is no clear consensus about how to handle multiple estimates reported from the same study in meta-analysis (Rosenberg et al. 2000). On one hand, we want to avoid nonindependence among effect sizes, but on the other hand, we do not want to lose biologically meaningful information by using only one estimate per study (e.g. the study's mean). The loss of information caused by omission of such effects may lead to more serious distortions of the results than those caused by their nonindependence (Gurevitch et al. 1992).

Therefore, we took multiple strategies to address possible bias caused by the nonindependence of multiple estimates per study. First, in cases where studies reported separate repeatability estimates on behaviours measured on more than two occasions, we did not include estimates that provided potentially redundant information (Bakker 1986; Hager & Teale 1994; Archard et al. 2006). For example, a study that measured individuals on three occasions could potentially report repeatability for the comparison between measures one and two, measures two and three, and measures one and three. In this case, we excluded the estimate of repeatability between measures two and three, as it would not provide additional information (for the purposes of our analysis) compared to the repeatability reported in times one and two. We did include the repeatability estimate between times one and three, however, as this represents a different interval between measures, one of the factors in which we were interested. Similarly, when studies reported repeatability for both separate and pooled groups (e.g. males, females, and males and females), we did not include the pooled estimate (Gil & Slater 2000; Archard et al. 2006; Battley 2006).

Second, we compared studies that reported different numbers of repeatability estimates (as in Nespolo & Franco 2007). We found no relationship between the number of estimates reported and the value of those estimates (slope = -0.002 , $Q_{\text{regression}} = 1.19$, $P = 0.28$). This suggests that the number of estimates reported by a study does not systematically change the effect size reported. Third, we removed, one at a time, studies that contributed the greatest number of estimates to the data set to evaluate whether they were primarily responsible for the observed patterns. Removing studies that reported the highest numbers of estimates did not change any of the main effects (results not shown).

Finally, because a large proportion of estimates were based on just two behaviours (courtship and mate preference, see Results), we reanalysed the data set when either courtship behaviours or mate preference behaviours were excluded. We paid particular attention to whether effect sizes changed in the reduced data sets to determine whether these widely studied behaviours disproportionately influenced the results.

Two studies (Hoffmann 1999; Serrano et al. 2005) in our data set measured a much larger number of individuals ($N = 972$ and $N = 1318$, respectively) to estimate repeatability and were therefore weighted much more heavily in the meta-analysis. For comparison, the average sample size of the remaining data set was 39. Serrano et al. (2005) measured habitat preference across years in adult kestrels in the field and found relatively high repeatability

for this behaviour. Hoffmann (1999) measured two courtship behaviours of male *Drosophila* in the laboratory and estimated relatively low repeatabilities.

On one hand, the purpose of meta-analysis is to take differences in power into consideration when evaluating across studies; therefore, it follows that these two studies should be weighted more heavily in our analysis. On the other hand, these two studies are not representative of most studies on repeatability (the next highest sample size after Serrano et al. 2005 in the data set is $N = 496$) and therefore they might bias our interpretation. For example, the repeatability estimate in the Serrano et al. (2005) was relatively high ($R = 0.58$) and was measured in the field. Therefore, this heavily weighted result might cause it to appear that repeatability is higher in the field than in the laboratory. To address the possibility that these particularly powerful studies were driving our results, we ran our analyses when the three estimates from these two studies were excluded.

To determine whether our data set was biased towards studies that found significant repeatability estimates (the ‘file drawer effect’), we constructed funnel plots (Light & Pillemer 1984) and calculated Rosenthal’s (1979) ‘fail-safe numbers’ in MetaWin. Funnel plots are useful for visualizing the distribution of effect sizes of sample sizes in the study. Funnel plots with wide openings at smaller sample sizes and with few gaps generally indicate less publication bias (Rosenberg et al. 2000). Fail-safe numbers represent the number of nonsignificant, missing or unpublished studies that would need to be added to the analysis to change the results from significant to nonsignificant (Rosenberg et al. 2000). If these numbers are high, relative to the number of observed studies, the results are probably representative of the true effects, even in the face of some publication bias (Rosenberg et al. 2000).

RESULTS

Summarizing the Data Set

We identified 759 estimates of repeatability that met our criteria (Fig. 1). The estimates are from 114 studies, representing 98 species (Table 1). The sample size (number of individuals measured) ranged from 5 to 1318. Most studies measured the subjects twice, although some studies measured individuals as many as 60 times, with a mean of 4.41 measures per individual. The majority of repeatability estimates (708 of 759) considered in this meta-analysis were calculated as suggested by Lessells & Boag (1987). As predicted, estimates that did not correct for different numbers of observations per individual (mean effect size = 0.47, 95% confidence limits = 0.43, 0.52; hereafter reported as 0.43–0.47–0.52) were higher than those that did correct for different numbers of observations per individual (0.35–0.37–0.38, $Q_b = 23.0$, $N = 759$, $P < 0.001$) (Lessells & Boag 1987). However, we found no evidence that this confounded our overall results.

Studies measured the repeatability of a wide variety of behaviours; courtship (327 estimates from 40 studies) and mate preference (148 estimates from 34 studies) were particularly well studied (Table 1, Fig. 2a). Most estimates came from studies of vertebrates (493 versus 266 estimates for invertebrates), with 201 estimates from birds alone (Fig. 2b). The majority of behaviours were studied in adults (706 versus 50 estimates on juveniles, 3 estimates on both adults and juveniles), and more estimates came from studies of males than females (388 versus 275; 95 estimates for both). Most studies measured individuals repeatedly within 1 year, although 69 estimates were based on an interval between observations that was greater than 1 year. Fewer estimates were made in the field (293 estimates) compared to the laboratory (466 estimates).

Altogether the data overwhelmingly support the hypothesis that behaviour is repeatable (Fig. 1). The average repeatability across all estimates was 0.37, and the weighted effect size across all estimates was significantly greater than zero (0.36 0.37 0.38, $Q_{\text{total}} = 3860.9$, $N = 759$, $P < 0.001$).

Evaluating Hypotheses

Are certain types of behaviour more repeatable than others?—Repeatability estimates varied widely across different classes of behaviour. The most repeatable classes of behaviour were mating, habitat selection and aggression, while the least repeatable behaviours were activity, mate preference and migration (Fig. 2a). The two best-studied behaviours, mate preference and courtship, had very different repeatabilities; courtship was more repeatable than mate preference.

Are certain taxa more repeatable than others?—There was not a clear difference in the repeatability of the behaviour of invertebrates compared to vertebrates ($Q_b = 2.79$, $N = 759$, $P = 0.0951$; Figs 2b, 3a), but further analyses suggested that the difference between invertebrates versus vertebrates might depend on the behaviour under consideration. On behaviours other than courtship, for example, invertebrates were more repeatable than vertebrates (0.41 0.45 0.48 versus 0.32 0.33 0.33; $Q_b = 33.6$, $N = 432$, $P < 0.001$; Table 2). For behaviours other than mate preference, on the other hand, vertebrates were more repeatable than invertebrates (0.42 0.43 0.45 versus 0.37 0.39 0.41; $Q_b = 13.7$, $N = 633$, $P < 0.001$; Table 2). It is likely that the interaction between taxonomic grouping and behaviour was influenced by the fact that mate preference behaviours, which generally had low repeatability, were typically measured on vertebrates.

As with heritability (Mousseau & Roff 1987), we found suggestive evidence that endothermic vertebrates were more repeatable than ectothermic vertebrates ($Q_b = 14.7$, $N = 493$, $P = 0.001$; Fig. 3b). This pattern depended on whether the animals were measured in the field or the laboratory: in the field, there was no difference (Table 2), but in the laboratory, endotherms were more repeatable (0.32 0.36 0.40 versus 0.22 0.24 0.27; $Q_b = 15.1$, $N = 186$, $P = 0.001$; Table 2). The large estimate reported in Serrano et al. (2005), which was measured in an endotherm in the field, may have been driving the overall difference between endotherms and ectotherms, but it seems unlikely that it was solely responsible for the difference because endotherms were more repeatable than ectotherms in laboratory studies only (Table 2).

Does repeatability decrease with the interval between observations?—Initially, it appeared that there was no difference in repeatability based on short versus long intervals between observations ($Q_b = 0.87$, $N = 759$, $P < 0.350$; Fig. 3c). However, closer analysis showed that this surprising result was probably caused by two particularly powerful and therefore heavily weighted studies in the meta-analysis: Hoffmann (1999) lowered effect sizes for short intervals, and Serrano et al. (2005) raised effect sizes for long intervals. When these studies were removed, repeatability estimates were higher for behaviours measured close together in time ($Q_b = 43.1$, $N = 755$, $P < 0.001$; Fig. 3c). This significant effect was robust to several other subsets of the data (Table 2).

Does repeatability increase with the number of observations per individual?—We found no evidence that repeatability estimates were affected by the number of observations per individual (slope = 0.008; $Q_{\text{regression}} = 0.42$, $N = 759$, $P = 0.516$; Fig. 4).

Does repeatability vary among age groups?—For this comparison, we did not consider adult-specific behaviours such as mate preference, mating, courtship and parental

behaviour. Overall, there was no difference in the repeatability of behaviour in juveniles or adults ($Q_b = 0.6166$, $N = 220$, $P = 0.4323$; Fig. 3d). However, certain subsets of the data set suggest that there might be important differences in the repeatability of behaviour of juveniles and adults. Among the subsets of the data set for which there was a statistically significant difference, the behaviour of juveniles was consistently more repeatable than the behaviour of adults. For example, among ectotherms, juvenile behaviour was more repeatable than adult behaviour ($Q_b = 13.19$, $N = 72$, $P = 0.0003$; Table 2).

