Novelty at second glance: a critical appraisal of the novel object paradigm based on meta-analysis

E. Takola [a, *] , E. Tobias Krause [b] , C. Müller [c] , H. Schielzeth [a]

[a] Population Ecology Group, Institute of Ecology and Evolution, Friedrich Schiller University Jena, Jena, Germany
[b] Institute of Animal Welfare and Animal Husbandry, Friedrich-Loeffler-Institute, Celle, Germany
[c] Department of Chemical Ecology, Bielefeld University, Bielefeld, Germany

The study of consistent individual differences has become an important focus in research on animal behaviour. These behavioural differences are typically measured through standardized testing procedures. One frequently used paradigm is the novel object test, in which animals are exposed to unfamiliar objects and their reaction is quantified. We used meta-analysis to evaluate how reliably novel object trials quantify individual differences. Overall, we found repeatability of responses to novel objects was strong and significant and was larger in short-term than in long-term studies. Average sample size and long-term estimates have both increased over the past three decades. Most short-term studies used different novel objects in repeated presentations, while long-term studies used either the same or different novel objects almost equally often. Novelty, the time interval between trials and their interaction together explained little of the total heterogeneity, while between-study heterogeneity remained large. Overall, novel object trials reliably estimate individual differences in behaviour, but results were very heterogeneous even within the same study species, suggesting susceptibility to unknown details in test conditions. Most studies that use novel object trials in a foraging context label the trait as neophobia, while novel object trials in a neutral context are labelled variously as shyness—boldness, exploration—avoidance or neophilia. To avoid ambiguity, we argue for the use of object—neophobia for trials near resources and object—neophilia for trials in a neutral context as the most specific labels for novel object responses.

© 2021 The Authors. Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Consistent individual differences in behaviour are widespread in nature. For a long time, individual differences were considered noise around an optimum niche value, but nowadays it is well established that intraspecific variation is of adaptive importance and can affect reproductive success (Smith & Blumstein, 2008), growth rates (Royauté et al., 2018), metabolic rates (Holtmann et al., 2016) and even population dynamics (Levin et al., 2000). It has also been shown that individual differences in behaviour have a heritable basis (Dochtermann et al., 2015; Stirling et al., 2002) and hence can evolve by natural selection. The study of individual differences has therefore become an important topic in behavioural ecology.

Temporally consistent individual differences in behaviour that are correlated across contexts are variously called animal personalities, behavioural syndromes, coping styles or temperament (Réale et al., 2007). We here refer to them as animal personalities without prejudice towards other terms. One hallmark of animal personality is that individual differences are stable over time (Kaiser & Müller, 2021; Sih et al., 2004). Most empirical studies on animal personality use standardized experimental set-ups with repeated measurements per individual to quantify temporal consistency. It is therefore important to evaluate the suitability of standardized experimental set-ups for the quantification of individual differences. We here focus on reliability of a specific testing paradigm, the novel object test (Yerkes & Yerkes, 1936), and evaluate sources of heterogeneity in study outcomes.

Novel object trials have become popular in recent years. In these trials, animals encounter an item that they had never seen before (thus a novel object) and their behavioural responses are quantified, often as approach latencies or approach distances (Greenberg, 1990; Guenther & Brust, 2017; Yerkes & Yerkes, 1936). Novel object trials are mostly used to quantify shyness—boldness, exploration—avoidance or

* Corresponding author.
  E-mail address: elina.takola@uni-jena.de (E. Takola).

neophilia/neophobia. Shyness–boldness and exploration–avoidance are generally seen as major axes of continuous personality types (which we indicate by the en dash Réale et al., 2007). It is less clear whether neophobia/neophilia represent the same or different personality axes (Mettke-Hofmann, 2014; Mettke-Hofmann et al., 2002) indicated by the slash. Variants of novel object trials aim to separate neophilia and neophobia by varying the context of testing in placing a novel object either in a neutral position (for neophilia) or close to an essential resource such as food for a quantification of neophobia (Greggor et al., 2015).

Shyness–boldness describes an animal's behaviour in a risky but not necessarily novel situation (Réale et al., 2007). Exploration–avoidance refers to the exploration of explicitly novel situations and is often used in a spatial exploration context (Réale et al., 2007). Neophobia/neophilia refer to responses to novelty per se and are sometimes considered as components of exploration–avoidance (Réale et al., 2007). The novel object test is not the only testing paradigm to measure shyness–boldness, exploration–avoidance or neophilia/neophobia. Shyness–boldness is also often quantified by startle response trials, emergence from shelter, response to predator (cues) or by mirror image trials (Ioannou et al., 2008; Noer et al., 2015). Exploratory behaviour is also often quantified by open field or novel-environment trials. Neophilia/neophobia is a more specific term to novel object trials, although it is sometimes also used for novel environment trials (Greggor et al., 2015). Mettke-Hofmann (2012), therefore, proposed to distinguish object neophilia/neophobia for novel object trials from spatial neophilia/neophobia for novel-environment trials.

Independent of the question of labelling is the question of repeated presentations and how they should be best embedded in the experimental design. While the first presentation of a novel object can generate the intended response, upon second presentation of the same item, objects are no longer novel. The second presentation may thus trigger a reduced behavioural response (Berlyne, 1966). The alternative is to use different unfamiliar objects, although these might trigger different responses if, for example, they differ in conspicuousness or perceived riskiness. Greggor et al. (2016) suggested that objects should be used that differ slightly but clearly. However, similarity and differences are ambiguous categories and what might be perceived as similar by some might be seen as different by other individuals. Furthermore, some species might habituate to novel stimuli per se (Réale et al., 2007), such that different novel objects do not trigger the same behavioural response upon second presentation.

The effect of using the same or different objects in repeated trials likely depends on the time interval between repeats. The degree of novelty in these repeated trials is the result of perception and memory and thus depends on the cognitive ability of individuals (Mettke-Hofmann, 2014), but our understanding of animal memory and cognition mechanisms is still incomplete, in particular when it comes to a large range of taxa. It is likely that the effects of novel objects differ between short-term replication (within hours, days or weeks) and long-term replication (after months or years). Therefore, the time interval between trials should be considered when assessing the role of the same or different objects in novel object trials.

We here review the reliability of the novel object paradigm using meta-analytic techniques (Gurevitch et al., 2018; Koricheva et al., 2013). Meta-analysis is a powerful tool for research synthesis in science, as it provides an objective and replicable quantitative overview of literature. Although a common criticism of meta-analytic methods highlights the pooling of incomparable effect sizes (also known as 'apples and oranges problem'), we address the issue of diverse study designs by adding variables as moderators, i.e. the equivalent to covariates in regression models, in meta-regression models, and accounting for multilevel heterogeneity. The use of moderators as fixed effects allows for the identification of context dependencies (such as testing context, wild versus captive populations, etc.) that affect the magnitude of individual differences. Moreover, we account for phylogenetic correlations, since closely related species might react similarly to the same stimuli (Nakagawa & Santos, 2012). Thus, we are able to explore the impact of various effects on the consistency of behavioural traits from multiple studies.

Besides the synthesis of effect sizes by meta-analysis, we present an overview of the most common terms used to describe the measured behaviours. We do not aim to take a stance on the validity of novel object trials to measure these traits, particularly since validity might differ between study systems. We rather aim to summarize how novel object responses are typically labelled in the published literature. By doing so, we take a phenotype-based approach that focuses on behaviour rather than on the underlying emotional, neurological and hormonal processes. We acknowledge that it would be desirable for the field to gain an in-depth process-based understanding for all study systems.

The main aim of our meta-analysis is to evaluate the reliability of the novel object test in quantifying individual differences in behaviour. We do so by conducting a meta-analysis of the magnitude of the correlation between repeated novel object trials, as well as testing context dependencies by moderator and subset analyses. Specifically, we test for (1) the dependency of the correlation on the time gap between measurements, (2) the effect of using the same or different objects during repeated novel object trials and (3) the effect of placing the object in a neutral context or next to a valuable resource (food or nest). Furthermore, we explore other sources of heterogeneity, such as differences between wild and captive individuals, as well as how the use of novel objects has changed over time in terms of sample sizes. Finally, we summarize and discuss variation in terminology when labelling response behaviours and present an overview of the most common response behaviours quantified in novel object trials.

METHODS

We used systematic reviewing techniques to evaluate the properties of the novel object paradigm for quantifying consistent individual differences in behaviour (Koricheva et al., 2013). Our methodology followed the preferred reporting items for systematic reviews and meta-analyses protocol (PRISMA), which aims to increase the robustness of meta-analyses and is based on a checklist (Moher et al., 2009).

Data Collection

We conducted a search in the Web of Science (WoS) Core Collection 5.24. The query included not only the term novel object (novel object*), but also words related to behavioural phenotypes (e.g. neophob*, neophil*, bold*, shy*) and the time range was set to 1990–2020 (see Appendix). The early 1990s were the time when novel object trials were first used systematically to quantify individual differences for context-general behavioural traits (Greenberg, 1990). We also initially searched for the term explorat*, but the number of hits was very large (more than 3000 additional publications). Thus, we used, instead, the combination explorat* and object*, which resulted in 572 additional publications. The WoS Category was limited to Behavioural Sciences and duplicates were removed, resulting in 3984 publications that were used for more detailed screening. The literature search was finalized on 15 March 2021.

Inclusion and Exclusion Criteria

We searched for empirical studies that used novel object trials and quantified the responses of individual animals to these objects.

A novel object should be unfamiliar to focal animals so that we do not expect an evolved attraction to these objects; we thus excluded objects that represent food resources of a species. We did include novel food sources in our analysis if the novel food was sufficiently different from the natural food of a species. This included studies that use artificial dyes to stain natural food if the novel food colour was considered sufficiently novel and unusual.

We screened studies based on the following criteria (Appendix Table A1). First, studies should be done with outbred, nonhuman animals with unimpaired physical condition. Second, studies should use a novel object paradigm, thus excluding presentations of mirror images, live conspecifics, taxidermy mount presentations and food resources of a species. Third, studies should have repeated novel object trials using focal individuals. Fourth, studies should report relevant correlations or repeatability as a measure of individual consistency.

We conducted the screening process in two stages. We first screened titles and abstracts, which excluded 3172 publications, mostly because they did not represent empirical studies, they were done on humans, they did not use systematic novel object trials or they did not study individual differences (Fig. 1). Only clearly nonfitting cases were excluded during abstract screening and ambiguous cases were taken forward to the next step. We next screened full texts of the remaining 812 publications. Screening of full texts was done independently by two people (E.T. and H.S.) and conflicts (10%) were resolved jointly. Full-text screening was focused on the same general criteria and on whether relevant effect and sample sizes were reported. Another 697 publications were excluded during full-text screening (Fig. 1). One study was opportunistically added to the final data set (see Appendix). Consequently, 115 studies matched our inclusion criteria and generated 265 effect sizes.

Data Extraction

We extracted pairwise correlation coefficients (Pearson, Spearman or Kendall) and repeatabilities (R or ICC), as we were interested in the temporal consistency of behavioural responses between trials. Effect sizes typically corresponded to two rounds of novel object trials with the same set of individuals. In cases where more than two rounds of testing were conducted or when multiple responses were quantified, multiple effect sizes were extracted from one study. When combined repeatabilities were reported for more than two trials, we extracted these repeatabilities as the relevant effect sizes. In five cases correlation measures were extracted from graphs using the metaDigitise package, version 1.0.1 (Pick et al., 2019).

For each effect size we extracted information related to (1) publication (year, authors and journal), (2) animals tested (species, sample size and domestication status), (3) testing conditions (novelty of the object in the repeated trials, time interval between trials and context of testing), (4) response behaviour being quantified (specific individual behaviours, response type, see below), (5) analyses being conducted (whether multiple personality traits were assayed, whether repeatabilities were calculated from non-Gaussian generalized linear models) and (6) the terms used to describe the behavioural phenotype (see Appendix Tables A2 and A3 for a detailed description).

The novelty of the object in repeated trials was a parameter of key interest in our analysis. When the novel objects were the same but of different colours, we considered them as different objects. Context of testing was categorized into (1) novel object in neutral position, (2) novel object close to food or (3) novel object close to nest. For domestication status we distinguished between (1) domestic animals tested in an artificial environment, (2) laboratory-reared animals tested in an artificial environment, (3) wild-caught animals tested in an artificial environment and (4) wild animals tested in a natural environment (Mathot et al., 2019). Regarding the response type and behaviour, we recorded the specific trait being quantified (if it was a single behavioural response), whether the response was a composite of multiple behaviours within the same trial (often principal component scores of multiple behaviours scored within the same trial or other synthetic response scores based on multiple components of behaviour) or whether the response was an average calculated across multiple (sub)trials. We did not record transformations being used, since we consider this a decision of individual researchers to best quantify the behaviour, similar to the researcher's decision to record a specific response behaviour and not another. For the same reason, we also did not distinguish between parametric (Pearson) and nonparametric (Spearman or Kendall) correlations. However, 10 studies analysed behavioural phenotypes as binary responses or using Poisson models, and these might produce systematically lower consistency measures; the type of model was therefore recorded.