Do repeatability estimates differ between the field and the laboratory?—

Overall, we found that behaviours measured in the field were more repeatable than behaviours measured in the laboratory (Fig. 3e). This pattern was robust across all subsets of the data set.

Do males and females differ in repeatability?—Overall, males were more repeatable in their behaviour than females (Table 2, Fig. 3f). The sex difference was observed in adults, but not in juveniles, and was true for all vertebrates (Table 2). However, there was an interaction between sex and the type of behaviour measured. When mate preference was omitted from the data set, the pattern was reversed and females were more repeatable than males, as judged both by the P value and by effect sizes (0.38–0.40–0.41 versus 0.43–0.47–0.51; $Q_b = 12.3$, $N = 538$, $P < 0.001$; Table 2, Fig. 3f). Therefore, it is likely that the very low repeatability of mate preference behaviours, which were typically measured on females (9 estimates of the repeatability of mate preference were for males versus 139 estimates for females), shifted the female average downwards.

Testing for Publication Bias

We found no evidence for publication bias based on either a visual inspection of our funnel plot (Fig. 5) or based on Rosenthal's fail-safe numbers. Our fail-safe numbers were very large relative to our observed sample sizes, with Rosenthal's numbers ranging from 100 to over 900 times the number of results in our analysis. Even when only the means from studies reporting multiple repeatability estimates were used, our Rosenthal's number was over 190 times as large as the number of included studies.

DISCUSSION

Our analysis provides strong support for consistent individual differences in behaviour. We found that the repeatability of behaviour was significantly greater than zero, and that roughly 35% of the variation among individuals in behaviour could be attributed to individual differences. Despite the heterogeneous nature of the data set, our analysis also uncovered some intriguing patterns.

We found strong evidence that not all types of behaviours were equally repeatable. Overall it is difficult to make inferences about the causes of variation in repeatability of behaviours (i.e. if some behaviours are more repeatable than others because they are the ones that are least influenced by the environment or the most canalized). However, one pattern that was robust among almost all subsets of the data was that individuals (typically females) were not consistent in their mate preferences; just because a female preferred a certain type of male on one occasion did not necessarily mean that she retained that preference on subsequent occasions. This result is consistent with a growing number of studies showing that what a female prefers in a mate is subject to change depending on her age, condition and the environment (reviewed in: Jennions & Petrie 1997; Cotton et al. 2006). In addition, by their very nature, repeatability studies allow the test subject to have more information about the distribution of quality of mates in the local environment in the second testing situation

compared to the first (Janetos 1980; Parker 1983; Real 1990; Dombrovsky & Perrin 1994). After receiving additional information in the first test, females might 'fine-tune' their preference in subsequent tests, therefore lowering repeatability.

Our analysis also suggests that not all types of taxa are equally repeatable. For example, endotherms were generally more repeatable than ectotherms, as has also been found for heritability estimates (Mousseau & Roff 1987). One interpretation of this pattern is that ectotherms are more sensitive to the environment, and, therefore, individuals are more likely to change their behaviour according to the environment (but not equally). It is intriguing that when we compared endotherms to ectotherms only in field studies, in which we presume there was more environmental variation compared to in laboratory studies, the difference between endotherms and ectotherms disappeared, contrary to our first interpretation. One possible explanation for this is that mate preference studies (which had very low repeatabilities) were typically conducted in the laboratory, and for mate preference behaviours, endotherms were more repeatable than ectotherms (0.24 0.28 0.33 versus 0.15 0.18 0.21, $N = 112$). In contrast, we found little evidence in support of the popular notion that invertebrates are more rigid in their behaviour than vertebrates.

We found strong support for the intuitive hypothesis that individuals are more consistent over short intervals compared to long intervals, at least when Hoffmann (1999) and Serrano et al. (2005) were excluded. Repeatability was significantly higher when the same individuals were measured for a second time within a year of the first measurement. Granted, greater than or less than 1 year is a fairly coarse measure, and one which does not take differences in life span into consideration. That is, a day in the life of a cricket that lives for only a few weeks (Kolluru 1999) represents a considerably longer fraction of its total life span compared to a long-lived organism such as an elephant seal (Sanvito & Galimberti 2003). This rough measure could therefore lead to bias if taxonomic differences were confounded with interval (i.e. short-lived organisms such as invertebrates are relatively repeatable and were also measured over relatively short intervals). However, we found no difference in the repeatability of behaviour of invertebrates versus vertebrate animals, and, therefore, do not consider taxonomic group to be a confounding variable. In addition, when we looked for relationships between repeatability and the interval between measurements while controlling for life span (and age at maturity), the effect of interval did not change (results not shown). As more data become available, it will be useful to carry out this type of broad comparison in the correct phylogenetic framework.

We found suggestive evidence that there might be systematic differences in the repeatability of behaviour of juveniles versus adults. At first glance, it appeared that there was no difference in the repeatability of behaviour of adults or juveniles. Unfortunately, there are only a few examples in the data set of repeatability estimates of juveniles and adults of the same species and they do not suggest a strong pattern (sticklebacks, *Gasterosteus aculeatus*: 0.68 juveniles versus 0.78 adults; Bakker 1986; big brown bat, *Eptesicus fuscus*: 0.51 juveniles versus 0.60 adults; Masters et al. 1995; godwit, *Limosa limosa baueri*: 0.4 juveniles versus 1.19 adults; Battley 2006; scorpionfly, *Panorpa vulgaris*: 0.30 juveniles versus 0.21 adults; Missoweit et al. 2007). Comparing the repeatability of behaviour of juveniles versus adults within the same species is an important, interesting and relatively unexplored question with no clear predictions about the direction of the effects. On one hand, we might expect juveniles to be undergoing dramatic developmental change and therefore not show repeatable behaviour. On the other hand, we might expect juveniles to be more repeatable because the costs of straying from a developmental trajectory are higher for juveniles (Biro & Stamps 2008).

Changes in repeatability with age might also reflect the action of selection on phenotypic variance. If there is directional or stabilizing selection on a particular behaviour, then phenotypic variance will decrease after selection. This could cause repeatability to decrease with age (if there is less variation among adults compared to juveniles). Alternatively, if traits expressed early in life are subject to stronger selection pressures than traits expressed later in life, then overall repeatability might increase with age (because there is more variation among adults compared to juveniles).

Contrary to our prediction, we found that behaviour was generally more repeatable in the field than the laboratory. Initially, we reasoned that greater environmental variance in the field would increase within-individual variation (s^2) and thereby decrease repeatability. Alternatively, greater environmental variance in the field might allow the expression of more behavioural variation among individuals (s^2_A), by creating micro-niches, and thereby increase repeatability. Provided there are advantages of behaving consistently (Dall et al. 2004; McElreath & Strimling 2006), this could explain why consistent individual differences in behaviour were greater in the field.

Our results suggest that there might be important sex differences in repeatability, but the direction of the difference probably depends on the behaviour under consideration. Overall, we found that males were more repeatable than females, which was consistent with our predictions (e.g. Andrew 1972; Wingfield 1994). For example, male house sparrows, *Passer domesticus*, were more repeatable with respect to parental behaviour compared to female house sparrows (Schwagmeyer & Mock 2003; Nakagawa et al. 2007). The lower repeatability of female parental care might be due to the greater responsiveness of females to the needs of the brood. Alternatively, males might be more repeatable than females because of selection favouring honest signalling (Nakagawa et al. 2007).

Closer examination revealed that the overall sex difference was driven by the extremely low repeatability of mate preference, which was typically measured on females (see discussion of mate preference above). Therefore, our evidence for sex differences in repeatability is inconclusive. This is in agreement with the data on the consistency of personality in humans as well, where sex differences in consistency are rarely observed (Robins et al. 2001). This lack of a consistent pattern across such a wide variety of classes of behaviours may, in fact, be expected if there are differences between the sexes and if the best overall behavioural strategy varies by gender.

Although the large collection of published estimates of repeatability offers an opportunity to look for patterns to explain variation in consistent individual differences in behaviour, there are limitations of the data set and several questions remain unanswered because of the heterogeneous and possibly biased nature of the collection of estimates. For example, one possible source of bias in the data set comes from the types of behaviours that were originally studied. Many studies estimated repeatability as a first step towards studying the genetic basis of the behaviour. Insofar as researchers only commenced studies of the genetic basis of behaviour on behaviours that they already suspected were heritable, then published estimates of repeatability might be biased upwards.

Another source of bias is measurement error. Most studies included in the data set did not distinguish measurement repeatability from true 'trait' repeatability (Falconer & Mackay 1996; Hoffmann 2000). Here, we are interested in using the published data to infer patterns about the underlying causes of deviation from consistency. However, one reason why it could appear that individuals do not behave consistently different from each other is because there is measurement error associated with each behavioural observation. This introduces a

possible source of bias in the data set if certain studies or certain types of behaviour have more measurement error associated with them than others.