The time interval was recorded in days, assuming 30 days in a month and 365 days in a year when converting from descriptions in publications. Since our data set included many species with different life histories, we also tried to standardize time intervals by dividing them by the species' life span (compiled from the AnAge database; Tacutu et al., 2018) to express the time interval as a proportion of life span. However, raw time interval measures and lifetime standardized measures were highly correlated ($r = 0.94$), and results were qualitatively unaffected, such that we used log-transformed time interval in days as a moderator in our analysis.

Effect Size and Weighting in Meta-analytic Models

We used R 3.6.3 for all analyses (R Core Team, 2020). Correlation and repeatability measures were transformed using Fisher's Z-transformation as implemented in the escalc function of the metafor package, ver. 2.4.0 (Viechtbauer, 2010). Since we dealt with pairs of observations, correlations and repeatabilities are equivalent; thus, when transformed back, we used r to indicate both. Effect sizes were weighted by the inverse of sampling variance in all analyses. We extracted multiple effect sizes from some studies and therefore it was possible to estimate heterogeneity (variability)
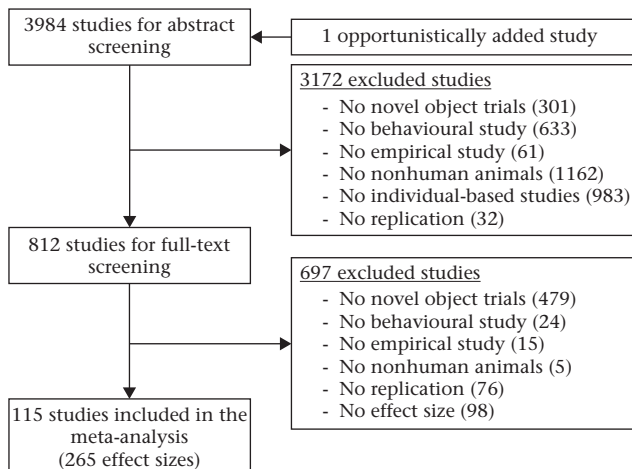

Figure 1. PRISMA diagram with abstract and full-text screening results. Numbers show the number of publications that were excluded or included.

across effect sizes (within-study) as well as between studies, species, etc (see below).

Meta-Analyses and Meta-regressions

We conducted a phylogenetic multilevel meta-analysis to estimate the overall effect. Phylogenetic information was downloaded from Open Tree of Life version ott3.2 (Hinchliff et al., 2015) using the rotl (ver. 3.0.11) R package (Michonneau et al., 2016). After constructing an ultrametric phylogenetic tree (Appendix Fig. A1) using the Grafen (1989) method, we converted the tree to a correlation matrix. This matrix was fitted as a random effect in our meta-analytic model, along with random effects for effect size ID, study ID and species ID. The analysis was performed first using the complete data set and then separately for major taxonomic groups (mammals, birds, fish, reptiles and insects). Weighted random-effect-only meta-analytic models were fitted using the rma function of the metafor package.

Besides the random-effect-only meta-analytic model, we also fitted a meta-regression with moderators (Appendix Table A3), once for the complete data set and once for every major taxonomic group represented by more than 10 publications in our data set (mammals and birds). As moderators we fitted the time interval between repeated trials (log-transformed), novelty (two levels), domestication status (four levels), correlation type (two levels), a binary indicator for non-Gaussian linear models, a binary indicator of whether multiple behavioural tests were performed in the study (other than the novel object), response type (three levels), testing context (three levels) and the interaction of novelty with time interval. The meta-regression models with moderators were also fitted for subsets of the testing context (neutral, food, nest) separately. As above, the random effects of the meta-regression were the effect size ID, study ID, phylogeny and species.

Heterogeneity (I^2) was examined for multiple levels in every model in our meta-analysis, including the subsets of different clades (Nakagawa & Santos, 2012). We also calculated marginal R^2 to estimate the proportion of variance explained by fixed effects (Nakagawa & Schielzeth, 2013). The variance explained by individual predictors was calculated by fitting only the predictor of interest (along with the random components) in a meta-regression model, followed by calculation of marginal R^2. We also calculated τ^2, which reflects true heterogeneity, as it is the measure of variation among effects observed in different studies.

Sensitivity Analyses

We conducted influence diagnostics and sensitivity analyses to evaluate the robustness of our results. For the influence diagnostics we used the influence function of the metafor package, version 2.4.0 (Viechtbauer, 2010), to identify influential studies using Cook's distance and the rstudent test. The diagnostics showed five potential outliers in the data set (Appendix Fig. A2). We therefore refitted the meta-analytic model again while excluding the five influential effect sizes. Since the overall estimate was not significantly affected, we present the analysis of the full data set.

Publication Bias

We tested for publication bias qualitatively through visual inspection of funnel plots and quantitatively by Egger's regression (Egger et al., 1997). Funnel plots were generated by plotting effect sizes against inverse sampling variance and inverse standard error. Egger's regression estimates funnel plot asymmetry as an indicator of publication bias. In addition, we examined the possibility of time lag bias, which is the decrease in effect sizes with increasing year of publication (Trikalinos & Ioannidis, 2006). The test for differences in effect sizes between studies that used novel object trials as the only personality-scoring paradigm versus studies that used multiple measures of personality traits also served as a test for publication bias. We expect studies with a single behavioural measure to be more likely to report statistically significant temporal consistency than studies that report multiple behavioural traits, out of which only a subset might be significantly repeatable.

RESULTS

Screening of 3984 abstracts and full texts resulted in 289 studies that used novel object trials to quantify individual behaviour in nonhuman animals. Of these studies, 213 (74%) replicated novel object trials for all or for a subset of individuals. After excluding 98 studies with repeated novel object trials that did not allow an extraction of effect sizes for temporal consistency, we found 265 effect sizes from 115 studies (Fig. 1) to be included in the analyses. This data set encompassed 70 species (22 mammal, 35 bird, five fish, four reptile and four insect species; Appendix Fig. A1).

Testing Practices

Sample size ranged from five to 567 individuals per effect size estimate (mean ± SD: 47.7 ± 57.6) and increased significantly by about 2.3% per year (effect of year of publication on log(N) sample size: $b = 0.023 \pm 0.008$, $t_{192} = 2.84$, $P = 0.005$; Fig. 2).

The time interval between two consecutive trials ranged between a few hours and 4 years (<0.1%−82% when expressed relative to the expected life span of the focal species). Seventy-two effect sizes (27%) refer to trials repeated on the same (eight effect sizes) or on consecutive days (64 effect sizes). Of the effect sizes, 62% were calculated from replications after at least 1 week, 42% after more than 1 month and 11% after at least 1 year. Studies over longer time periods became more popular over the years with an increase in the time interval between trials of about 14% per year (effect of year of publication on log(time interval): $b = 0.139 \pm 0.023$, $t_{192} = 6.08$, $P < 10^{-5}$; Fig. 2). In the following, we operationally define effect sizes calculated from repeats less than 1 month apart as short-term replications and those with longer intervals as long-term replications.

Seventy-five studies used different objects in repeated trials, 33 used the same objects and seven used both. Most short-term studies (83% of effect sizes for short-term repeatabilities) used different objects, while the same 'novel' objects were used more often when addressing long-term consistencies (only 34% different objects among estimates for long-term repeatabilities; Appendix Table A2). Domestic animals were represented by 40 effect sizes (15%), laboratory-reared animals accounted for 42%, wild-caught animals for 29% and wild animals tested in the wild only 14%.

Eighty-two studies conducted novel object trials in a neutral context (74% of effect sizes), 30 next to a food source (20% of effect sizes) and nine inside or close to the nest (5.6% of effect sizes). Most studies calculated individual consistencies for a specific response behaviour (76% of effect sizes), while some used principal component or other composite scores calculated from multiple behavioural components measured in the same trial (12% of effect sizes) or calculated individual temporal consistencies after averaging across multiple trials (12% of effect sizes). Most studies (86%) used novel object trials along with other standardized personality assays (such as open field trials, startle responses or intruder trials), while only 16 studies (14%) focused on the behavioural consistency for novel object trials only (Appendix Table A2).
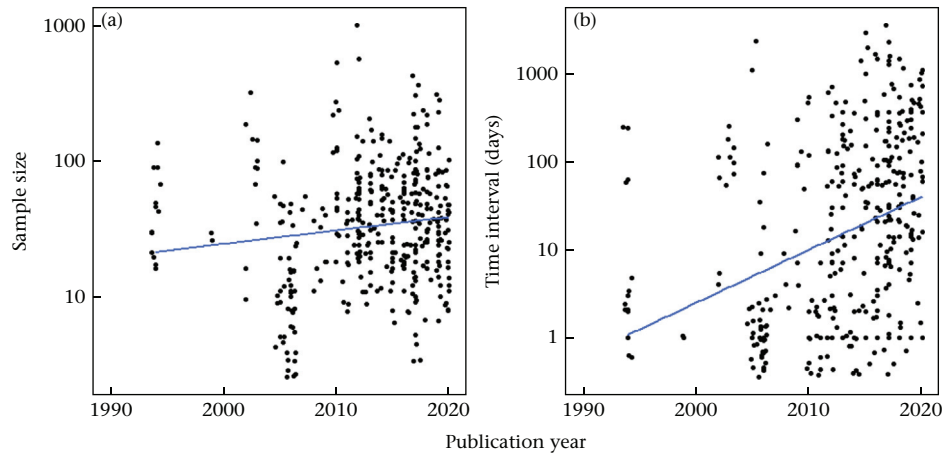
Figure 2. Temporal trends of (a) sample size and (b) time interval between repeated trials. Black dots show raw values of sample sizes and time intervals. The sample size and time interval are shown on a log scale.

Overall Effect Sizes and Heterogeneities

The overall effect of the phylogenetically controlled meta-analysis was strong and significantly greater than zero ($\beta_0 = 0.52$, confidence interval, CI = [0.46, 0.57]), which is equivalent to a correlation of $r = 0.47$. Heterogeneity among effect sizes was high ($I^2_{total} = 80\%$). Variation among studies and among effect sizes accounted for 54% and 25% of this heterogeneity, respectively, while species identity and phylogenetic relationships explained a negligible part. The average short-term repeatability was $r = 0.52$ (equally for time intervals up to 1 week and for time intervals between 1 week and 1 month; Fig. 3), while the average long-term repeatability was $r = 0.40$ ($r = 0.41$ for time intervals of 1 month to 1 year and $r = 0.39$ for time intervals of more than 1 year).

We repeated the analysis separately for the subsets of mammals, birds, fish, reptiles and insects. Mammals, bird, fish and insects showed strong and significant consistencies of behaviour (all $r > 0.40$), while individual consistency was low and nonsignificant for reptiles ($r = 0.074$; Table 1). Total heterogeneity was particularly high in the subsets of mammals, birds and fish (all $I^2_{total} > 80\%$), but not for insects and reptiles ($I^2_{total} < 4\%$; Table 2). Between-study heterogeneity was particularly high for the subset of mammals and fish ($I^2_{study} > 62\%$), moderate for birds ($I^2_{study} = 32\%$) and low

for insects and reptiles (Table 2). The amount of real heterogeneity, expressed by tau-squared, was $\tau^2 = 0.078$, CI = [0.07, 0.11].

We also fitted the meta-analytic model for each subset of testing context (neutral, food, nest). When the object was placed in a neutral position repeatability was $r = 0.46$, whereas when it was placed close to food or the nest repeatabilities were higher ($r = 0.55$ and $r = 0.53$, respectively). However, the large estimate for objects close to the nest was partly caused by a larger proportion of short-term studies (nine of 15 effect sizes) in this subset.

The Impact of Novelty and Time Interval

The amount of total heterogeneity in overall effect indicated scope for effects of moderators. We therefore fitted a meta-regression with novelty, time and their interaction as moderators. This meta-regression was first fitted for the full data set and then for the subsets of taxa with >10 studies (i.e. mammals and birds). The moderators explained in total 3% of the variance and did not have a significant effect on the correlation ($Q_M = 5.35$, $P = 0.15$). Novelty had a low and nonsignificant effect on behavioural consistency and, as expected, time yielded a negative estimate (shorter time intervals resulted in higher repeatability estimates). The estimate for the interaction was negative (the effect of time interval was stronger if objects were different), but not significantly different from zero ($\beta_{int} = -0.0001$, CI = [-0.0414, 0.0412], $P = 0.92$). Similar trends were observed in the subsets of mammals and birds (Fig. 4). In the overall model and the subset of birds, these moderators explained around 4%, but in the subset of mammals, they explained 7%. Even though meta-regression did not show a significant effect of time, long-term consistencies seem to be markedly lower than short-term consistencies when the data are broken down to time interval classes (Fig. 3).

The Impact of Other Moderators

We explored effects of additional moderators in the meta-regression model by fitting each of them in a meta-regression model. As for novelty and time interval, we fitted these meta-regressions once for the whole data set and once for each taxon with >10 studies (i.e. mammals and birds). Domestication status accounted for a low fraction of variance ($R^2_{dom} = 1\%$) and was not significantly correlated with the overall effect size. In the subset of mammals, domestication status explained 2% of variation and for birds 1%. However, the levels of domestication status did not show consistent estimates across different subsets of the data (Fig. 4).
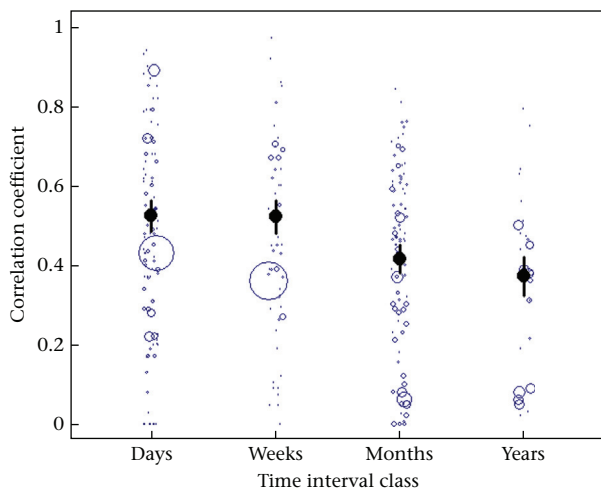

Figure 3. Correlation coefficients for four classes of time intervals between repeated trials. Open dots show correlation coefficients (dot size scaled by sample size) and black dots and bars show random-effect-only meta-analytic model estimates ± SE. Days = 0–6 days, weeks = 7–30 days, months = 31–364 days, years = 365 + days.

Table 1
Summaries and results from phylogenetic multilevel meta-analyses

	$N_{E.S.}$	$N_{studies}$	$N_{species}$	Zr	SE	LCI	UCI	z	P	r
Overall	265	115	70	0.52	0.03	0.46	0.57	17.37	<0.0001	0.47
Mammals	93	36	22	0.54	0.05	0.43	0.65	9.69	<0.0001	0.49
Birds	141	62	35	0.54	0.042	0.46	0.62	12.75	<0.0001	0.49
Fish	19	10	5	0.52	0.15	0.22	0.81	3.43	<0.001	0.46
Reptiles	8	3	4	0.07	0.05	−0.02	0.17	1.49	0.17	0.074
Insects	4	4	4	0.43	0.07	0.28	0.58	5.70	<0.0001	0.40

LCI and UCI indicate the lower and upper limits of the 95% confidence intervals; $N_{E.S.}$ indicates the number of effect sizes; $N_{studies}$ indicates the number of studies; $N_{species}$ indicates the number of species; Zr indicates the Z-transformed correlation calculated by the meta-analytic model; r indicates the correlation (back-transformed) calculated by the meta-analytic model.

Table 2
Total heterogeneity in effect sizes (%) across hierarchical levels of random effects for the overall data set and for subsets of the data

	$I^2_{species}$	I^2_{phylo}	I^2_{study}	$I^2_{e.s.}$	I^2_{total}
Overall	0	0	54	26	80
Mammals	0	0	62	19	81
Birds	6.2	6.2	32	35	80
Fish	0.1	0.1	73	7	80
Reptiles	0	0	4	0	4
Insects	0	0	0	0	0

Accuracy is given to one decimal only for effects <10%. $I^2_{species}$ indicates the heterogeneity observed at the level of species; I^2_{phylo} indicates the heterogeneity observed at the level of phylogeny; I^2_{study} indicates the heterogeneity observed at the level of study ID; $I^2_{e.s.}$ indicates the heterogeneity observed at the level of effect size ID; I^2_{total} indicates the total heterogeneity.

Testing context explained only 1% of the total heterogeneity. The type of response (single behaviours, aggregates of multiple components and averages across trials) had no significant effect and the effect of estimation by non-Gaussian models was also nonsignificant. All moderators explained less than 4% in all cases except in the subset of mammals in which the response type explained 11%.

Publication Bias and Sensitivity Analysis

For sensitivity analysis, we refitted the overall meta-analytic model without five particularly influential studies (Appendix Fig. A2). The estimate of the overall effect marginally decreased from $\beta_0 = 0.52$ to 0.49 (CI = [0.44, 0.54]), whereas the total heterogeneity dropped from 80% to 72%.

Visual inspection of the funnel plot showed only weak asymmetry of effect sizes (Fig. 5). However, Egger's test identified significant asymmetry ($t_{199} = 3.04$, $P = 0.003$) but a subsequent trim-and-fill method estimated zero missing effect sizes. We tested for time lag bias by fitting a meta-regression with publication year as a predictor. The slope showed a negative trend ($\beta = -0.02$, CI = [−0.04, 0.0036], $Q_M = 5.41$, $P = 0.02$) which explained 3.5% of variance. Studies that reported multiple behavioural traits had nonsignificantly larger consistency estimates than studies that focused on novel object trials. This result is thus not indicative of publication bias.

Reproducibility Within Species

The amount of heterogeneity explained by species was estimated to be zero in the overall meta-analysis. However, most species were used only in one or a few studies. Three species, though,
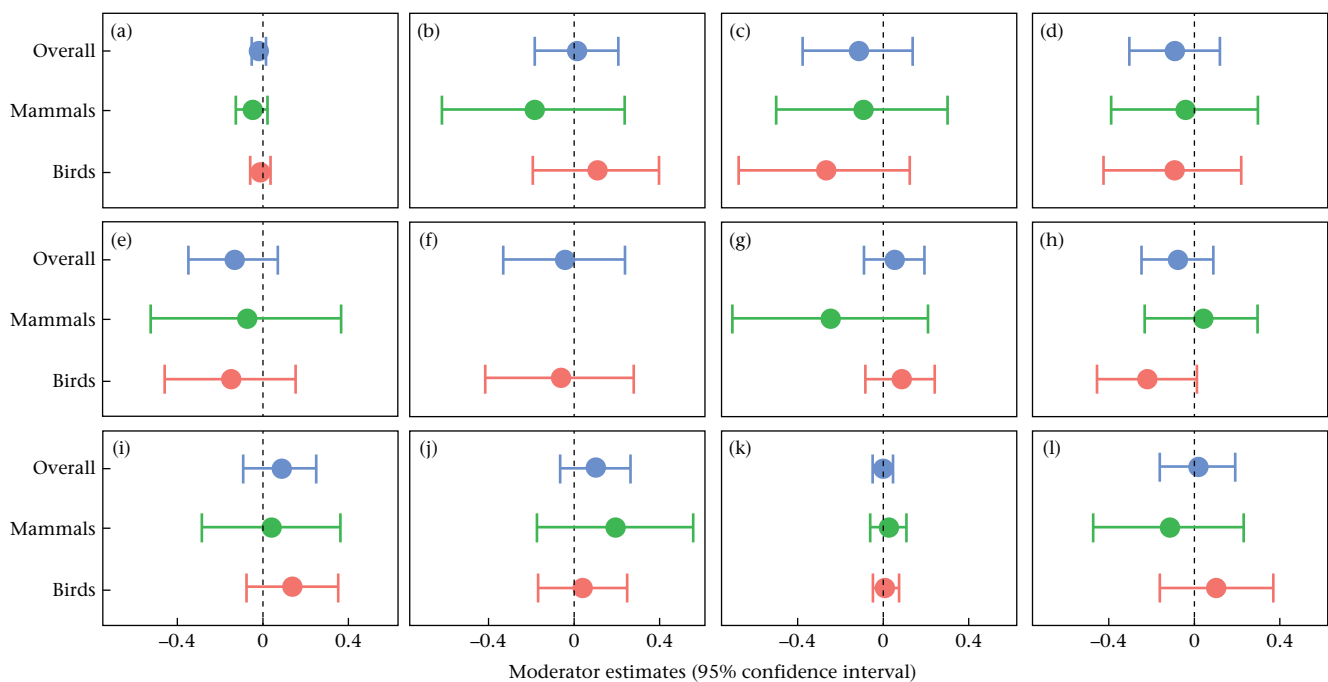

Figure 4. Forest plot showing the results of meta-regressions using the full data set and subsets of bird and mammal species. Moderators were (a) time interval, (b) novelty, (c), (d), (e) the domestication gradient of testing and (f), (g) testing context (position of novel object relative to resources), (h) non-Gaussian models, (i) PCA scores, (j) composite measures of behaviour, (k) the interaction of novelty and time and (l) multiple behavioural assays during the study. The reference category combination in the model was the same object, wild-caught animals tested in the wild, neutral context, Gaussian models, single behaviour, repeatability estimate and only novel objects as the only personality trait being assayed.
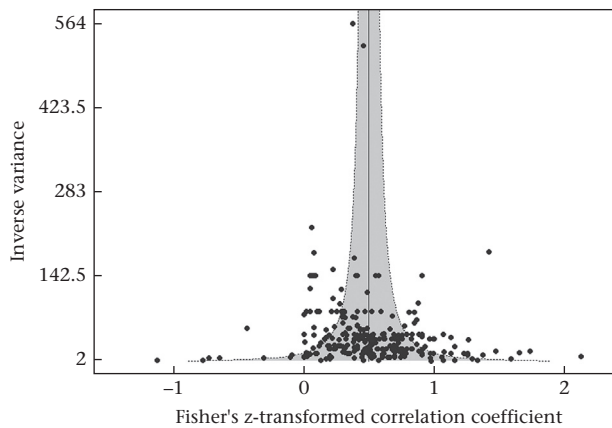
Figure 5. Funnel plot of inverse sampling variation against effect sizes for an assessment of publication bias.

were used in more than three studies, so we inspected the consistency of estimates within these species (guinea pig, *Cavia aperea*, zebra finch, *Taeniopygia guttata* and great tit, *Parus major*) more closely. Estimates of individual consistency in response to novelty of the guinea pig were done with two laboratory populations (domestic guinea pigs and wild-derived cavies) in the same research laboratory, all with the novel object in a neutral context and they used either latency to approach or the number of touches as a response. Nevertheless, estimates varied widely (Fig. 6). Estimates with zebra finches were all done in seven different outbred captive laboratory populations (including the study with the second largest sample size in our data set) and were performed either in a neutral context or close to food. Estimates varied widely (Fig. 6) within contexts and even within the same population with multiple estimates. Estimates for the great tit were particularly heterogeneous in context (neutral, near food or near nest) and they were conducted in the wild, in the laboratory with wild-caught birds or with laboratory-bred individuals. However, the scatter of estimates was similar to the cases of guinea pigs and zebra finches (Fig. 6).

Terminology

Most studies (48 studies, 42%) labelled responses to novel objects as either neophobia (38 studies, 32%), neophilia (10 studies, 9%), shyness—boldness (31 studies, 27%) or exploration—avoidance (22 studies, 19%), while more rarely occurring labels were fearfulness (five studies), approach—avoidance (two), risk responsiveness (two) and activity (one) (Appendix Tables A4—A6). Eight studies did

not use any general labels for the traits being measured. Labelling was associated with testing context, with an even stronger bias towards neophobia when novel objects were placed next to food or nests (70% across these two contexts) and a more even distribution across neophobia/neophilia, shyness—boldness and exploration—avoidance when the novel object was in a neutral place (Appendix Table A4, Fig. A3).

DISCUSSION

Our meta-analysis of individual temporal consistencies, as quantified in novel object trials, revealed an overall strong and significant repeatability in responses to novel objects ($r = 0.47$). This estimate is substantially larger than an estimate of average repeatability in behaviour, $r = 0.37$ (Bell et al., 2009), which demonstrates that the novel object paradigm is a useful and reliable way to quantify consistent individual differences between individuals. The long-term repeatability was overall lower than the short-term repeatability. Any state-dependent causes of individual differences are likely to be temporally autocorrelated, such that short-term repeatabilities are expected to be higher than long-term repeatabilities. Furthermore, environmental variables also tend to be temporally autocorrelated, which can lead to pseudorepeatability, in particular when individuals select their different microenvironments or individualized niches (resulting in 'recurrent environments'; Dupré, 2014). Environmental autocorrelation is likely to affect short-term repeatabilities more strongly than longer-term repeatabilities.