An additional factor that can potentially reduce repeatability is mean-level change between measurements (e.g. on average, the population is more aggressive the second time it is observed compared to the first time; Hayes & Jenkins 1997). If mean-level change causes more within-individual than between-individual change between observations, then repeatability will be low. Mean-level change might have contributed to our finding that repeatability declines as a function of test–retest intervals because mean-level changes in behaviour are more likely to occur over longer periods than over shorter periods. In general, however, mean-level change does not preclude the possibility that repeatability will be significantly different from zero. So long as between-individual differences are large relative to within-individual differences, a behaviour can still be repeatable despite mean-level change.

Our results offer several suggestions for the design and analysis of future research. First, repeatability does not appear to depend on the number of times that individuals are measured. Indeed, it seems likely that increasing the number of observations per individual will decrease the error around the estimate, rather than the repeatability estimate itself. This result suggests that if researchers want to estimate repeatability of a behaviour, they have more to gain by measuring more individuals on fewer occasions rather than fewer individuals on more occasions. Second, to facilitate comparisons across studies, it would be helpful if future papers report statistics such as n_o , whether there was mean-level change between observations, and whether variance among individuals was the same at each measurement (an assumption of the intraclass correlation coefficient statistic but rarely reported).

On a related note, repeatability statistics say little about whether individuals themselves are repeatable; the statistic is a property of the population of individuals. It is likely that in most studies, individuals differ in how much their behaviour changes between observations. That is, even when a repeatability statistic is significantly different from zero, it does not necessarily mean that all of the individuals within the population behaved equally consistently; some individuals were probably more consistent than others. Indeed, the literature on coping styles has emphasized that consistency is a trait that varies among individuals; the behaviour of proactive individuals, which tend to be rigid and routinized, is more repeatable than the behaviour of reactive individuals, which tend to be more responsive to cues in the environment (e.g. Benus et al. 1990, 1991; Marchetti & Drent 2000). An important direction for future studies in this area is to define conditions responsible for individual differences in plasticity (Nussey et al. 2007).

This study reveals some interesting, and sometimes surprising, results when repeatability was assessed across a wide variety of behaviours, species and experiments. For example, while higher repeatability over short intervals might be expected, higher repeatability in the field versus the laboratory prompts us to take a more nuanced look at the specific behaviours being measured in these different settings. Variation in repeatability among classes of behaviour could also influence how we interpret these behaviours. The patterns we found not only show what is known to date about repeatability and behaviour, but also serve to stimulate questions about why repeatability varies across behaviours, ages, sexes or taxa. Our results suggest that particularly interesting but relatively unexplored questions include comparing the repeatability of behaviour between different age classes, between males and females and under different ecological conditions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Thanks to Stephanie DeFlorio for technical help, Charles Roseman for informative conversations and to Judy Stamps, Chad Johnson, David Sinn and two anonymous referees for helpful comments on the manuscript.

References

- Adolph SC, Hardin JS. Estimating phenotypic correlations: correcting for bias due to intraindividual variability. *Functional Ecology*. 2007; 21:178–184.
- Allen GR. Diel calling activity and field survival of the bushcricket, *Sciarasaga quadrata* (Orthoptera: Tettigoniidae): a role for sound-locating parasitic flies? *Ethology*. 1998; 104:645–660.
- Andrew RJ. Recognition processes and behavior with special reference to effects of testosterone on persistence. *Advances in the Study of Behaviour*. 1972; 4:175–208.
- Archard GA, Cuthill IC, Partridge JC. Condition-dependent mate choice in the guppy: a role for short-term food restriction? *Behaviour*. 2006; 143:1317–1340.
- Aspbury AS, Basolo AL. Repeatable female preferences, mating order and mating success in the poeciliid fish, *Heterandria formosa*. *Behavioral Ecology and Sociobiology*. 2002; 51:238–244.
- Aspi J, Hoikkala A. Laboratory and natural heritabilities of male courtship song characters in *Drosophila montana* and *D. littoralis*. *Heredity*. 1993; 70:400–406.
- Bakker TCM. Aggressiveness in sticklebacks (*Gasterosteus aculeatus* L. : a behaviour-genetic study. *Behaviour*. 1986; 98:145–167.
- Banbura J. Mate choice by females of the swallow *Hirundo rustica*: is it repeatable? *Journal of Ornithology*. 1992; 133:125–132.
- Battley PF. Consistent annual schedules in a migratory shorebird. *Biology Letters*. 2006; 2:517–520. [PubMed: 17148277]
- Bee MA, Gerhardt HC. Neighbour–stranger discrimination by territorial male bullfrogs (*Rana catesbeiana*): I. Acoustic basis. *Animal Behaviour*. 2001; 62:1129–1140.
- Beeler AE, Rauter CM, Moore AJ. Pheromonally mediated mate attraction by males of the burying beetle *Nicrophorus orbicollis*: alternative calling tactics conditional on both intrinsic and extrinsic factors. *Behavioral Ecology*. 1999; 10:578–584.
- Bell AM. Future directions in behavioral syndromes research. *Proceedings of the Royal Society of London*. 2007; 274:755–761. Series B.
- Benesh DP, Valtonen ET, Seppala O. Multidimensionality and intra-individual variation in host manipulation by an acanthocephalan. *Parasitology*. 2008; 135:617–626. [PubMed: 18294426]
- Benus RF, Den Daas S, Koolhaas JM, Van Oortmerssen GA. Routine formation and flexibility in social and non-social behaviour of aggressive and nonaggressive male mice. *Behaviour*. 1990; 112:176–193.
- Benus RF, Bohus B, Koolhaas JM, van Oortmerssen GA. Heritable variation for aggression as a reflection of individual coping strategies. *Experientia*. 1991; 47:1008–1019. [PubMed: 1936199]
- Bety J, Giroux JF, Gauthier G. Individual variation in timing of migration: causes and reproductive consequences in greater snow geese (*Anser caerulescens atlanticus*). *Behavioral Ecology and Sociobiology*. 2004; 57:1–8.
- Birkhead TR, Fletcher F. Male phenotype and ejaculate quality in the zebra finch *Taeniopygia guttata*. *Proceedings of the Royal Society of London*. 1995; 262:329–334. Series B.
- Biro PA, Stamps JA. Are animal personality traits linked to life-history productivity? *Trends in Ecology & Evolution*. 2008; 23:361–368. [PubMed: 18501468]
- Blanckenhorn WU, Perner D. Heritability and repeatability of behavioral attributes affecting foraging success and fitness in water striders. *Animal Behaviour*. 1994; 48:169–176.
- Boake CRB. Repeatability: its role in evolutionary studies of mating behavior. *Evolutionary Ecology*. 1989; 3:173–182.

- Boake, CRB. *Quantitative Genetic Studies of Behavioral Evolution*. University of Chicago Press; Chicago: 1994.
- Bonte D, Deblauwe I, Maelfait JP. Environmental and genetic background of tiptoe-initiating behaviour in the dwarfspider *Erigone atra*. *Animal Behaviour*. 2003; 66:169–174.
- Bosch J, Marquez R. Female preference function related to precedence effect in an amphibian anuran (*Alytes cisternasii*): tests with non-overlapping calls. *Behavioral Ecology*. 2002; 13:149–153.
- Bosch J, Marquez R, Boyero L. Behavioural patterns, preference, and motivation of female midwife toads during phonotaxis tests. *Journal of Ethology*. 2003; 21:61–66.
- Brandt Y, Allen JR. Persistence of individually distinctive display patterns in fatigued side-blotched lizards (*Uta stansburiana*). *Behavioral Ecology and Sociobiology*. 2004; 55:257–265.
- Brandt LSE, Ludwar BC, Greenfield MD. Co-occurrence of preference functions and acceptance thresholds in female choice: mate discrimination in the lesser wax moth. *Ethology*. 2005; 111:609–625.
- Brodie ED, Russell NH. The consistency of individual differences in behaviour: temperature effects on antipredator behaviour in garter snakes. *Animal Behaviour*. 1999; 57:445–451. [PubMed: 10049485]
- Brooks R, Endler JA. Female guppies agree to differ: phenotypic and genetic variation in mate-choice behavior and the consequences for sexual selection. *Evolution*. 2001; 55:1644–1655. [PubMed: 11580024]
- Brooks R. Variation in female mate choice within guppy populations: population divergence, multiple ornaments and the maintenance of polymorphism. *Genetica*. 2002; 116:343–358. [PubMed: 12555789]
- Brown WD, Smith AT, Moskalik B, Gabriel J. Aggressive contests in house crickets: size, motivation and the information content of aggressive songs. *Animal Behaviour*. 2006; 72:225–233.
- Butlin RK, Hewitt GM. Heritability estimates for characters under sexual selection in the grasshopper, *Chorthippus brunneus*. *Animal Behaviour*. 1986; 34:1256–1261.
- Castellano S, Cuatto B, Rinella R, Rosso A, Giacoma C. The advertisement call of the European treefrog: a multilevel study of variation. *Ethology*. 2002; 108:75–89.
- Clark DC, Moore AJ. Variation and repeatability of male agonistic hiss characteristics and their relationship to social rank in *Gromphadorhina portentosa*. *Animal Behaviour*. 1995; 50:719–729.
- Cotton S, Small J, Pomiankowski A. Sexual selection and condition-dependent mate preferences. *Current Biology*. 2006; 16:R755–R765. [PubMed: 16950102]
- Cummings M, Mollaghan D. Repeatability and consistency of female preference behaviours in a northern swordtail, *Xiphophorus nigrensis*. *Animal Behaviour*. 2006; 72:217–224.
- Dall SRX, Houston AI, McNamara JM. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters*. 2004; 7:734–739.
- Dingemans NJ, Both C, Drent PJ, Van Oers K, Van Noordwijk AJ. Repeatability and heritability of exploratory behaviour in great tits from the wild. *Animal Behaviour*. 2002; 64:929–938.
- Dingemans NJ, Reale D. Natural selection and animal personality. *Behaviour*. 2005; 142:1159–1184.
- Dohm MR, Mautz WJ, Looby PG, Gellert KS, Andrade JA. Effects of ozone on evaporative water loss and thermoregulatory behavior of marine toads (*Bufo marinus*). *Environmental Research*. 2001; 86:274–286. [PubMed: 11453678]
- Dohm MR. Repeatability estimates do not always set an upper limit to heritability. *Functional Ecology*. 2002; 16:273–280.
- Dombrovsky Y, Perrin N. On adaptive search and optimal stopping in sequential mate choice. *American Naturalist*. 1994; 144:355–361.
- Du JW, Lofstedt C, Lofqvist J. Repeatability of pheromone emissions from individual female ermine moths *Yponomeuta padellus* and *Yponomeuta rorellus*. *Journal of Chemical Ecology*. 1987; 13:1431–1441. [PubMed: 24302244]
- Edvardsson M, Arnqvist G. No apparent indirect genetic benefits to female red flour beetles preferring males with intense copulatory courtship. *Behavior Genetics*. 2006; 36:775–782. [PubMed: 16557363]
- Falconer, DS.; Mackay, TFC. *Introduction to Quantitative Genetics*. Longman; Essex: 1996.

- Forstmeier W, Birkhead TR. Repeatability of mate choice in the zebra finch: consistency within and between females. *Animal Behaviour*. 2004; 68:1017–1028.
- Freeman-Gallant CR, Rothstein MD. Apparent heritability of parental care in Savannah sparrows. *Auk*. 1999; 116:1132–1136.
- Fuiman LA, Cowan JH. Behavior and recruitment success in fish larvae: repeatability and covariation of survival skills. *Ecology*. 2003; 84:53–67.
- Gabor CR, Aspbury AS. Non-repeatable mate choice by male sailfin mollies *Poecilia latipinna*, in a mating complex. *Behavioral Ecology*. 2008; 19:871–878.
- Gamble S, Lindholm AK, Endler JA, Brooks R. Environmental variation and the maintenance of polymorphism: the effect of ambient light spectrum on mating behaviour and sexual selection in guppies. *Ecology Letters*. 2003; 6:463–472.
- Garamszegi LZ, Hegyi G, Heylen D, Ninni P, De Lope F, Eens M, Møller AP. The design of complex sexual traits in male barn swallows: associations between signal attributes. *Journal of Evolutionary Biology*. 2006a; 19:2052–2066. [PubMed: 17040402]
- Garamszegi LZ, Rosivall B, Hegyi G, Szollosi E, Torok J, Eens M. Determinants of male territorial behavior in a Hungarian flycatcher population: plumage traits of residents and challengers. *Behavioral Ecology and Sociobiology*. 2006b; 60:663–671.
- Gerhardt HC, Dyson ML, Tanner SD. Dynamic properties of the advertisement calls of gray tree frogs: patterns of variability and female choice. *Behavioral Ecology*. 1996; 7:7–18.
- Gerhardt HC, Tanner SD, Corrigan CM, Walton HC. Female preference functions based on call duration in the gray tree frog (*Hyla versicolor*). *Behavioral Ecology*. 2000; 11:663–669.
- Gil D, Slater PJB. Song organization and singing patterns of the willow warbler, *Phylloscopus trochilus*. *Behaviour*. 2000; 137:759–782.
- Gillham MC, Devrijer PWF. Patterns of variation in the acoustic calling signals of chlorionid planthoppers (Homoptera, Delphacidae) coexisting on the common reed *Phragmites australis*. *Biological Journal of the Linnean Society*. 1995; 54:245–269.
- Godin JGJ, Dugatkin LA. Variability and repeatability of female mating preference in the guppy. *Animal Behaviour*. 1995; 49:1427–1433.
- Gosling SD. From mice to men: what can we learn about personality from animal research? *Psychological Bulletin*. 2001; 127:45–86. [PubMed: 11271756]
- Gray CM, Brooke MDL, Hamer KC. Repeatability of chick growth and food provisioning in Manx shearwaters *Puffinus puffinus*. *Journal of Avian Biology*. 2005; 36:374–379.
- Greenfield MD, Siegfried E, Snedden WA. Variation and repeatability of female choice in a chorusing katydid, *Ephippiger ephippiger*: an experimental exploration of the precedence effect. *Ethology*. 2004; 110:287–299.
- Gurevitch J, Morrow LL, Wallace A, Walsh JS. A meta-analysis of competition in field experiments. *American Naturalist*. 1992; 140:539–572.
- Hager BJ, Teale SA. Repeatability of female response to ipsdienol enantiomeric mixtures by pine engraver, *Ips pini* (Coleoptera: Scolytidae). *Journal of Chemical Ecology*. 1994; 20:2611–2622. [PubMed: 24241835]
- Hayes JP, Jenkins SH. Individual variation in mammals. *Journal of Mammalogy*. 1997; 78:274–293.
- Helfenstein F, Wagner RH, Danchin E, Rossi JM. Functions of courtship feeding in black-legged kittiwakes: natural and sexual selection. *Animal Behaviour*. 2003; 65:1027–1033.
- Higgins LA, Waugaman RD. Sexual selection and variation: a multi-variate approach to species-specific calls and preferences. *Animal Behaviour*. 2004; 68:1139–1153.
- Hoffmann AA. Is the heritability for courtship and mating speed in *Drosophila* (fruit fly) low? *Heredity*. 1999; 82:158–162. [PubMed: 10098264]
- Hoffmann, AA. Laboratory and field heritabilities: some lessons from *Drosophila*. In: Mousseau, TA.; Sinervo, B.; Endler, JA., editors. *Adaptive Genetic Variation in the Wild*. Oxford University Press; New York: 2000. p. 200–218.
- Holveck MJ, Riebel K. Preferred songs predict preferred males: consistency and repeatability of zebra finch females across three test contexts. *Animal Behaviour*. 2007; 74:297–309.

- Honza M, Pozgayova M, Prochazka P, Tkadlec E. Consistency in egg rejection behaviour: responses to repeated brood parasitism in the blackcap (*Sylvia atricapilla*). *Ethology*. 2007; 113:344–351.
- Howard RD, Young JR. Individual variation in male vocal traits and female mating preferences in *Bufo americanus*. *Animal Behaviour*. 1998; 55:1165–1179. [PubMed: 9632502]
- Howard RD, Martens RS, Innis SA, Drnevich JM, Hale J. Mate choice and mate competition influence male body size in Japanese medaka. *Animal Behaviour*. 1998; 55:1151–1163. [PubMed: 9632501]
- Hoysak DJ, Godin JGJ. Repeatability of male mate choice in the mosquitofish, *Gambusia holbrooki*. *Ethology*. 2007; 113:1007–1018.
- Isoherranen E, Aspi J, Hoikkala A. Variation and consistency of female preferences for simulated courtship songs in *Drosophila virilis*. *Animal Behaviour*. 1999; 57:619–625. [PubMed: 10196051]
- Janetos AC. Strategies of female mate choice: a theoretical analysis. *Behavioral Ecology and Sociobiology*. 1980; 7:107–112.
- Janzen FJ, Morjan CL. Repeatability of microenvironment-specific nesting behaviour in a turtle with environmental sex determination. *Animal Behaviour*. 2001; 62:73–82.
- Jennions MD, Petrie M. Variation in mate choice and mating preferences: a review of causes and consequences. *Biological Reviews of the Cambridge Philosophical Society*. 1997; 72:283–327. [PubMed: 9155244]
- Jennions MD, Backwell PRY, Passmore NI. Repeatability of mate choice: the effect of size in the African painted reed frog, *Hyperolius marmoratus*. *Animal Behaviour*. 1995; 49:181–186.
- Johnsen TS, Zuk M. Repeatability of mate choice in female red jungle fowl. *Behavioral Ecology*. 1996; 7:243–246.
- Johnson JC, Sih A. Fear, food, sex and parental care: a syndrome of boldness in the fishing spider, *Dolomedes triton*. *Animal Behaviour*. 2007; 74:1131–1138.
- Kamel SJ, Mrosovsky N. Nest site selection in leatherbacks, *Dermochelys coriacea*: individual patterns and their consequences. *Animal Behaviour*. 2004; 68:357–366.
- Kamel SJ, Mrosovsky N. Repeatability of nesting preferences in the hawksbill sea turtle, *Eretmochelys imbricata*, and their fitness consequences. *Animal Behaviour*. 2005; 70:819–828.
- Kent JW, Rankin MA. Heritability and physiological correlates of migratory tendency in the grasshopper *Melanoplus sanguinipes*. *Physiological Entomology*. 2001; 26:371–380.
- Kodric-Brown A, Nicoletto PF. Repeatability of female choice in the guppy: response to live and videotaped males. *Animal Behaviour*. 1997; 54:369–376. [PubMed: 9268469]
- Kokko H. Should advertising parental care be honest? *Proceedings of the Royal Society of London*. 1998; 265:1871–1878. Series B.
- Kolliker M. Ontogeny in the family. *Behavior Genetics*. 2005; 35:7–18. [PubMed: 15674529]
- Kolluru GR. Variation and repeatability of calling behavior in crickets subject to a phonotactic parasitoid fly. *Journal of Insect Behavior*. 1999; 12:611–626.
- Koolhaas JM, Korte SM, De Boer SF, Van Der Vegt BJ, Van Reenen CG, Hopster H, De Jong IC, Ruis MAW, Blokhuis HJ. Coping styles in animals: current status in behavior and stress-physiology. *Neuroscience & Biobehavioral Reviews*. 1999; 23:925–935. [PubMed: 10580307]
- Koteja P, Carter PA, Swallow JG, Garland T. Food wasting by house mice: variation among individuals, families, and genetic lines. *Physiology & Behavior*. 2003; 80:375–383. [PubMed: 14637238]
- Kralj-Fiser S, Scheiber IBR, Blejec A, Moestl E, Kotrschal K. Individualities in a flock of free-roaming greylag geese: behavioral and physiological consistency over time and across situations. *Hormones and Behavior*. 2007; 51:239–248. [PubMed: 17196200]
- Lehtonen TK, Lindström K. Repeatability of mating preferences in the sand goby. *Animal Behaviour*. 2008; 75:55–61.
- Lemon WC. Heritability of selectively advantageous foraging behavior in a small passerine. *Evolutionary Ecology*. 1993; 7:421–428.
- Lessells CM, Boag PT. Unrepeatable repeatabilities: a common mistake. *Auk*. 1987; 104:116–121.
- Light, RJ.; Pillemer, DB. *Summing Up: the Science of Reviewing Research*. Harvard University Press; Boston: 1984.