General Evaluation of the Novel Object Paradigm

The rather high overall repeatability in response to novel objects shows that novel object trials provide a generally suitable paradigm for the quantification of temporal consistencies and behavioural differences between individuals. However, we found substantial heterogeneity in effect sizes, mostly between studies and to a lesser degree between species. The large heterogeneity poses the question of whether differences between studies reflect genuine differences between populations or whether they reflect differences in the uncontrolled aspects of the experimental set-up. There are many reasons why populations may differ in the relative magnitude of individual differences. For example, populations might have been exposed to different selective regimes, such as urban versus rural populations (Miranda et al., 2013), captive versus wild populations (Herborn et al., 2010) or different housing conditions among captive populations (Zocher et al., 2020). In addition,
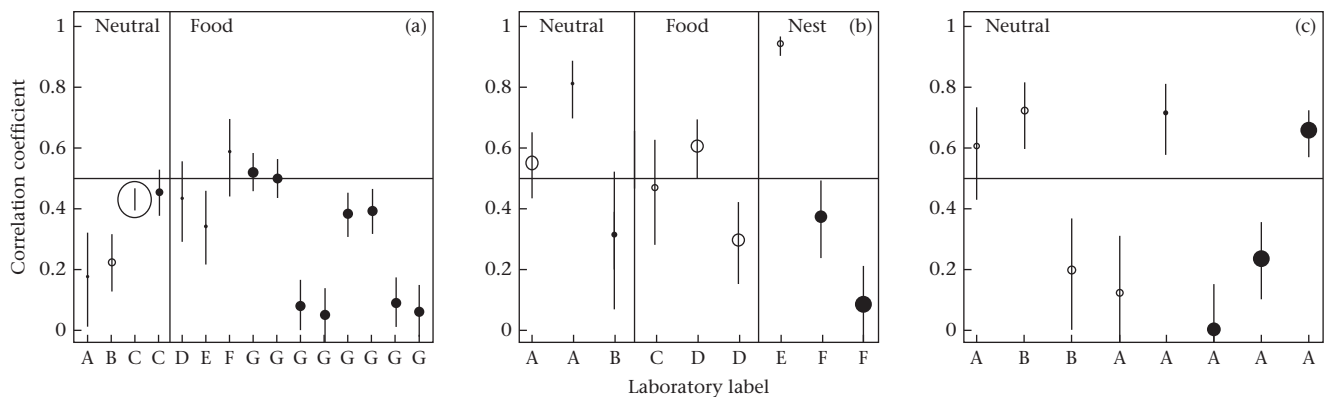
Figure 6. Repeatability (estimates ± SE) of behaviour in different testing contexts (position of novel object relative to food or the nest) for the most popular species in our data set: (a) zebra finch, *T. guttata*, (b) great tit, *P. major*, and (c) guinea pig, *C. aperea*. Open and filled dots are used to indicate short and long time intervals, respectively. The size of the dots is scaled by sample size. Different letters for the laboratory label mark different populations of animals. Horizontal lines represent zero.

population size might affect the amount of standing genetic variation and thus the phenotypic variation for behavioural traits. Moreover, the season or other environmental differences might affect the magnitude of state-dependent individual differences (Sih et al., 2015), which might arguably be larger in the wild than in captivity, although empirical evidence is scarce. Any such differences in population background, population size and the magnitude of state variation could give raise to heterogeneity in effect sizes. Hence, heterogeneity might well have a biological origin that is relevant for our understanding of variation in individual behavioural traits related to personality.

However, it is also important to consider the nonexclusive alternative, that experimental set-ups of novel object trials differ in how reliably they capture individual differences. This is an important concern, since most studies used response to novelty as a trait to be correlated with other behaviours (Guenther & Brust, 2017), endocrine measurements (Arnold et al., 2016) or reproductive success (Schielzeth et al., 2011) and these relationships might be systematically underestimated if behavioural measurements contain substantial measurement error. For example, experimental set-ups might assess different responses depending on short-term state fluctuation (e.g. in the state of hunger). Furthermore, we usually know far too little about which objects might trigger sufficient interest in animals and which objects are perceived as intimidating, which is likely to be influenced by size, colour, shape and odour of the object as well as familiarity with similar-looking objects. Objects that are perceived as scary or intimidating might trigger fear responses, while neutral objects might trigger more of an explorative response. Some of the heterogeneity in effect sizes might not represent differences in behaviour between individuals, but rather variation in novel object trials themselves, thus potentially impairing robustness of the paradigm.

Under the premise that novel object trials are designed to measure context-general personality traits, we would expect consistent findings within species or at least within populations of the same species. However, the between-species component of heterogeneity was very low and replicate studies within three specific species (guinea pig and cavies, zebra finch and great tit) show substantial differences in estimates (Fig. 6). It could be argued that these reflect genuine population differences in the case of the zebra finch and great tit, since these studies were all done mostly in different populations. This explanation seems unlikely in the case of the guinea pig, however, as all studies were performed in the same laboratory and replicated with two different populations of animals. The inconsistency across these studies is thus of potential concern.

Specific Design Decisions

Choice of novel objects is very important, as it can induce different reactions (Greggor et al., 2015). Interestingly, experimental design decision such as the use of the same or different novel objects for the test replications seems to play a very minor role in influencing the magnitude of individual differences, since on average estimates were not significantly affected. However, the vast majority of short-term repeatability estimates used different novel objects (Appendix Fig. A4). This is a useful decision for the test set-up for two reasons. First, shorter time intervals will make it more likely that individuals remember specific objects (Bell et al., 2009). Second, novel object trials are intended to quantify context-general aspects of behaviour; hence, it is the repeatable component in response to different objects that matters in most cases. Over extended time periods, however, it seems less likely that individuals would remember a specific encounter. Indeed, about half of the long-term studies used the same novel objects, with retesting done months or years after a first encounter. This design had no

systematic effect on the magnitude of consistent individual differences, suggesting that the quantified behaviours are as comparable as trials with different novel objects.

The phylogenetic relationship matrix that we fitted in the meta-analytic model did not explain a significant amount of variation. However, when splitting the data by classes of animals, we found not only that mammals and birds were the most popular subjects in novel object trials, but also that they showed higher average repeatabilities. This might be due to sampling bias within groups but might also indicate that these groups are particularly suitable for testing novel object responses. It seems plausible that highly visual organisms, such as birds and many day-active mammals, are particularly suitable for novel object trials. The biased use of different groups is in agreement with the uneven representation of taxonomic classes observed by Rosenthal et al. (2017). Our view on the consistency of responses to novel objects is thus strongly dominated by these two groups of vertebrates.

Overall, we found only minor publication bias in the published record. Furthermore, we found no difference in the magnitude of repeatability estimates between studies that focus on novel object responses as the sole behaviour as compared to the large number of studies that combined multiple testing paradigms to evaluate personality dimensions. The robustly large amount of individual variation in response to novel objects reliably produces significant repeatabilities, such that there is little scope for selective reporting and thus publication bias (Forstmeier et al., 2017). Encouragingly, both the average sample size of repeatedly tested individuals and the time interval between the test repeats have increased over the years. In recent years, a typical sample was around 50–60 individuals retested after about 1–2 months. If this trend continues, it will reveal more reliable estimates and provide more data on long-term behavioural consistency.

Terminology

Besides the question of how well novel object trials allow a quantification of consistent individual differences, another important question is which personality axis they are best ascribed to: a problem of labelling and validity. Many publications in our analysis dive straight into labelling. Many published abstracts, for example, use terms like 'boldness' and 'exploration' without stating how these were quantified. However, mentioning the label is usually not conclusive enough (Kaiser & Müller, 2021). Boldness and exploration are particularly ambiguous labels, since they are also often used for startle response and open field tests, respectively. Neophilia, or even more precisely object neophilia, is a less ambiguous term that is almost exclusively used for behaviour in novel object trails. We suggest that abstracts, and not only methods sections, should clearly state the testing paradigms that were used in the quantification of individual differences.

Neophobia/neophilia might be seen as a component of exploration–avoidance. Neophobia, in particular, might also be interpreted as a behavioural response to a risky situation. It is often unclear whether an animal will perceive a novel object as risky or neutral. If this were clear, one could draw a fine line between neophobia as response to risky novelty (more in line with shyness–boldness) and neophilia as response to neutral novelty (more in line with exploration–avoidance). Réale et al. (2007) indeed grouped neophobia/neophilia with exploration–avoidance and excluded novel situations from the definition of shyness–boldness. However, our survey shows that many empirical papers (including those published after 2007) do not follow this definition, since novel object responses are often interpreted as a measure of shyness–boldness. In most cases, how animals perceive the situation will not be known and a differentiation will thus remain ambiguous.

The most frequent terms used to describe the animals' reactions to a novel object were neophobia (37 studies), neophilia (10 studies), shyness—boldness (31 studies) and exploration—avoidance (22 studies). A few studies used multiple labels. An important difference between these terms is the testing context used for their assessment. The term neophobia was mostly used when the novel object was placed in or close to a food source or close to the nest (thus amplifying the risk aspect). This seems suitable if animals are motivated to approach a food source or a nest but are prevented from approaching by 'fear of the new'. The effect is likely to be even stronger when the object is placed close to the nest than when placed close to food. When the novel object was placed in a neutral position (e.g. in the middle of the testing cage), the use of terms was distributed more evenly, which can be interpreted as novelty being seen as something to be discovered and explored (thus amplifying the exploratory aspect), or as a risky situation that induces neophobia and thus requires boldness to approach.

It may be worth studying whether novel object responses in a neutral context are better correlated with exploration and novel object responses close to food or the nest better correlated with startle responses. However, we are not aware of any systematic review on that matter. For the time being, it seems best to label responses to novel objects as object neophilia (in a neutral context for nonscary objects) and object neophobia (in a non-neutral context) and to clearly specify whether objects were placed close to a resource. The non-neutral context might be either a food source or a nest. Systematic studies across a range of taxa are needed to establish whether object neophobia/neophilia is best grouped with shyness—boldness or exploration—avoidance or kept separately as one axis neophobia—neophilia or two axes (neophobia and neophilia). We predict the best solution will depend on subtleties of the set-up in how neophobic/neophilic tendencies affect behaviour in risky and novel situations.

Over- and under-labelling in personality research gave rise to a phenomenon called the jingle—jangle fallacy (Carter et al., 2013) where there is either over-pooling (jingle) or over-separating of terms (jangle). Both cases can be found in the novel object literature. More experimental evidence is needed to disentangle or to merge behavioural terms, due to unknown underlying mechanisms. For the time being, we think it is overall better to err on the jangle side and to over-separate rather than conflate different personality axes.

Conclusions

We evaluated current practices of the novel object test and estimated average effects when novel object trials are used to assess the magnitude of temporally consistent individual differences. We found that most studies replicated novel object trials, that sample sizes have increased significantly over time and that there are more long-term than short-term assessments of behavioural consistencies. This illustrates overall good and improving research practice. Average consistencies tended to be even slightly larger than average behavioural consistencies across different testing paradigms, illustrating that the novel object paradigm is suitable for the quantification of individual differences in behaviour. Moderators in our analysis did not explain a large amount of heterogeneity. Almost all short-term studies used different novel objects for the trial repeats, which seems important, while long-term studies used either the same or different novel objects. Our results suggest that the latter decision does not affect the results. While there is some variation in how behavioural traits are labelled, the most specific description would be object neophobia/neophilia, which can be interpreted as a component of shyness—boldness or exploration—avoidance. This suggestion is based on the level of the behavioural response only and cognitive and neurophysiological studies are needed to explore the processes that lead to these behaviours. Such studies might suggest a different categorization (splitting or pooling of terms), but we think it is important to establish a clear labelling at the level of behaviour in the first place. Owing to overlap of labels with other testing paradigms, we suggest that abstracts of published papers specify the testing set-up rather than referring only to labels.

Author Contributions

H.S. and E.T.K. conceived the project. E.T. conducted data collection and analysis with the help of HS. E.T. drafted the manuscript. All authors contributed to the interpretation of results and to the revision of the manuscript.

Acknowledgments

References

Amy, M., Ung, D., Beguin, N., & Leboucher, G. (2017). Personality traits and behavioural profiles in the domestic canary are affected by sex and photoperiod. *Ethology, 123*, 885—893.

An, Y. S., Kriengwatana, B., Newman, A. E., MacDougall-Shackleton, E. A., & MacDougall-Shackleton, S. A. (2011). Social rank, neophobia and observational learning in black-capped chickadees. *Behaviour, 148*, 55—69.

Arnold, K. E., Herborn, K. A., Henderson, L. J., Adam, A., Alexander, L., & Evans, N. (2016). Individual variation in corticosterone and personality traits in the blue tit Cyanistes caeruleus. *Behaviour, 153*, 1611—1637.

Basic, D., Winberg, S., Schjolden, J., Krogdahl, A., & Hoglund, E. (2012). Context-dependent responses to novelty in Rainbow trout (*Oncorhynchus mykiss*), selected for high and low post-stress cortisol responsiveness. *Physiology & Behavior, 105*, 1175—1181.

Baxter-Gilbert, J., Riley, J. L., & Whiting, M. J. (2019). Bold new world: Urbanization promotes an innate behavioral trait in a lizard. *Behavioral Ecology and Sociobiology, 73*, 105.

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: A meta-analysis. *Animal Behaviour, 77*, 771—783.

Berlyne, D. E. (1966). Curiosity and exploration. *Science, 153*, 25—33.

Bibi, N., Wei, Y. S., Xu, H. W., Liang, J. N., Hussain, I., Ahmad, M. S., & Wang, H. T. (2019). Personality is associated with dominance in a social feeding context in the great tit. *Behaviour, 156*, 1419—1434.

Boogert, N. J., Reader, S. M., & Laland, K. N. (2006). The relation between social rank, neophobia and individual learning in starlings. *Animal Behaviour, 72*, 1229—1239.

Brust, V., & Guenther, A. (2017). Stability of the Guinea pigs personality - cognition - linkage over time. *Behavioural Processes, 134*, 4—11.

Burns, J. G. (2008). The validity of three tests of temperament in guppies (*Poecilia reticulata*). *Journal of Comparative Psychology, 122*, 344—356.

Carere, C., Drent, P. J., Privitera, L., Koolhaas, J. M., & Groothuis, T. G. G. (2005). Personalities in great tits, *Parus major*: Stability and consistency. *Animal Behaviour, 70*, 795—805.

Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: What are behavioural ecologists measuring? *Biological Reviews, 88*, 465—475.

Christensen, J., Beblein, C., & Malmkvist, J. (2020). Development and consistency of fearlessness in horses from foal to adult. *Applied Animal Behaviour Science, 232*, Article 105106. https://doi.org/10.1016/j.applanim.2020.105106

Collins, S. M., Hatch, S. A., Elliott, K. H., & Jacobs, S. R. (2019). Boldness, mate choice and reproductive success in *Rissa tridactyla. Animal Behaviour, 154*, 67—74.

Coutant, T., Bagur, S., & Gilbert, C. (2018). Development of an observational quantitative temperament test in three common parrot species. *Applied Animal Behaviour Science, 202*, 100—111.

Damas-Moreira, I., Riley, J. L., Harris, D. J., & Whiting, M. J. (2019). Can behaviour explain invasion success? A comparison between sympatric invasive and native lizards. *Animal Behaviour, 151*, 195—202.

Dammhahn, M., & Almeling, L. (2012). Is risk taking during foraging a personality trait? A field test for cross-context consistency in boldness. *Animal Behaviour, 84*, 1131—1139.

Dardenne, S., Ducatez, S., Cote, J., Poncin, P., & Stevens, V. M. (2013). Neophobia and social tolerance are related to breeding group size in a semi-colonial bird. *Behavioral Ecology and Sociobiology, 67*, 1317—1327.

David, M., Auclair, Y., & Cezilly, F. (2011). Personality predicts social dominance in female zebra finches, *Taeniopygia guttata*, in a feeding context. *Animal Behaviour, 81*, 219—224.

DeRango, E. J., Schwarz, J. F. L., Kalberer, S., Piedrahita, P., Páez-Rosas, D., & Krüger, O. (2019). Intrinsic and maternal traits influence personality during early life in Galapagos sea lion, *Zalophus wollebaeki*, pups. *Animal Behaviour, 154*, 111—120.

Devost, I., Jones, T. B., Cauchoix, M., Montreuil-Spencer, C., & Morand-Ferron, J. (2016). Personality does not predict social dominance in wild groups of black-capped chickadees. *Animal Behaviour, 122*, 67—76.

Dochtermann, N. A., Schwab, T., & Sih, A. (2015). The contribution of additive genetic variation to personality variation: Heritability of personality. *Proceedings of the Royal Society B, 282*, 20142201.

Dupré, J. (2014). The role of behaviour in the recurrence of biological processes. *Biological Journal of the Linnean Society, 112*, 306—314.

Edwards, H. A., Burke, T., & Dugdale, H. L. (2017). Repeatable and heritable behavioural variation in a wild cooperative breeder. *Behavioral Ecology, 28*, 668—676.

Edwards, H. A., Dugdale, H. L., Richardson, D. S., Komdeur, J., & Burke, T. (2018). Extra-pair parentage and personality in a cooperatively breeding bird. *Behavioral Ecology and Sociobiology, 72*, 37.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629—634.

Ensminger, A. L., & Westneat, D. F. (2012). Individual and sex differences in habituation and neophobia in house sparrows (*Passer domesticus*). *Ethology, 118*, 1085—1095.

Exnerova, A., Jezova, D., Stys, P., Doktorovova, L., Rojas, B., & Mappes, J. (2015). Different reactions to aposematic prey in 2 geographically distant populations of great tits. *Behavioral Ecology, 26*, 1361—1370.

Farrell, T. M., Weaver, K., An, Y. S., & MacDougall-Shackleton, S. A. (2012). Song bout length is indicative of spatial learning in European starlings. *Behavioral Ecology, 23*, 101—111.

Finkemeier, M. A., Trillmich, F., & Guenther, A. (2016). Match-mismatch experiments using photoperiod expose developmental plasticity of personality traits. *Ethology, 122*, 80—93.

Forstmeier, W., Wagenmakers, E.-J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings — a practical guide. *Biological Reviews, 92*, 1941—1968.

Fox, R. A., & Millam, J. R. (2010). The use of ratings and direct behavioural observation to measure temperament traits in Cockatiels (*Nymphicus hollandicus*). *Ethology, 116*, 59—75.

Frost, A. J., Thomson, J. S., Smith, C., Burton, H. C., Davis, B., Watts, P. C., & Sneddon, L. U. (2013). Environmental change alters personality in the rainbow trout, *Oncorhynchus mykiss*. *Animal Behaviour, 85*, 1199—1207.

Funghi, C., Leitão, A. V., Ferreira, A. C., Mota, P. G., & Cardoso, G. C. (2015). Social dominance in a gregarious bird is related to body size but not to standard personality assays. *Ethology, 121*, 84—93.

Gabriel, P. O., & Black, J. M. (2010). Behavioural syndromes in steller's jays: The role of time frames in the assessment of behavioural traits. *Animal Behaviour, 80*, 689—697.

Garamszegi, L. Z., Markó, G., Szász, E., Zsebők, S., Azcárate, M., Herczeg, G., & Török, J. (2015). Among-year variation in the repeatability, within- and between-individual, and phenotypic correlations of behaviors in a natural population. *Behavioral Ecology and Sociobiology, 69*, 2005—2017.

Garamszegi, L. Z., Rosivall, B., Rettenbacher, S., Markó, G., Zsebők, S., Szöllősi, E., Eens, M., Potti, J., & Török, J. (2012). Corticosterone, avoidance of novelty, risk-taking and aggression in a wild bird: No evidence for pleiotropic effects. *Ethology, 118*, 621—635.

Grace, J. K., & Anderson, D. J. (2014). Personality correlates with contextual plasticity in a free-living, long-lived seabird. *Behaviour, 151*, 1281—1311.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society B, 326*, 119—157.

Greenberg, R. (1990). Feeding neophobia and ecological plasticity: A test of the hypothesis with captive sparrows. *Animal Behaviour, 39*, 375—379.

Greenberg, J. R., & Holekamp, K. E. (2017). Human disturbance affects personality development in a wild carnivore. *Animal Behaviour, 132*, 303—312.

Greggor, A., Jolles, J. W., Thornton, A., & Clayton, N. S. (2016). Seasonal changes in neophobia and its consistency in rooks: The effect of novelty type and dominance position. *Animal Behaviour, 121*, 11—20.

Greggor, A. L., Masuda, B., Flanagan, A. M., & Swaisgood, R. R. (2020). Age-related patterns of neophobia in an endangered island crow: Implications for conservation and natural history. *Animal Behaviour, 160*, 61—68.

Greggor, A. L., Thornton, A., & Clayton, N. (2015). Neophobia is not only avoidance: Improving neophobia tests by combining cognition and ecology. *Current Opinion in Behavioral Sciences, 6*, 82—89.

Grindstaff, J. L., Hunsaker, V. R., & Cox, S. N. (2012). Maternal and developmental immune challenges alter behavior and learning ability of offspring. *Hormones and Behavior, 62*, 337—344.

Guenther, A., & Brust, V. (2017). Individual consistency in multiple cognitive performance: Behavioural versus cognitive syndromes. *Animal Behaviour, 130*, 119—131.

Guenther, A., Brust, V., Dersen, M., & Trillmich, F. (2014). Learning and personality types are related in cavies (*Cavia aperea*). *Journal of Comparative Psychology, 128*, 74—81.

Guenther, A., Finkemeier, M. A., & Trillmich, F. (2014). The ontogeny of personality in the wild Guinea pig. *Animal Behaviour, 90*, 131—139.

Guenther, A., Groothuis, A. G. G., Krüger, O., & Goerlich-Jansson, V. C. (2018). Cortisol during adolescence organises personality traits and behavioural syndromes. *Hormones and Behavior, 103*, 129—139.

Guenther, A., & Trillmich, F. (2013). Photoperiod influences the behavioral and physiological phenotype during ontogeny. *Behavioral Ecology, 24*, 402—411.

Guido, J. M., Biondi, L. M., Vassalo, A. I., & Muzio, R. N. (2017). Neophobia is negatively related to reversal learning ability in females of a generalist bird of prey, the chimango caracara, *Milvago chimango*. *Animal Cognition, 20*, 591—602.

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature, 555*, 175—182.

Gyuris, E., Feró, O., & Barta, Z. (2012). Personality traits across ontogeny in firebugs, *Pyrrhocoris apterus*. *Animal Behaviour, 84*, 103—109.

Haage, M., Bergvall, U. A., Maran, T., Kiik, K., & Angerbjörn, A. (2013). Situation and context impacts the expression of personality: The influence of breeding season and test context. *Behavioural Processes, 100*, 103—109.

Hebert, O. L., Lavin, L. E., Marks, J. M., & Dzieweczynski, T. L. (2014). The effects of 17α-ethinyloestradiol on boldness and its relationship to decision making in male Siamese fighting fish. *Animal Behaviour, 87*, 203—212.

Herborn, K. A., Macleod, R., Miles, W. T. S., Schofield, A. N. B., Alexander, L., & Arnold, K. E. (2010). Personality in captivity reflects personality in the wild. *Animal Behaviour, 79*, 835—843.

Hinchliff, C. E., Smith, S. A., Allman, J. F., Burleigh, J. G., Chaudhary, R., Coghill, L. M., Crandall, K. A., Deng, J., Drew, B. T., Gazis, R., Gude, K., Hibbett, D. S., Katz, L. A., Laughinghouse, H. D., McTavish, E. J., Midford, P. E., Owen, C. L., Ree, R. H., Rees, J. A., et al. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences USA, 112*, 12764—12769.

Hirata, M., & Arimoto, C. (2018). Novel object response in beef cattle grazing a pasture as a group. *Behavioural Processes, 157*, 315—319.

Holtmann, B., Lagisz, M., & Nakagawa, S. (2016). Metabolic rates, and not hormone levels, are a likely mediator of between-individual differences in behaviour: A meta-analysis. *Functional Ecology, 31*, 685—696.

Hopkins, W. D., & Bennett, A. J. (1994). Handedness and approach: Avoidance behavior in chimpanzees (*Pan*). *Journal of Experimental Psychology-Animal Behavior Processes, 20*, 413—418.

Ioannou, C. C., Payne, M., & Krause, J. (2008). Ecological consequences of the bold—shy continuum: The effect of predator boldness on prey risk. *Oecologia, 157*, 177.

Jäger, J., Schradin, C., Pillay, N., & Rimbach, R. (2017). Active and explorative individuals are often restless and excluded from studies measuring resting metabolic rate: Do alternative metabolic rate measures offer a solution? *Physiology & Behavior, 174*, 57—66.

Janczak, A. M., Pedersen, L. J., & Bakken, M. (2003). Aggression, fearfulness and coping styles in female pigs. *Applied Animal Behaviour Science, 81*, 13—28.

Johnson, Z., Brent, L., Alvarenga, J. C., Comuzzie, A. G., Shelledy, W., Ramirez, S., Cox, L., Mahaney, M. C., Huang, Y.-Y., Mann, J. J., Kaplan, J. R., & Rogers, J. (2015). Genetic influences on response to novel objects and dimensions of personality in *Papio* baboons. *Behavior Genetics, 45*, 215—227.

Jolles, J. W., Ostojic, L., & Clayton, N. S. (2013). Dominance, pair bonds and boldness determine social-foraging tactics in rooks, *Corvus frugilegus*. *Animal Behaviour, 85*, 1261—1269.

Jolly, C. J., Webb, J. K., Gillespie, G. R., Hughes, N. K., & Phillips, B. L. (2019). Bias averted: Personality may not influence trappability. *Behavioral Ecology and Sociobiology, 73*, 129.

Kaiser, M. I., & Müller, C. (2021). What is an animal personality? *Biology and Philosophy, 36*, 1—25.

Kerman, K., Miller, L., & Sewall, K. (2018). The effect of social context on measures of boldness: Zebra finches (*Taeniopygia guttata*) are bolder when housed individually. *Behavioural Processes, 157*, 18—23.

Koricheva, J., Gurevitch, J., & Mengersen, K. (2013). *Handbook of meta-analysis in ecology and evolution*. Princeton, NJ: Princeton University Press.

Krams, I. A., Vrublevska, J., Sepp, T., Abolins-Abols, M., Rantala, M. J., & Mierauskas, P. (2014). Sex-specific associations between nest defence, exploration and breathing rate in breeding pied flycatchers. *Ethology, 120*, 492—501.

Krause, E. T., Krüger, O., & Schielzeth, H. (2017). Long-term effects of early nutrition and environmental matching on developmental and personality traits in zebra finches. *Animal Behaviour, 128*, 103—115.

Krebs, R., Linnenbrink, M., & Guenther, A. (2019). Validating standardised personality tests under semi-natural conditions in wild house mice (*Mus musculus domesticus*). *Ethology, 125*, 761—773.

Kurvers, R., de Hoog, S. I. V., van Wieren, S. E., Ydenberg, R. C., & Prins, H. H. T. (2012). No evidence for negative frequency-dependent feeding performance in relation to personality. *Behavioral Ecology, 23*, 51—57.