- MacColl ADC, Hatchwell BJ. Heritability of parental effort in a passerine bird. *Evolution*. 2003; 57:2191–2195. [PubMed: 14575340]
- McDonald PG, Heathcote CF, Clarke MF, Wright J, Kazem AJN. Provisioning calls of the cooperatively breeding bell miner *Manorina melanophrys* encode sufficient information for individual discrimination. *Journal of Avian Biology*. 2007; 38:113–121.
- McElreath R, Strimling P. How noisy information and individual asymmetries can make ‘personality’ an adaptation: a simple model. *Animal Behaviour*. 2006; 72:1135–1139.
- Magellan K, Magurran AE. Behavioural profiles: individual consistency in male mating behaviour under varying sex ratios. *Animal Behaviour*. 2007; 74:1545–1550.
- Malmos KB, Sullivan BK, Lamb T. Calling behavior and directional hybridization between two toads (*Bufo microscaphus***B. woodhousii*) in Arizona. *Evolution*. 2001; 55:626–630.
- Marchetti C, Drent PJ. Individual differences in the use of social information in foraging by captives. *Animal Behaviour*. 2000; 60:131–140. [PubMed: 10924212]
- Martin JGA, Reale D. Temperament, risk assessment and habituation to novelty in eastern chipmunks, *Tamias striatus*. *Animal Behaviour*. 2008; 75:309–318.
- Martins CIM, Schrama JW, Verreth JAJ. The consistency of individual differences in growth, feed efficiency and feeding behaviour in African catfish *Clarias gariepinus* (Burchell 1822) housed individually. *Aquaculture Research*. 2005; 36:1509–1516.
- Masters WM, Raver KAS, Kazial KA. Sonar signals of big brown bats, *Eptesicus fuscus*, contain information about individual identity, age and family affiliation. *Animal Behaviour*. 1995; 50:1243–1260.
- Meffert, LM.; Hagenbuch, KL. The genetic architecture of house fly mating behavior. In: Moscona, AA.; Monroy, A., editors. *Current Topics in Developmental Biology*. Vol. 66. Academic Press; New York, New York: 2005. p. 189–213.
- Michalak P. Repeatability of mating behaviour in Montandon’s newt, *Triturus montandoni* (Caudata salamandridae). *Ethology Ecology & Evolution*. 1996; 8:19–27.
- Missoweit M, Engels S, Sauer KP. Foraging ability in the scorpionfly *Panorpa vulgaris*: individual differences and heritability. *Behavioral Ecology and Sociobiology*. 2007; 61:487–492.
- Møller AP. Repeatability of female choice in a monogamous swallow. *Animal Behaviour*. 1994; 47:643–648.
- Møller AP. Heritability of arrival date in a migratory bird. *Proceedings of the Royal Society of London. Series B*. 2001; 268:203–206.
- Moreno J, Veiga JP, Romasanta M, Sanchez S. Effects of maternal quality and mating status on female reproductive success in the polygynous spotless starling. *Animal Behaviour*. 2002; 64:197–206.
- Morris MR, Nicoletto PF, Hesselman E. A polymorphism in female preference for a polymorphic male trait in the swordtail fish *Xiphophorus cortezi*. *Animal Behaviour*. 2003; 65:45–52.
- Mousseau TA, Roff DA. Natural selection and the heritability of fitness components. *Heredity*. 1987; 59:181–197. [PubMed: 3316130]
- Murphy CG, Gerhardt HC. Mating preference functions of individual female barking tree frogs, *Hyla gratiosa*, for two properties of male advertisement calls. *Evolution*. 2000; 54:660–669. [PubMed: 10937241]
- Nakagawa S, Gillespie DOS, Hatchwell BJ, Burke T. Predictable males and unpredictable females: sex difference in repeatability of parental care in a wild bird population. *Journal of Evolutionary Biology*. 2007; 20:1674–1681. [PubMed: 17714284]
- Nemiroff L, Despland E. Consistent individual differences in the foraging behaviour of forest tent caterpillars (*Malacosoma disstria*). *Canadian Journal of Zoology-Revue Canadienne de Zoologie*. 2007; 85:1117–1124.
- Nespolo RF, Franco M. Whole-animal metabolic rate is a repeatable trait: a meta-analysis. *Journal of Experimental Biology*. 2007; 210:2000–2005. [PubMed: 17515425]
- Nussey DH, Wilson AJ, Brommer JE. The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*. 2007; 20:831–844. [PubMed: 17465894]

- van Oers K, Drent PJ, de Goede P, van Noordwijk AJ. Realized heritability and repeatability of risk-taking behaviour in relation to avian personalities. *Proceedings of the Royal Society of London*. 2004; 271:65–73. Series B.
- van Oers K, de Jong G, van Noordwijk AJ, Kempenaers B, Drent PJ. Contribution of genetics to the study of animal personalities: a review of case studies. *Behaviour*. 2005; 142:1185–1206.
- Otronen M. Variation in sperm precedence during mating in male flies, *Drosophila melanogaster*. *Animal Behaviour*. 1997; 53:1233–1240. [PubMed: 9236019]
- Parker, GA. Mate quality and mating decisions. In: Bateson, P., editor. *Mate Choice*. Cambridge University Press; Cambridge: 1983. p. 141–166.
- Pavlova DZ, Pinxten R, Eens M. Seasonal singing patterns and individual consistency in song activity in female European starlings (*Sturnus vulgaris*). *Behaviour*. 2007; 144:663–680.
- Petersen MR. Reproductive ecology of emperor geese: survival of adult females. *Condor*. 1992; 94:398–406.
- Potti J. Arrival time from spring migration in male pied flycatchers: individual consistency and familial resemblance. *Condor*. 1998; 100:702–708.
- Potti J, Moreno J, Merino S. Repeatability of parental effort in male and female pied flycatchers as measured with doubly labeled water. *Canadian Journal of Zoology-Revue Canadienne de Zoologie*. 1999; 77:174–179.
- Preziosi RF, Fairbairn DJ. Sexual size dimorphism and selection in the wild in the waterstrider *Aquarius remigis*: body size, components of body size and male mating success. *Journal of Evolutionary Biology*. 1996; 9:317–336.
- Quinn JL, Cresswell W. Personality, anti-predation behaviour and behavioural plasticity in the chaffinch *Fringilla coelebs*. *Behaviour*. 2005; 9:1377–1402.
- Radwan J. Heritability of sperm competition success in the bulb mite, *Rhizoglyphus robini*. *Journal of Evolutionary Biology*. 1998; 11:321–327.
- Real L. Search theory and mate choice. I. Models of single-sex discrimination. *American Naturalist*. 1990; 136:376.
- Reale D, Reader SM, Sol D, McDougall PT, Dingemans NJ. Integrating animal temperament within ecology and evolution. *Biological Reviews*. 2007; 82:291–318. [PubMed: 17437562]
- Reinhold K, Reinhold K, Jacoby KJ. Dissecting the repeatability of female choice in the grasshopper *Chorthippus biguttulus*. *Animal Behaviour*. 2002; 64:245–250.
- Rice WR. Analyzing tables of statistical tests. *Evolution*. 1989; 43:223–225.
- Riddell BE, Swain DP. Competition between hatchery and wild coho salmon (*Oncorhynchus kisutch*): genetic variation for agonistic behavior in newly-emerged wild fry. *Aquaculture*. 1991; 98:161–172.
- Riebel K. Early exposure leads to repeatable preferences for male song in female zebra finches. *Proceedings of the Royal Society of London*. 2000; 267:2553–2558. Series B.
- Rivero A, Alatalo RV, Kotiaho JS, Mappes J, Parri S. Acoustic signalling in a wolf spider: can signal characteristics predict male quality? *Animal Behaviour*. 2000; 60:187–194. [PubMed: 10973720]
- Roberts BW, DelVecchio WF. The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin*. 2000; 126:3–25. [PubMed: 10668348]
- Robins RW, Fraley RC, Roberts BW, Trzesniewski KH. A longitudinal study of personality change in young adulthood. *Journal of Personality*. 2001; 69:617–640. [PubMed: 11497032]
- Rosenberg, MS.; Adams, DC.; Gurevitch, J. *Metawin: Statistical Software for Meta-analysis*. Sinauer; Sunderland, Massachusetts: 2000. Version 2.0
- Rosenthal R. The ‘file drawer problem’ and tolerance for null results. *Psychological Bulletin*. 1979; 86:638–641.
- Roulin A. Nonrandom pairing by male barn owls (*Tyto alba*) with respect to a female plumage trait. *Behavioral Ecology*. 1999; 10:688–695.
- Runkle LS, Wells KD, Robb CC, Lance SL. Individual, nightly, and seasonal variation in calling behavior of the gray tree frog, *Hyla versicolor*: implications for energy expenditure. *Behavioral Ecology*. 1994; 5:318–325.