Kurvers, R., Eijkelenkamp, B., van Oers, K., van Lith, B., van Wieren, S. E., Ydenberg, R. C., & Prins, H. H. T. (2009). Personality differences explain leadership in barnacle geese. *Animal Behaviour, 78*, 447—453.

Le Vin, A. L., Mable, B. K., Taborsky, M., Heg, D., & Arnold, K. E. (2011). Individual variation in helping in a cooperative breeder: Relatedness versus behavioural type. *Animal Behaviour, 82*, 467—477.

Lermite, F., Peneaux, C., & Griffin, A. S. (2017). Personality and problem-solving in common mynas (*Acridotheres tristis*). *Behavioural Processes, 134*, 87—94.

Levin, P. S., Tolimieri, N., Nicklin, M., & Sale, P. F. (2000). Integrating individual behavior and population ecology: The potential for habitat-dependent population regulation in a reef fish. *Behavioral Ecology, 11*, 565–571.

Malmkvist, J., & Hansen, S. W. (2002). Generalization of fear in farm mink, *Mustela vison*, genetically selected for behaviour towards humans. *Animal Behaviour, 64*, 487–501.

Martin-Wintle, M. S., Shepherdson, D., Zhang, G., Huang, Y., Luo, B., & Swaisgood, R. R. (2017). Do opposites attract? Effects of personality matching in breeding pairs of captive giant pandas on reproductive success. *Biological Conservation, 207*, 27–37.

Mathot, K. J., Dingemanse, N. J., & Nakagawa, S. (2019). The covariance between metabolic rate and behaviour varies across behaviours and thermal types: meta-analytic insights. *Biological Reviews, 94*, 1056–1074.

Mazza, V., Dammhahn, M., Eccard, J. A., Palme, R., Zaccaroni, M., & Jacob, J. (2019). Coping with style: Individual differences in responses to environmental variation. *Behavioral Ecology and Sociobiology, 73*, 142.

Mazza, V., Eccard, J. A., Zaccaroni, M., Jacob, J., & Dammhahn, M. (2018). The fast and the flexible: Cognitive style drives individual variation in cognition in a small mammal. *Animal Behaviour, 137*, 119–132.

McCune, K., Jablonski, P., Lee, S.-i., & Ha, R. (2018). Evidence for personality conformity, not social niche specialization in social jays. *Behavioral Ecology, 29*, 910–917.

Meagher, R. K., von Keyserlingk, M. A. G., Atkinson, D., & Weary, D. M. (2016). Inconsistency in dairy calves' responses to tests of fearfulness. *Applied Animal Behaviour Science, 185*, 15–22.

Medina-Garcia, A., Jawor, J. M., & Wright, T. F. (2017). Cognition, personality, and stress in budgerigars, *Melopsittacus undulatus. Behavioral Ecology, 28*, 1504–1516.

Meehan, C. L., & Mench, J. A. (2002). Environmental enrichment affects the fear and exploratory responses to novelty of young Amazon parrots. *Applied Animal Behaviour Science, 79*, 75–88.

Mettke-Hofmann, C. (2012). Head colour and age relate to personality traits in Gouldian finches. *Ethology, 118*, 906–916.

Mettke-Hofmann, C. (2014). Cognitive ecology: Ecological factors, life-styles, and cognition. *WIREs Cognitive Science, 5*, 345–360.

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

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

Michelena, P., Sibbald, A. M., Erhard, H. W., & McLeod, J. E. (2009). Effects of group size and personality on social foraging: The distribution of sheep across patches. *Behavioral Ecology, 20*, 145–152.

Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact with the Open Tree of Life data. *Methods in Ecology and Evolution, 7*, 1476–1481.

Miller, K. A., Garner, J. P., & Mench, J. A. (2005). The test-retest reliability of four behavioural tests of fearfulness for quail: A critical evaluation. *Applied Animal Behaviour Science, 92*, 113–127.

Miller, K. A., Garner, J. P., & Mench, J. A. (2006). Is fearfulness a trait that can be measured with behavioural tests? A validation of four fear tests for Japanese quail. *Animal Behaviour, 71*, 1323–1334.

Miranda, A. C., Schielzeth, H., Sonntag, T., & Partecke, J. (2013). Urbanization and its effects on personality traits: A result of microevolution or phenotypic plasticity? *Global Change Biology, 19*, 2634–2644.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology, 62*, 1006–1012.

Moldoff, D. E., & Westneat, D. F. (2017). Foraging sparrows exhibit individual differences but not a syndrome when responding to multiple kinds of novelty. *Behavioral Ecology, 28*, 732–743.

Monestier, C., Morellet, N., Verheyden, H., Gaillard, J.-M., Bideau, E., Denailhac, A., Lourtet, B., Cebe, N., Picot, D., Rames, J.-L., & Hewison, A. J. M. (2017). Neophobia is linked to behavioural and haematological indicators of stress in captive roe deer. *Animal Behaviour, 126*, 135–143.

Morinay, J., Daniel, G., Gustafsson, L., & Doligez, B. (2019). No evidence for behavioural syndrome and genetic basis for three personality traits in a wild bird population. *Animal Behaviour, 153*, 69–82.

Nakagawa, S., & Santos, E. S. A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology, 26*, 1253–1274.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*, 133–142.

Noer, C. L., Needham, E. K., Wiese, A. S., Balsby, T. J., & Dabelsteen, T. (2015). Context matters: Multiple novelty tests reveal different aspects of shyness-boldness in farmed american mink (*Neovison vison*). *PLoS One, 10*, Article e0130474.

Noer, C. L., Needham, E. K., Wiese, A. S., Balsby, T. J. S., & Dabelsteen, T. (2016). Personality matters: Consistency of inter-individual variation in shyness-boldness across non-breeding and pre-breeding season despite a fall in general shyness levels in farmed American mink (*Neovison vison*). *Applied Animal Behaviour Science, 181*, 191–199.

Overington, S. E., Cauchard, L., Cote, K. A., & Lefebvre, L. (2011). Innovative foraging behaviour in birds: What characterizes an innovator? *Behavioural Processes, 87*, 274–285.

Pedersen, V. (1994). Long-term effects of different handling procedures on behavioral, physiological and production-related parameters in silver foxes. *Applied Animal Behaviour Science, 40*, 285–296.

Perals, D., Griffin, A. S., Bartomeus, I., & Sol, D. (2017). Revisiting the open-field test: What does it really tell us about animal personality? *Animal Behaviour, 123*, 69–79.

Pick, J. L., Nakagawa, S., & Noble, D. W. A. (2019). Reproducible, flexible and high-throughput data extraction from primary literature: The metaDigitise R package. *Methods in Ecology and Evolution, 10*, 426–431.

Pogány, A., Vincze, E., Szurovecz, Z., Kosztolanyi, A., Barta, Z., Székely, T., & Riebel, K. (2018). Personality assortative female mating preferences in a songbird. *Behaviour, 155*, 481–503.

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Rangassamy, M., Athari, S. K., Monclus, R., Boissier, M. C., Bessis, N., & Rodel, H. G. (2016). Personality modulates proportions of CD4(+) regulatory and effector T cells in response to socially induced stress in a rodent of wild origin. *Physiology & Behavior, 167*, 255–264.

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews, 82*, 291–318.

Rockwell, C., Gabriel, P. O., & Black, J. M. (2012). Bolder, older, and selective: Factors of individual-specific foraging behaviors in steller's jays. *Behavioral Ecology, 23*, 676–683.

Rohrer, K. N., & Ferkin, M. H. (2020). Long-term repeatability and stability of three personality traits in meadow voles. *Ethology, 126*, 791–802.

Rosenthal, M. F., Gertler, M., Hamilton, A. D., Prasad, S., & Andrade, M. C. B. (2017). Taxonomic bias in animal behaviour publications. *Animal Behaviour, 127*, 83–89.

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

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

Schielzeth, H., Bolund, E., Kempenaers, B., & Forstmeier, W. (2010). Quantitative genetics and fitness consequences of neophilia in zebra finches. *Behavioral Ecology, 22*, 126–134.

Schielzeth, H., Bolund, E., Kempenaers, B., & Forstmeier, W. (2011). *Quantitative genetics and fitness consequences of neophilia in zebra finches* Behavioral Ecology, 22, 126–134.

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

Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004). Behavioral syndromes: An integrative overview. *Quarterly Review of Biology, 79*, 241–277.

Sih, A., Mathot, K. J., Moirón, M., Montiglio, P., Wolf, M., & Dingemanse, N. J. (2015). Animal personality and state–behaviour feedbacks: A review and guide for empiricists. *Trends in Ecology & Evolution, 30*, 50–60.

Siviter, H., Deeming, D. C., Rosenberger, J., Burman, O. H. P., Moszuti, S. A., & Wilkinson, A. (2017). The impact of egg incubation temperature on the personality of oviparous reptiles. *Animal Cognition, 20*, 109–116.

Smith, B. R., & Blumstein, D. T. (2008). Fitness consequences of personality: A meta-analysis. *Behavioral Ecology, 19*, 448–455.

Smith, B. R., & Blumstein, D. T. (2012). Structural consistency of behavioural syndromes: Does predator training lead to multi-contextual behavioural change? *Behaviour, 149*, 187–213.

Soha, J. A., Peters, S., Anderson, R. C., Searcy, W. A., & Nowicki, S. (2019). Performance on tests of cognitive ability is not repeatable across years in a songbird. *Animal Behaviour, 158*, 281–288.

Sol, D., Griffin, A. S., & Bartomeus, I. (2012). Consumer and motor innovation in the common myna: The role of motivation and emotional responses. *Animal Behaviour, 83*, 179–188.

Spake, J. R., Gray, K. A., & Cassady, J. P. (2012). Relationship between backtest and coping styles in pigs. *Applied Animal Behaviour Science, 140*, 146–153.

Stirling, D. G., Réale, D., & Roff, D. A. (2002). Selection, structure and the heritability of behaviour. *Journal of Evolutionary Biology, 15*, 277–289.

Stöwe, M., Bugnyar, T., Heinrich, B., & Kotrschal, K. (2006). Effects of group size on approach to novel objects in ravens (*Corvus corax*). *Ethology, 112*, 1079–1088.

Stöwe, M., Bugnyar, T., Loretto, M., Schloegl, C., Range, F., & Kotrschal, K. (2006). Novel object exploration in ravens (*Corvus corax*): Effects of social relationships. *Behavioural Processes, 73*, 68–75.

Stuber, E. F., Araya-Ajoy, Y. G., Mathot, K. J., Mutzel, A., Nicolaus, M., Wijmenga, J. J., Mueller, J. C., & Dingemanse, N. J. (2013). Slow explorers take less risk: A problem of sampling bias in ecological studies. *Behavioral Ecology, 24*, 1092–1098.

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J. W., Fraifeld, V. E., & de Magalhaes, J. P. (2018). Human ageing genomic resources: New and updated databases. *Nucleic Acids Research, 46*, D1083–D1090.

Thodberg, K., Jensen, K. H., & Herskin, M. (1999). A general reaction pattern across situations in prepubertal gilts. *Applied Animal Behaviour Science, 63*, 103–119.

Tan, M. K., & Tan, H. T. W. (2019). Individual- and population-level personalities in a floriphilic katydid. *Ethology, 125*, 114–121.

Tobler, M., & Sandell, M. I. (2007). Yolk testosterone modulates persistence of neophobic responses in adult zebra finches, *Taeniopygia guttata. Hormones and Behavior, 52*, 640–645.

Tremmel, M., & Müller, C. (2013). Insect personality depends on environmental conditions. *Behavioral Ecology, 24*, 386–392.

Tremmel, M., & Müller, C. (2014). Diet dependent experience and physiological state shape the behavior of a generalist herbivore. *Physiology & Behavior, 129*, 95–103.

Trikalinos, T. A., & Ioannidis, J. P. A. (2006). Assessing the evolution of effect sizes over time. In H. Rothstein, A. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis* (pp. 241–259). Chichester, U.K.: J. Wiley.

Trompf, L., & Brown, C. (2014). Personality affects learning and trade-offs between private and social information in guppies, *Poecilia reticulata. Animal Behaviour, 88*, 99–106.

Valros, A., Pedersen, L. J., Poytakangas, M., & Jensen, M. B. (2017). Evaluating measures of exploratory behaviour in sows around farrowing and during lactation - a pilot study. *Applied Animal Behaviour Science, 194*, 1–6.

Verbeek, M. E. M., Drent, P. J., & Wiepkema, P. R. (1994). Consistent individual differences in early exploratory behaviour of male great tits. *Animal Behaviour, 48*, 1113–1121.

Vernouillet, A., & Kelly, D. M. (2020). Individual exploratory responses are not repeatable across time or context for four species of food-storing corvid. *Scientific Reports, 10*, 394.

Vetter, S. G., Brandstätter, C., Macheiner, M., Suchentrunk, F., Gerritsmann, H., & Bieber, C. (2016). Shy is sometimes better: Personality and juvenile body mass affect adult reproductive success in wild boars, *Sus scrofa. Animal Behaviour, 115*, 193–205.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48.

Vrublevska, J., Krama, T., Rantala, M. J., Mierauskas, P., Freeberg, T. M., & Krams, I. A. (2015). Personality and density affect nest defence and nest survival in the great tit. *Acta Ethologica, 18*, 111–120.

Williams, L. J., King, A. J., & Mettke-Hofmann, C. (2012). Colourful characters: Head colour reflects personality in a social bird, the gouldian finch, *Erythrura gouldiae. Animal Behaviour, 84*, 159–165.

Wilson, A. D. M., & Stevens, E. D. (2005). Consistency in context-specific measures of shyness and boldness in rainbow trout, *Oncorhynchus mykiss. Ethology, 111*, 849–862.

Winter, G. D., Martins, H. R., Trovo, R. A., & Chapman, B. B. (2016). Different behaviour-body length correlations in two populations of juvenile three-spined stickleback (*Gasterosteus aculeatus*). *Behavioural Processes, 122*, 75–79.

Yerkes, R. M., & Yerkes, A. W. (1936). Nature and conditions of avoidance (fear) response in chimpanzee. *Journal of Comparative Psychology, 21*, 53–66.

Yuen, C. H., Pillay, N., Heinrichs, M., Schoepf, I., & Schradin, C. (2015). Personality does not constrain social and behavioural flexibility in African striped mice. *Behavioral Ecology and Sociobiology, 69*, 1237–1249.

Yuen, C. H., Pillay, N., Heinrichs, M., Schoepf, I., & Schradin, C. (2016). Personality traits are consistent when measured in the field and in the laboratory in African striped mice (*Rhabdomys pumilio*). *Behavioral Ecology and Sociobiology, 70*, 1989–1990.

Zidar, J., Balogh, A., Favati, A., Jensen, P., Leimar, O., & Løvlie, H. (2017). A comparison of animal personality and coping styles in the red junglefowl. *Animal Behaviour, 130*, 209–220.

Zidar, J., Balogh, A., Favati, A., Jensen, P., Leimar, O., Sorato, E., & Lovlie, H. (2018). The relationship between learning speed and personality is age- and task-dependent in red junglefowl. *Behavioral Ecology and Sociobiology, 72*, 168.

Zocher, S., Schilling, S., Grzyb, A. N., Adusumilli, V. S., Bogado Lopes, J., Günther, S., Overall, R. W., Winter, Y., & Kempermann, G. (2020). Early-life environmental enrichment generates persistent individualized behavior in mice. *Science Advances, 6*, Article eabb1478.

Appendix

We conducted a literature search in the Web of Science using the following search query:

TS = ((novel object*) OR
    (neophob* AND explorat*) OR
    (neophi* AND explorat*) OR
    ((bold* OR shy*) AND explorat*) OR
    (neophob*) OR
    (neophi*) OR
    (bold* OR shy*) OR
    (explorat* AND object*))
PY = 1990-2020
WC = 'Behavioral Sciences'
Language: (English)
Document types: (Article)

Table A1
Exclusion criteria of title/abstract and full-text screening

Exclusion criterion	Justification	Code name
Not empirical data	Papers do not present original empirical data to be meta-analysed	No empirical study
Studies on human subjects	We were specifically interested in animal studies where methodological approaches are very different from studies on humans	No nonhuman animals
Studies with no novel stimuli	When stimuli are either familiar (already existing in the animal's environment) or irrelevant to the novel object test (e.g. maze), they are not considered novel object trials	No novel object test
Novel object test was conducted to assess other attributes but not any behaviour or personality traits	Several papers assess cognition, but cognitive response might be very different from the behavioural response in which we are interested here	No behavioural study
Novel object trials used to test for differences between treatment groups or clonal lines	If only clonal lines are compared, then the correlation is more within lines than within individuals	No individual-based study
Novel object test was conducted only once per individual	At least two trials of novel objects are needed to calculate the within-individual correlation	No replication
Correlation and/or repeatability measures were not reported	The effect size of interest is needed for meta-analysis	No effect size

Table A2
Descriptive measures of the data set 1990–2020

	No. of effect sizes $N = 265$	No. of studies $N = 115$
Use of different novel objects		
Yes	164	75
No	101	33
Both	–	7
Domestication level and testing context		
Domestic animals tested in artificial environment	40	13
Laboratory-reared animals tested in artificial environment	110	54
Captive wild animals tested in artificial environment	77	29
Wild animals tested in natural environment	38	21
Testing context		
Novel object in neutral position	196	82
Novel object close to/inside nest	15	8
Novel object close to/inside feeder	54	30
Multiple assays		
Yes	215	99
No	50	16

Table A3
Input variables of the meta-analytic model

Variable	Type	Code	Levels	Explanation
Study ID	Random	StudyID	Categorical: 109 studies	Unique identifier for each study
Effect Size ID	Random	EffectSizeID	Categorical: 201 effect sizes	Unique identifier for each effect size
Species	Random	Species	Categorical: 67 species	Unique species names
Sample Size	Fixed	Sample	Numerical	Number of individuals tested repeatedly
Time	Fixed	Time	Numerical	Time interval between two trials
Novelty	Fixed	Novelty	Binary: 0: no; 1: yes	Same or different objects used in repeated trials
Domestication status	Fixed	Domestication	Categorical: 1: domestic animals tested in lab; 2: lab-reared animals tested in lab; 3: wild-caught animals tested in lab; 4: wild animals tested in field	Domestication of species might affect the explorative behaviour of individuals. The place of testing might also has an impact on behavioural responses
GLMM	Fixed	GLMM	Binary: 0: no non-Gaussian linear model; 1: non-Gaussian linear model	Non-Gaussian linear models (e.g. Poisson and binomial models) often lead to lower repeatabilities
Response type	Fixed	Response	Categorical: single (behaviour); composite (of multiple behaviours); average (across multiple trials)	Composite measures and in particular averaged behaviours are expected to yield higher repeatabilities, because measurement error is reduced
Multiple assays	Fixed	MultiAssaysYN	Binary: 0: only novel object; 1: multiple behavioural tests	Whether the novel object test was the only behavioural test of the study might influence the bias towards reporting only statistically significant results
Context	Fixed	Context	Categorical: food; nest, neutral	Whether the novel object was placed next to a food item, next to or inside a nest or in a neutral spot might affect the estimate of individual differences

Table A4
Number of studies corresponding to combinations of terminology and testing contexts

Term	Neutral	Food	Nest	Total (unique)
Neophobia/Neophilia	**23**	**22**	**4**	**48**
Neophobia	13	22	3	
Neophilia	10	0	0	
Boldness−shyness	**26**	**5**	**1**	**32**
Boldness	25	5	1	
Shyness	1	0	0	
Exploration	**19**	**2**	**2**	**23**
Other terms	**10**	**0**	**2**	**12**
Activity	1	0	0	
Approach−avoidance	2	0	2	
Risk responsiveness	2	0	0	
Fearfulness	5	0	0	
No specific terms	**9**	**0**	**0**	**9**
Total (unique studies)	**82**	**30**	**9**	

Table A5
Number of studies per terminology used and moderators tested

	Neophobia/Neophilia	Boldness−shyness	Exploration
Clade			
Birds	32	7	9
Mammals	2	13	5
Fish	1	5	2
Insects	0	4	0
Reptiles	2	0	0
Context			
Food	19	4	1
Nest	3	1	1
Neutral	17	24	14
Novelty			
Different objects	32	21	10
Same object	6	9	7
Domestication status			
Domesticated	5	2	2
Laboratory-reared	10	20	5
Wild-caught (captive)	18	3	5
Wild	5	4	4

Table A6
Publications included in the final data set

Study ID	Authors (Year)	Species	Sample Size	Novel object item	Behaviour measured	Behavioural term	metaDigitize
1	Amy et al. (2017)	*Serinus canaria*	48	Red plastic round key ring, one blue oblong keyring and two gold and green metallic Eiffel towers	Latency to feed	Neophobia	No
2	An, Kriengwatana, Newman, MacDougall-Shackleton, and MacDougall-Shackleton (2011)	*Poecile atricapillus*	21	Deflated green balloon, plastic pink kazoo, fruit loops	Latency to approach	Neophobia	No
3	Arnold et al. (2016)	*Cyanistes caeruleus*	69	Pink plastic frog, half purple rubber ball	Latency to approach	Neophobia	No
4	Basic, Winberg, Schjolden, Krogdahl, and Hoglund (2012)	*Oncorhynchus mykiss*	18	Yellow rubber stopper	Latency to approach, locomotor activity, number of approaches	Behavioural response	No
5	Baxter-Gilbert, Riley, and Whiting (2019)	*Intellagama lesueurii*	83–228	Paper coffee cups, aluminium pie tins, water bottle, bag of potato chips, soft drink can	Latency to approach	Neophilia	No
6	Bibi et al. (2019)	*P. major*	24	Penlight battery, pink ball	Latency to feed	Neophobia	No
7	Boogert, Reader, and Laland (2006)	*Sturnus vulgaris*	15	Five coloured clothes-pegs, styrofoam on a cardboard plate, yellow reflective material, white opaque tube cap, white spool of light purple elastic wire, bright green opaque tube cap	Latency to feed	Object neophobia	No
8	Brust and Guenther (2017)	*C. aperea*	22	Green eggcup, yellow plastic duck	Latency to contact	Boldness	No
9	Burns (2008)	*P. reticulata*	11–36	Plastic cylinder tower with black and white stripes, purple plastic block	Latency to approach	Neophobia	No
10	Carere, Drent, Privitera, Koolhaas, and Groothuis (2005)	*P. major*	19	Penlight battery, pink rubber toy	Latency to approach	Exploration	No
11	Christensen et al. (2020)	*Equus caballus*	25	Four plastic boxes	Vigilance	Alertness	
12	Collins, Hatch, Elliott, and Jacobs (2019)	*Rissa tridactyla*	42	Ball of tissue and duct tape attached to thick wire	Latency to approach	Boldness	No
13	Coutant, Bagur, and Gilbert (2018)	*Amazona aestiva*	18–31	Plastic key toys	Behaviour scores	Neophilia	No
14	Damas-Moreira et al. (2019)	*Podarcis sicula* *Podarcis virescens*	26 29	White nonperfumed candles in foil, yellow cupcake paper, blue plastic clothes-peg	Latency to approach	Neophobia	Yes
15	Dammhahn and Almeling (2012)	*Microcebus murinus*	29–30	White wooden box with white mesh lid	Behaviour scores	Boldness	No
16	Dardenne et al. (2013)	*Hirundo rustica*	77	Christmas decorations	Latency to approach	Neophobia	No
17	David et al. (2011)	*T. guttata*	42	Small bag, soccer figurine	Latency to feed	Neophobia	No
18	DeRango et al. (2019)	*Zalophus wollebaeki*	14–33	Red, blue and yellow balls	Behaviour scores	Boldness	No
19	Devost et al. (2016)	*P. atricapillus*	78	Pink cardboard box	Latency to approach	Neophilia	No
20	Edwards et al. (2017)	*Acrocephalus sechellensis*	177	Pink toy	Behaviour scores	Exploration	No
21	Edwards et al. (2018)	*A. sechellensis*	185	Pink toy	Number of approaches	Exploration	No
22	Ensminger and Westneat (2012)	*Passer domesticus*	27	Blue ceramic vase, black, white and orange sports action figure	Latency to feed	Neophobia	No
23	Exnerova et al. (2015)	*P. major*	50	Bright blue pen, pink plastic clothes-peg	Behaviour scores	Exploration	No
24	Farrell et al. (2012)	*S. vulgaris*	40	Food dish, deflated blue balloon, corn chips	Latency to feed	Neophobia	No
25	Feenders and Bateson	*S. vulgaris*	31	Green or red light on right or left key	Latency to contact	Neophobia	No
26	Finkemeier et al. (2016)	*C. aperea*	63	Green plastic eggcup	Number of approaches	Boldness	No
27	Fox and Millam (2010)	*Nymphicus hollandicus*	45	Plastic chains, small swings, bells, mirrors, coloured wood	Behaviour scores	Behavioural response	No
28	Frost et al. (2013)	*O. mykiss*	12	Lego duplo blocks	Latency to approach	Boldness	No
29	Funghi et al. (2015)	*Estrilda astrild*	42	Green clothes-peg, pink marker	Latency to approach	Neophobia	No
30	Gabriel and Black (2010)	*Cyanocitta stelleri*	29	Feeding apparatus	Behaviour scores	Exploration	No

Table A6 (*continued*)