- Rushbrook BJ, Dingemanse NJ, Barber I. Repeatability in nest construction by male three-spined sticklebacks. *Animal Behaviour*. 2008; 75:547–553.
- Ryan MJ, Rand AS. Sexual selection in female perceptual space: how female tungara frogs perceive and respond to complex population variation in acoustic mating signals. *Evolution*. 2003; 57:2608–2618. [PubMed: 14686535]
- Sanvito S, Galimberti F. Source level of male vocalisations in the genus *Mirounga*: repeatability and correlates. *Bioacoustics*. 2003; 14:45–57.
- Sattman DA, Cocroft RB. Phenotypic plasticity and repeatability in the mating signals of *Enchenopa* treehoppers, with implications for reduced gene flow among host-shifted populations. *Ethology*. 2003; 109:981–994.
- Schwagmeyer PL, Mock DW. How consistently are good parents good parents? Repeatability of parental care in the house sparrow, *Passer domesticus*. *Ethology*. 2003; 109:303–313.
- Semlitsch RD, Scott DE, Pechmann JHK, Gibbons JW. Phenotypic variation in the arrival time of breeding salamanders: individual repeatability and environmental influences. *Journal of Animal Ecology*. 1993; 62:334–340.
- Serrano D, Tella JL, Ursua E. Proximate causes and fitness consequences of hatching failure in lesser kestrels *Falco naumanni*. *Journal of Avian Biology*. 2005; 36:242–250.
- Sheldahl LA, Martins EP. The territorial behavior of the western fence lizard, *Sceloporus occidentalis*. *Herpetologica*. 2000; 56:469–479.
- Sih A, Bell AM, Johnson JC. Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution*. 2004a; 19:372–378. [PubMed: 16701288]
- Sih A, Bell AM, Johnson JC, Ziemba R. Behavioral syndromes: an integrative overview. *Quarterly Review of Biology*. 2004b; 79:241–277. [PubMed: 15529965]
- Smith BR, Blumstein DT. Fitness consequences of personality: a meta-analysis. *Behavioral Ecology*. 2008; 19:448–455.
- Smith GR, Doupnik BL. Habitat use and activity level of large American bullfrog tadpoles: choices and repeatability. *Amphibia-Reptilia*. 2005; 26:549–552.
- Smith MJ, Hunter D. Temporal and geographic variation in the advertisement call of the booroolong frog (*Litoria booroolongensis*: Anura: Hylidae). *Ethology*. 2005; 111:1103–1115.
- Spencer RJ, Thompson MB. The significance of predation in nest site selection of turtles: an experimental consideration of macro- and microhabitat preferences. *Oikos*. 2003; 102:592.
- Stirling DG, Reale D, Roff DA. Selection, structure and the heritability of behaviour. *Journal of Evolutionary Biology*. 2002; 15:277–289.
- Sullivan BK. Sexual selection and calling behavior in the american toad (*Bufo americanus*). *Copeia*. 1992; 1992:1–7.
- Sullivan BK, Hinshaw SH. Female choice and selection on male calling behaviour in the grey treefrog *Hyla versicolor*. *Animal Behaviour*. 1992; 44:733–744.
- Tallamy DW, Darlington MB, Pesek JD, Powell BE. Copulatory courtship signals male genetic quality in cucumber beetles. *Proceedings of the Royal Society of London*. 2003; 270:77–82. Series B.
- Tarano Z. Variation in male advertisement calls in the neotropical frog *Physalaemus enesefae*. *Copeia*. 2001; 2001:1064–1072.
- Thornhill R. Cryptic female choice and its implications in the scorpionfly *Harpobittacus nigriceps*. *American Naturalist*. 1983; 122:765.
- Travis J, Woodward BD. Social context and courtship flexibility in male sailfin mollies, *Poecilia latipinna* (Pisces: Poeciliidae). *Animal Behaviour*. 1989; 38:1001–1011.
- Verburgt L, Ferguson JWH, Weber T. Phonotactic response of female crickets on the Kramer treadmill: methodology, sensory and behavioural implications. *Journal of Comparative Physiology A*. 2008; 194:79–96.
- Wagner JWE, Sullivan BK. Sexual selection in the Gulf Coast toad, *Bufo valliceps*: female choice based on variable characters. *Animal Behaviour*. 1995; 49:305–319.
- Wagner JWE, Murray AM, Cade WH. Phenotypic variation in the mating preferences of female field crickets, *Gryllus integer*. *Animal Behaviour*. 1995; 49:1269–1281.

- Whittingham LA, Dunn PO, Stapleton MK. Repeatability of extrapair mating in tree swallows. *Molecular Ecology*. 2006; 15:841–849. [PubMed: 16499706]
- Wilkinson GS, Kahler H, Baker RH. Evolution of female mating preferences in stalk-eyed flies. *Behavioral Ecology*. 1998; 9:525–533.
- Wingfield JC. Control of territorial aggression in a changing environment. *Psychoneuroendocrinology*. 1994; 19:709–721. [PubMed: 7938366]

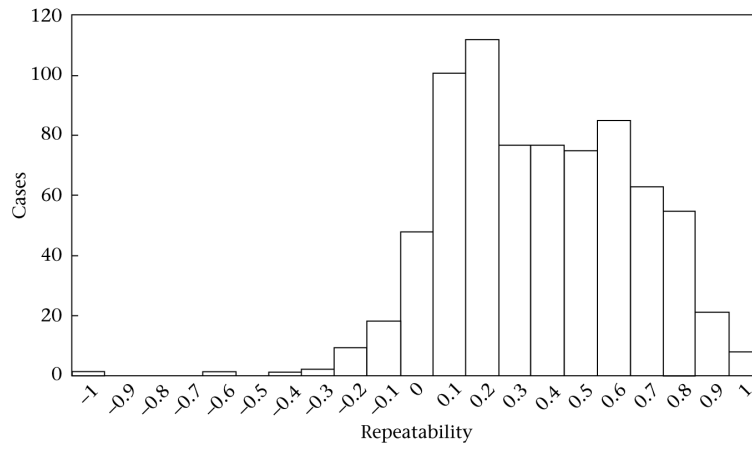


Figure 1. The distribution of repeatability from published studies. Estimates less than zero occur when there are large standard error bars surrounding estimates of between-individual variance components.

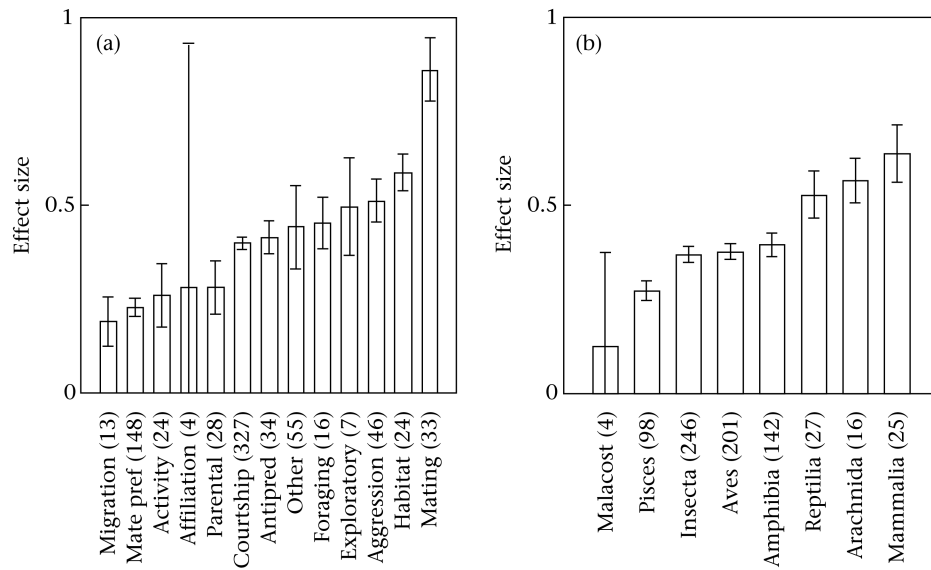


Figure 2. Average effect sizes by (a) behaviour and (b) taxonomic class. Numbers in parentheses indicate sample sizes and error bars denote 95% confidence intervals.

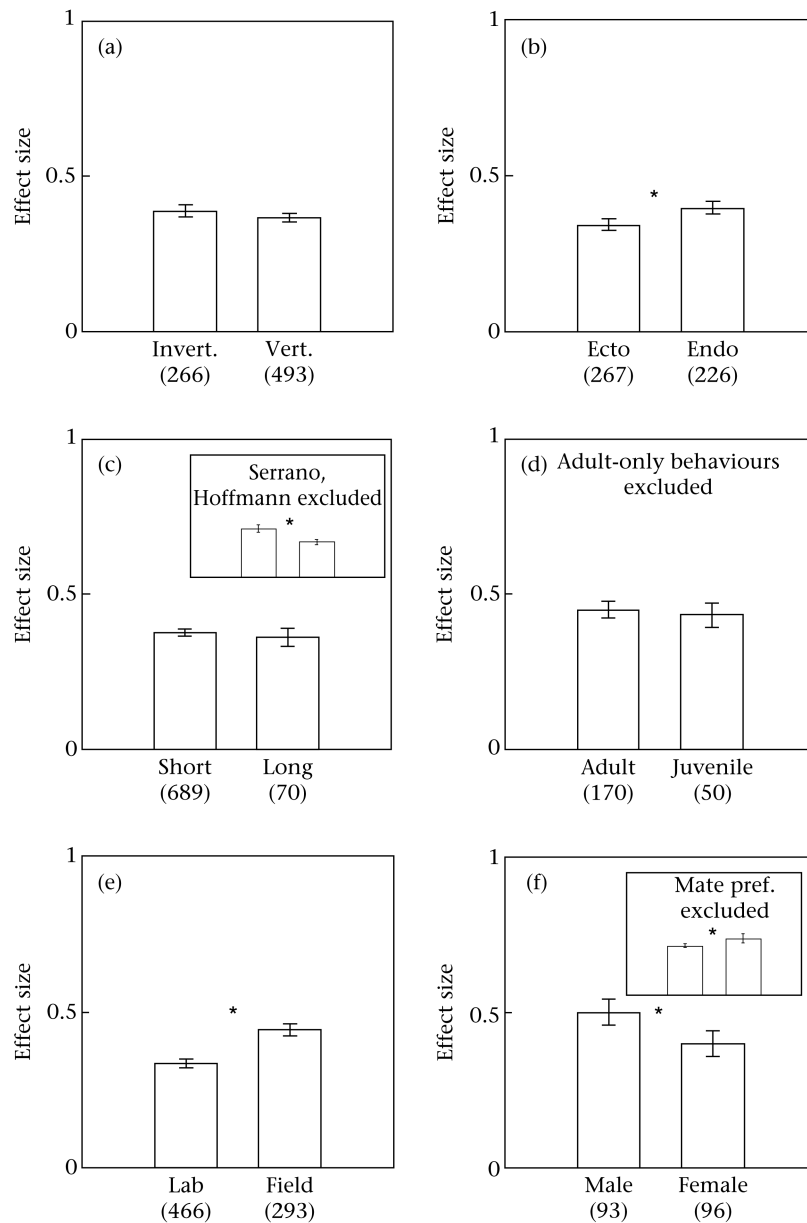


Figure 3. Average effect sizes for each of the grouping variables. The bars show means and 95% confidence intervals and numbers in parentheses indicate sample sizes. An inset graph is included if the overall result changed when a subset of the data set was excluded. (a) Vertebrates versus invertebrates; (b) ectotherms versus endotherms; (c) interval; (d) developmental stage; (e) location; (f) sex.

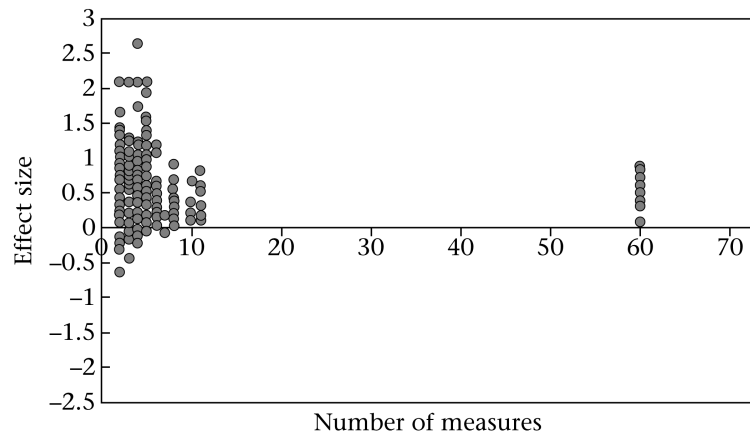


Figure 4. Relationship (nonsignificant) between effect size and the number of times that individuals were measured.

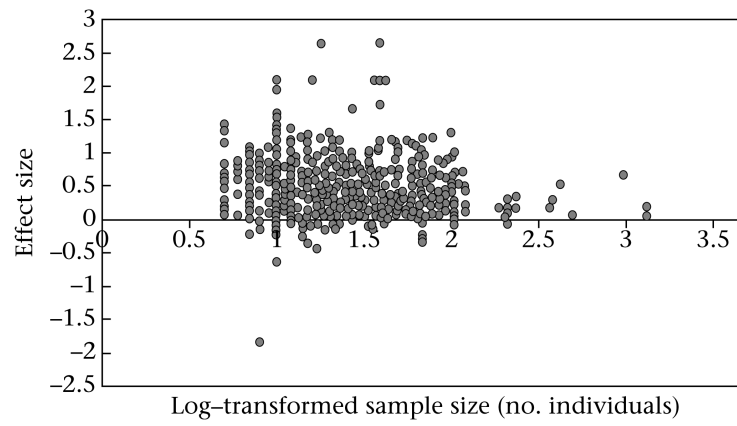


Figure 5. Funnel plot showing effect size as a function of the number of individuals measured (sample size). Note that the X axis is log-transformed.

Table 1

Studies included in meta-analysis organized by behavioural class and species

Author	Species	Taxonomic class
Activity		
Benesh et al. 2008	<i>Asellus aquaticus</i>	Malacostraca
Nemiroff & Despland 2007	<i>Malacosoma disstna</i>	Insecta
Smith & Doupnik 2005	<i>Rana catesbeiana</i>	Amphibia
Kralj-Fiser et al. 2007	<i>Anser anser</i>	Aves
Affiliation		
Kralj-Fiser et al. 2007	<i>Anser anser</i>	Aves
Aggression		
Brown et al. 2006	<i>Acheta domesticus</i>	Insecta
Clark & Moore 1995	<i>Gromphadorhina portentosa</i>	Insecta
Bakker 1986	<i>Gasterosteus aculeatus</i>	Pisces
Riddell & Swain 1991	<i>Oncorhynchus kisutch</i>	Pisces
Kralj-Fiser et al. 2007	<i>Anser anser</i>	Aves
Garamszegi et al. 2006b	<i>Ficedula albicollis</i>	Aves
Pavlova et al. 2007	<i>Sturnus vulgaris</i>	Aves
Antipredator		
Johnson & Sih 2007	<i>Dolomedes triton</i>	Arachnida
Bonte et al. 2003	<i>Erigone atra</i>	Arachnida
Fuiman & Cowan 2003	<i>Sciaenops ocellatus</i>	Pisces
Brodie & Russell 1999	<i>Thamnophis sirtalis sirtalis</i>	Reptilia
Kralj-Fiser et al. 2007	<i>Anser anser</i>	Aves
van Oers et al. 2004	<i>Parus major</i>	Aves
Courtship		
Thornhill 1983	<i>Harpobittacus nigriceps</i>	Insecta
Hager & Teale 1994	<i>Ips pini</i>	Insecta
Aspi & Hoikkala 1993	<i>Drosophila montana, D. littoralis</i>	Insecta
Beeler et al. 1999	<i>Nicrophorus orbicollis</i>	Insecta
Brandt et al. 2005	<i>Achroia grisella</i>	Insecta
Hoffmann 1999	<i>Drosophila melanogaster</i>	Insecta
Kolluru 1999	<i>Teleogryllus oceanicus</i>	Insecta
Meffert & Hagenbuch 2005	<i>Musca domestica</i>	Insecta
Sattman & Cocroft 2003	<i>Enchenopa binotata</i>	Insecta
Allen 1998	<i>Sciarasaga quadrata</i>	Insecta
Butlin & Hewitt 1986	<i>Chorthippus brunneus</i>	Insecta
Gillham & Devrijer 1995	<i>Chloriona spp.</i>	Insecta
Higgins & Waugaman 2004	<i>Gryllus texensis & G. rubens</i>	Insecta
Du et al. 1987	<i>Yponomeuta rorellus</i>	Insecta
Rivero et al. 2000	<i>Hygrolycosa rubrofasciata</i>	Arachnida
Bee & Gerhardt, 2001	<i>Rana catesbeiana</i>	Amphibia

Author	Species	Taxonomic class
Gerhardt et al. 1996	<i>Hyla versicolor</i>	Amphibia
Runkle et al. 1994	<i>Hyla versicolor</i>	Amphibia
Howard & Young 1998	<i>Bufo americanus</i>	Amphibia
Sullivan 1992	<i>Bufo americanus</i>	Amphibia
Ryan & Rand 2003	<i>Physalaemus pustulosus</i>	Amphibia
Smith & Hunter 2005	<i>Litoria booroolongensis</i>	Amphibia
Sullivan & Hinshaw 1992	<i>Hyla versicolor</i>	Amphibia
Tarano 2001	<i>Physalaemus enesefae</i>	Amphibia
Wagner & Sullivan 1995	<i>Bufo valliceps</i>	Amphibia
Michalak 1996	<i>Triturus montandoni</i>	Amphibia
Malmos et al. 2001	<i>Bufo microscaphus</i>	Amphibia
Gamble et al. 2003	<i>Poecilia reticulata</i>	Pisces
Travis & Woodward 1989	<i>Poecilia latipinna</i>	Pisces
Rushbrook et al. 2008	<i>Gasterosteus aculeatus</i>	Pisces
Garamszegi et al. 2006a	<i>Hirundo rustica</i>	Aves
Gil & Slater 2000	<i>Phylloscopus trochilus</i>	Aves
Birkhead & Fletcher 1995	<i>Taeniopygia guttata</i>	Aves
Forstmeier & Birkhead 2004	<i>Taeniopygia guttata</i>	Aves
Helfenstein et al. 2003	<i>Rissa tridactyla</i>	Aves
Sanvito & Galimberti 2003	<i>Mirounga lionina</i> , <i>M. angustirostris</i>	Mammalia
Exploratory behaviour		
Dingemanse et al. 2002	<i>Parus major</i>	Aves
Quinn & Cresswell 2005	<i>Fringilla coelebs</i>	Aves
Foraging		
Missoweit et al. 2007	<i>Panorpa vulgaris</i>	Insecta
Martins et al. 2005	<i>Clarias gariepinus</i>	Pisces
Koteja et al. 2003	<i>Mus domesticus</i>	Mammalia
Habitat selection		
Blanckenhorn & Perner 1994	<i>Aquarius remigis</i>	Insecta
Dohm et al. 2001	<i>Bufo marinus</i>	Amphibia
Smith & Doupnik 2005	<i>Rana catesbeiana</i>	Amphibia
Sheldahl & Martins 2000	<i>Sceloporus occidentalis</i>	Pisces
Kamel & Mrosovsky 2005	<i>Eretmochelys imbricata</i>	Reptilia
Kamel & Mrosovsky 2004	<i>Dermochelys coriacea</i>	Reptilia
Janzen & Morjan 2001	<i>Chrysemys picta</i>	Reptilia
Spencer & Thompson 2003	<i>Emydura macquarii</i>	Reptilia
Serrano et al. 2005	<i>Falco naumanni</i>	Aves
Mate preference		
Brandt et al. 2005	<i>Achroia grisella</i>	Insecta
Greenfield et al. 2004	<i>Ephippiger ephippiger</i>	Insecta
Hager & Teale 1994	<i>Ips pini</i>	Insecta
Isoherranen et al. 1999	<i>Drosophila virilis</i>	Insecta

Author	Species	Taxonomic class
Reinhold et al. 2002	<i>Chorthippus biguttulus</i>	Insecta
Wagner et al. 1995	<i>Gryllus integer</i>	Insecta
Wilkinson et al. 1998	<i>Cyrtodiopsis whitei</i> , <i>C. dalmanni</i>	Insecta
Verburgt et al. 2008	<i>Gryllus bimaculatus</i>	Insecta
Bosch & Marquez 2002	<i>Alytes cisternasii</i>	Amphibia
Bosch et al. 2003	<i>Alytes cisternasii</i> , <i>A. obstetricans</i>	Amphibia
Gerhardt et al. 2000	<i>Hyla versicolor</i>	Amphibia
Howard & Young 1998	<i>Bufo americanus</i>	Amphibia
Jennions et al. 1995	<i>Hyperolius marmoratus</i>	Amphibia
Michalak 1996	<i>Triturus montandoni</i>	Amphibia
Murphy & Gerhardt 2000	<i>Hyla gairdneri</i>	Amphibia
Archard et al. 2006	<i>Poecilia reticulata</i>	Pisces
Aspbury & Basolo 2002	<i>Heterandria formosa</i>	Pisces
Brooks, 2002	<i>Poecilia reticulata</i>	Pisces
Brooks & Endler 2001	<i>Poecilia reticulata</i>	Pisces
Kodric-Brown & Nicoletto 1997	<i>Poecilia reticulata</i>	Pisces
Cummings & Mollaghan 2006	<i>Xiphophorus nigrensis</i>	Pisces
Godin & Dugatkin 1995	<i>Poecilia reticulata</i>	Pisces
Howard et al. 1998	<i>Oryzias latipes</i>	Pisces
Kodric-Brown & Nicoletto 1997	<i>Poecilia reticulata</i>	Pisces
Morris et al. 2003	<i>Xiphophorus nigrensis</i>	Pisces
Hoysak & Godin 2007	<i>Gambusia holbrooki</i>	Pisces
Gabor 2008	<i>Poecilia latipinna</i>	Pisces
Lehtonen & Lindström 2008	<i>Pomatoschistus minutus</i>	Pisces
Banbura 1992	<i>Hirundo rustica</i>	Aves
Møller 1994	<i>Hirundo rustica</i>	Aves
Roulin, 1999	<i>Tyto alba</i>	Aves
Forstmeier & Birkhead 2004	<i>Taeniopygia guttata</i>	Aves
Holveck & Riebel 2007	<i>Taeniopygia guttata</i>	Aves
Johnsen & Zuk 1996	<i>Gallus gallus</i>	Aves
Riebel 2000	<i>Taeniopygia guttata</i>	Aves
Mating		
Edvardsson & Arnqvist 2006	<i>Tribolium castaneum</i>	Insecta
Radwan 1998	<i>Rhizoglyphus robini</i>	Insecta
Tallamy et al. 2003	<i>Diabrotica undecimpunctata howardi</i>	Insecta
Otronen 1997	<i>Dryomyza anilis</i>	Insecta
Michalak 1996	<i>Triturus montandoni</i>	Amphibia
Magellan & Magurran 2007	<i>Poecilia reticulata</i>	Pisces
Travis & Woodward 1989	<i>Poecilia latipinna</i>	Pisces
Whittingham et al. 2006	<i>Tachycineta bicolor</i>	Aves
Migration		

Author	Species	Taxonomic class
Kent & Rankin 2001	<i>Melanoplus sanguinipes fabricius</i>	Insecta
Semlitsch et al. 1993	<i>Ambystoma talpoideum</i>	Amphibia
Battley 2006	<i>Limosa limosa baueri</i>	Aves
Bety et al. 2004	<i>Anser caerulescens atlanticus</i>	Aves
Petersen 1992	<i>Chen canagicus</i>	Aves
Møller 2001	<i>Hirundo rustica</i>	Aves
Potti 1998	<i>Ficedula hypoleuca</i>	Aves
Other		
Brandt & Allen 2004	<i>Uta stansburiana</i>	Reptilia
Kralj-Fiser et al. 2007	<i>Anser anser</i>	Aves
McDonald et al. 2007	<i>Manorina melanophrys</i>	Aves
Masters et al. 1995	<i>Eptesicus fuscus</i>	Mammalia
Parental behaviour		
Honza et al. 2007	<i>Sylvia atricapilla</i>	Aves
Gray et al. 2005	<i>Puffinus puffinus</i>	Aves
Nakagawa et al. 2007	<i>Passer domesticus</i>	Aves
Schwagmeyer & Mock 2003	<i>Passer domesticus</i>	Aves
Freeman-Gallant & Rothstein 1999	<i>Passerculus sandwichensis</i>	Aves
MacColl & Hatchwell 2003	<i>Aegithalos caudatus</i>	Aves
Potti et al. 1999	<i>Ficedula hypoleuca</i>	Aves
Lemon 1993	<i>Taeniopygia guttata</i>	Aves

Table 2

Summary of meta-analysis results for repeatability and behaviour

Variables included in analysis	<i>N</i>	Invert/vert	Ectotherm/endoth ^{er} m*	Test interval (short/long) [†]	Developmental stage (juvenile/adult) [‡]	Location (field/lab)	Sex (male/female)
Entire data set	759	NS	Ec<En	NS	NS	L<F	F<M
Males	388	I<V	Ec<En	L<S	NS	L<F	—
Females	275	V<I	Ec<En	NS	A<J	L<F	—
Invertebrates	266	—	—	?	NS	L<F	M<F
Vertebrates	493	—	Ec<En	NS	NS	L<F	F<M
Ectotherms*	267	—	—	L<S	A<J	L<F	F<M
Endotherms*	226	—	—	NS	NS	L<F	F<M
Short interval	690	NS	NS	—	NS	L<F	F<M
Long interval	69	?	Ec<En	—	?	L<F	NS
Juveniles	50	NS	NS	?	—	L<F	NS
Adults	706	NS	Ec<En	NS	—	L<F	F<M
Field	293	V<I	NS	L<S	A<J [§]	—	NS
Lab	466	V<I	Ec<En	L<S	NS	—	F<M
Excluding courtship	432	V<I	Ec<En	NS	—	L<F	F<M
Excluding mate preference	611	I<V	En<Ec [§]	NS	—	L<F	M<F
Excluding: Serrano et al. 2005; Hoffmann 1999	756	V<I	NS	L<S	NS	L<F	F<M

All comparisons except those with NS were significant at $P < 0.05$ according to the heterogeneity test. Sample sizes reflect the number of estimates considered in all comparisons except sex and developmental stage, where estimates that were based on measures of both males and females or both juveniles and adults were excluded. '?: not enough estimates were available for the comparison.

* Only vertebrates were included in this analysis.

[†] Interval between observations was either less than 1 year ('short') or greater than 1 year ('long').

[‡] Adult-specific behaviours (mate preference, mating, courtship and parental behaviours) were excluded from these comparisons.

[§] Nonsignificant following a sequential Bonferroni correction for multiple comparisons.