Study ID	Authors (Year)	Species	Sample Size	Novel object item	Behaviour measured	Behavioural term	metaDigitize
31	Garamszegi et al. (2015)	*Ficedula albicollis*	16—27	Paper sheet with small random drawings in variable colours	Latency to approach	Approach—avoidance	No
32	Garamszegi et al. (2012)	*F. albicollis*	52	White paper sheet	Latency to contact	Approach—avoidance	No
33	Grace and Anderson (2014)	*Sula granti*	86—157	Red Bull can, plastic crate	Behaviour scores	Behavioural response	Yes
34	Greenberg and Holekamp (2017)	*Crocuta crocuta*	14	Cooler, funnel, stool, bucket	Latency to contact	Neophobia & exploration	No
35	Greggor et al. (2016)	*Corvus frugilegus*	16—17	Paper and plastic customized	Latency to contact	Neophobia	No
36	Greggor et al. (2020)	*Corvus hawaiiensis*	102	Paper and plastic customized	Latency to feed	Neophobia	No
37	Grindstaff, Hunsaker, and Cox (2012)	*T. guttata*	109	AA battery, blue plastic frog	Latency to contact	Neophobia	No
38	Guenther and Brust (2017)	*C. aperea*	24	Plastic duck, Lego bricks	Latency to contact	Boldness	No
39	Guenther and Trillmich (2013)	*C. aperea*	50	Yellow plastic cup, green eggcup, red plastic pig	Latency to contact	Boldness	No
40	Guenther, Brust et al. (2014)	*C. aperea*	21	Yellow plastic cup, green eggcup	Latency to contact	Boldness	No
41	Guenther, Finkemeier et al. (2014)	*C. aperea*	62	Green eggcup, yellow plastic duck	Latency to contact	Boldness	No
42	Guenther et al. (2018)	*C. aperea*	30	Yellow plastic duck, grey plastic cylinder, green eggcup	Number of approaches	Boldness	No
43	Guido et al. (2017)	*Milvago chimango*	9	Feeding apparatus	Latency to feed	Neophobia	No
44	Gyuris et al. (2012)	*Pyrrhocoris apterus*	40	Plastic plugs	Locomotor activity	Exploration	No
45	Haage et al. (2013)	*Mustela lutreola*	68—80	Dog toys	Latency to approach	Boldness	No
46	Hebert et al. (2014)	*Betta splendens*	25	Not mentioned	Latency to feed offspring, Latency to approach, Locomotor activity	Boldness	No
47	Herborn et al. (2010)	*C. caeruleus*	43—125	Pink plastic frog, half purple rubber ball	Latency to feed, latency to approach	Neophobia	No
48	Hirata and Arimoto (2018)	*Bos taurus*	25	Plastic lanterns, frog swim ring, hand fans, elephant watering cans, plastic containers, plastic baskets, fan blades	Latency to approach	Boldness	No
49	Hopkins and Bennett (1994)	*Pan troglodytes*	49	Paintbrush, plastic cooler, dolly with three wheels, plastic board with holes, metal battery clamp, PVC configuration, metal paint roller, garage door spring, plastic pegboard, plastic seriation buckets	Latency to approach	Approach—avoidance	No
50	Jäger et al. (2017)	*Rhabdomys pumilio*	73	Plastic toys, table tennis ball	Latency to contact	Exploration	No
51	Janczak et al. (2003)	*Sus domesticus*	88	Bucket	Latency to approach, latency to contact, number of approaches, locomotor activity	Fearfulness	No
52	Johnson et al. (2015)	*Papio anubis*	43	Truck, plastic bear	latency to contact	Response to novelty	No
53	Jolles et al. (2013)	*C. frugilegus*	19	Not mentioned	Latency to approach	Boldness	No
54	Jolly et al. (2019)	*Melomys burtoni*	30	Plastic bowl	Latency to move	Boldness	No
55	Kerman et al. (2018)	*T. guttata*	30	Toys neutral colours	Latency to feed	Boldness	No
56	Krams et al. (2014)	*Ficedula hypoleuca*	40	Purple tennis ball	Latency to approach	Neophobia & exploration	No
57	Krause et al. (2017)	*T. guttata*	147	Blue AA battery	Latency to feed, number of approaches	Neophobia & exploration	No
58	Krebs et al. (2019)	*Mus domesticus*	30	Lego toy	Latency to contact, locomotor activity, latency to approach	Behavioural response	No
59	Kurvers et al. (2009)	*Branta leucopsis*	18	Green plastic mat, brown deep-pile rug	Latency to approach	Boldness	No
60	Kurvers et al. (2012)	*B. leucopsis*	44	Green plastic mat, brown deep-pile rug	Latency to approach	Boldness	No
61	Le Vin et al. (2011)	*Neolamprologus pulcher*	28	Purple plastic half sphere, red Buddha figurine	Latency to approach	Risk responsiveness	No

Table A6 (*continued*)

Study ID	Authors (Year)	Species	Sample Size	Novel object item	Behaviour measured	Behavioural term	metaDigitize
62	Lermite et al. (2017)	*Acridotheres tristis*	49	Green plastic hairbrush, pink doorstop	Latency to feed, latency to approach	Neophobia neophilia	No
63	Malmkvist and Hansen (2002)	*Neovison vison*	187	Wooden cube	Latency to approach	Fearfulness	No
64	Martin-Wintle et al. (2017)	*Ailuropoda melanoleuca*	18	Ice blocks, ice blocks with apple and carrot, tubs of water with two apple halves and one carrot, rubber ball	Behaviour scores	Neophobia	No
65	Mazza et al. (2019)	*Myodes glareolus*	86	Plastic toy horse, plastic toy duck	Latency to approach	Boldness	No
66	Mazza et al. (2018)	*M. glareolus*	86	Plastic toy horse, plastic toy duck	Latency to approach	Boldness	No
67	McCune et al. (2018)	*Aphelocoma californica*	16–18	Yellow plastic duck, rock	Latency to approach	Boldness	No
68	Meagher et al. (2016)	*B. taurus*	11–24	Brightly coloured ball	Latency to approach	Fearfulness	No
69	Medina-Garcia et al. (2017)	*Melopsittacus undulatus*	32	Yellow plastic cup, foraging device, small plastic statue of red dragon	Latency to contact	Neophobia	Yes
70	Meehan and Mench (2002)	*Amazona amazonica*	16	Stuffed toy chicken, miniature artificial pine tree, large plastic funnel, string of seashells, bunch of hot pink feathers, plastic action figure, red silk flower, child's sandal, screwdriver, bunch of measuring spoons, rubber duck, small woven basket	Latency to contact	Behavioural response	No
71	Mettke-Hofmann (2012)	*Erythrura gouldiae*	24	White cotton mop, brown cardboard tube with holes	Latency to feed, latency to approach	Neophobia neophilia	No
72	Mettke-Hofmann et al. (2005)	*Sylvia melanocephala* *Sylvia borin*	11 9	Cotton mop, tube with holes	Behaviour scores	Neophobia	No
73	Michelena et al. (2009)	*Ovis aries*	40	Plastic boxes with perforated lids containing different fresh herbs, coffee powder, baby's rattle, bottle brush, various baby teething rings	Locomotor activity	Exploration	No
74	Miller et al. (2005)	*Coturnix coturnix japonica*	35	Small green courgette, bright yellow plastic cup, two brown pinecones	Latency to contact	Fearfulness	No
75	Miller et al. (2006)	*C. c. japonica*	48	Small green courgette, bright yellow plastic cup, two brown pinecones	Latency to contact, latency to approach	Fearfulness	No
76	Moldoff and Westneat (2017)	*P. domesticus*	36	Blue plastic cup, clear glass jar	Latency to approach	Neophobia	No
77	Monestier et al. (2017)	*Capreolus capreolus*	21	10 geometric polystyrene shapes (circle, diamond, square, triangle) painted with contrasting colours	Number of approaches	Neophobia	No
78	Morinay et al. (2019)	*F. albicollis*	65	Coloured figurine	Latency to approach	Neophobia	No
79	Noer et al. (2016)	*N. vison*	60	Green cone-shaped dog toy, red circular dog toy	Behaviour scores	Shyness	No
80	Overington et al. (2011)	*Quiscalus lugubris*	36	Bright orange rubber ball with three nails, multicoloured ball with spikes attached to a black film canister	Latency to feed	Neophobia	No
81	Pedersen (1994)	*Vulpes vulpes*	16	Orange rubber glove mounted on a stick	Behaviour scores	Behavioural response	No
82	Perals et al. (2017)	*A. tristis*	58	Blue, green and red Lego pieces	Latency to contact	Neophobia	No
83	Pogány et al. (2018)	*T. guttata*	59	Small unpainted metal flag, small flag painted black and yellow stripes	Latency to feed	Neophobia	No
84	Rangassamy et al. (2016)	*Mus spicilegus*	37	Artificial hamburger plasticized PVC, kidney-shaped metallic box	Latency to contact	Neophobia	No
85	Rockwell et al. (2012)	*C. stelleri*	63–57	Wooden platform	Locomotor activity	Neophobia	No
86	Rohrer and Ferkin (2020)	*Microtus pennsylvanicus*	42	Oat cereal ring on a hook, glass slide smeared with white clover pulp	Latency to approach, latency to move, number of approaches, locomotor activity	Boldness Activity	No

Table A6 (*continued*)

| Study ID	Authors (Year)	Species	Sample Size	Novel object item	Behaviour measured	Behavioural term	metaDigitize
87	Ruuskanen and Laaksonen (2010)	*F. hypoleuca*	57	Pink and gold toy rubber duck	Latency to approach	Neophobia	No
88	Schielzeth et al. (2010)	*T. guttata*	119–530	Dry perennial herbs, half an apple, toy ball	Behaviour scores	Neophilia	No
89	Schürch and Heg (2010)	*N. pulcher*	15–21	Model bird, beetle kitchen magnet, screwdriver, blue clamp	Latency to approach	Boldness	No
90	Siviter et al. (2017)	*Pogona vitticeps*	13	Plastic toy police van, china garden fairy, pottery wellington boot, blue plastic lamp	Latency to approach	Behavioural response	No
91	Smith and Blumstein (2012)	*Poecilia reticulata*	37	Pink soap dish, plastic baseball	Number of approaches	Exploration	No
92	Soha et al. (2019)	*Melospiza melodia*	18	Foraging grid	Latency to feed	Neophobia	No
93	Sol et al. (2012)	*A. tristis*	60	Yellow tape, green hairbrush	Latency to feed	Neophobia	No
94	Spake et al. (2012)	*S. domesticus*	567	Bucket, tennis shoe	Latency to contact	Exploration	No
95	Stöwe, Bugnyar, Heinrich et al. (2006) (A)	*Corvus corax*	6–11	Bottles, boxes, bags, cans, candles, cups	Behaviour scores, number of approaches, latency to approach	Approach -avoidance	No
96	Stöwe, Bugnyar, Loretto et al. (2006) (B)	*C. corax*	11–12	Bottles, boxes, bags, cans, candles, cups	Behaviour scores	Neophobia	No
97	Stuber et al. (2013)	*P. major*	49–61	Miniature video camera	Behaviour scores	Exploration	No
98	Tan and Tan (2019)	*Phaneroptera brevis*	48	Dog food pellet	Latency to approach	Boldness & exploration	No
99	Thodberg et al. (1999)	*S. domesticus*	26	Bucket	Latency to contact	Object exploration	
100	Tobler and Sandell (2007)	*T. guttata*	44	Iceberg lettuce, painted toy plastic dinosaurs with black eyespots	Latency to approach	Neophobia	No
101	Tremmel and Müller (2013)	*Phaedon cochleariae*	48	Red rubber plugs	Locomotor activity	Boldness	No
102	Tremmel and Müller (2014)	*Galeruca tanaceti*	51	Red rubber plugs	Locomotor activity	Boldness	No
103	Trompf and Brown (2014)	*P. reticulata*	79	Pink or yellow plastic peg, pink, blue, green or yellow pieces of Lego	Latency to approach	Boldness	No
104	Valros et al. (2017)	*Sus scrofa*	10	White plastic flowerpots, plastic cups, plastic spaghetti spoon	Latency to contact	Neophilia	No
105	Verbeek et al. (1994)	*P. major*	17–46	Penlight battery, pink panther toy	Latency to contact	Exploration	No
106	Vernouillet and Kelly (2020)	*Gymnorhinus cyanocephalus* *Nucifraga columbiana*	11 12	Red cup, green poker chip, black bottle, yellow plastic duck	Latency to approach, latency to move	Exploration	No
107	Vetter et al. (2016)	*S. scrofa*	57	Bucket, football, booster eat, toy tube, plastic basket, traffic cone, plastic crow, potato bag, watering can	Latency to contact	Exploration	No
108	Vrublevska et al. (2015)	*P. major*	21	Purple tennis ball, yellow tennis ball	Latency to feed	Neophobia	No
109	Williams et al. (2012)	*E. gouldiae*	18	Blue cork, white half cork	Latency to approach, latency to feed	Neophilia neophobia	No
110	Wilson and Stevens (2005)	*O. mykiss*	5	Feeding apparatus	latency to feed	Boldness	Yes
111	Winter et al. (2016)	*Gasterosteus aculeatus*	58	Dark plastic rain gutter	Locomotor activity	Exploration	No
112	Yuen et al. (2015)	*R. pumilio*	37	Plastic toy, white table tennis ball	Latency to contact object	Exploration	No
113	Yuen et al. (2016)	*R. pumilio*	18	Animal toy	Behaviour scores	Exploration	Yes
114	Zidar et al. (2017)	*Gallus gallus*	87–100	Spherical brown beige plush toy with large yellow and black eyes	Escape	Boldness	No
115	Zidar et al. (2018)	*G. gallus*	87	Spherical brown and beige plush toy with large yellow and black eyes	Vigilance	Neophilia	No
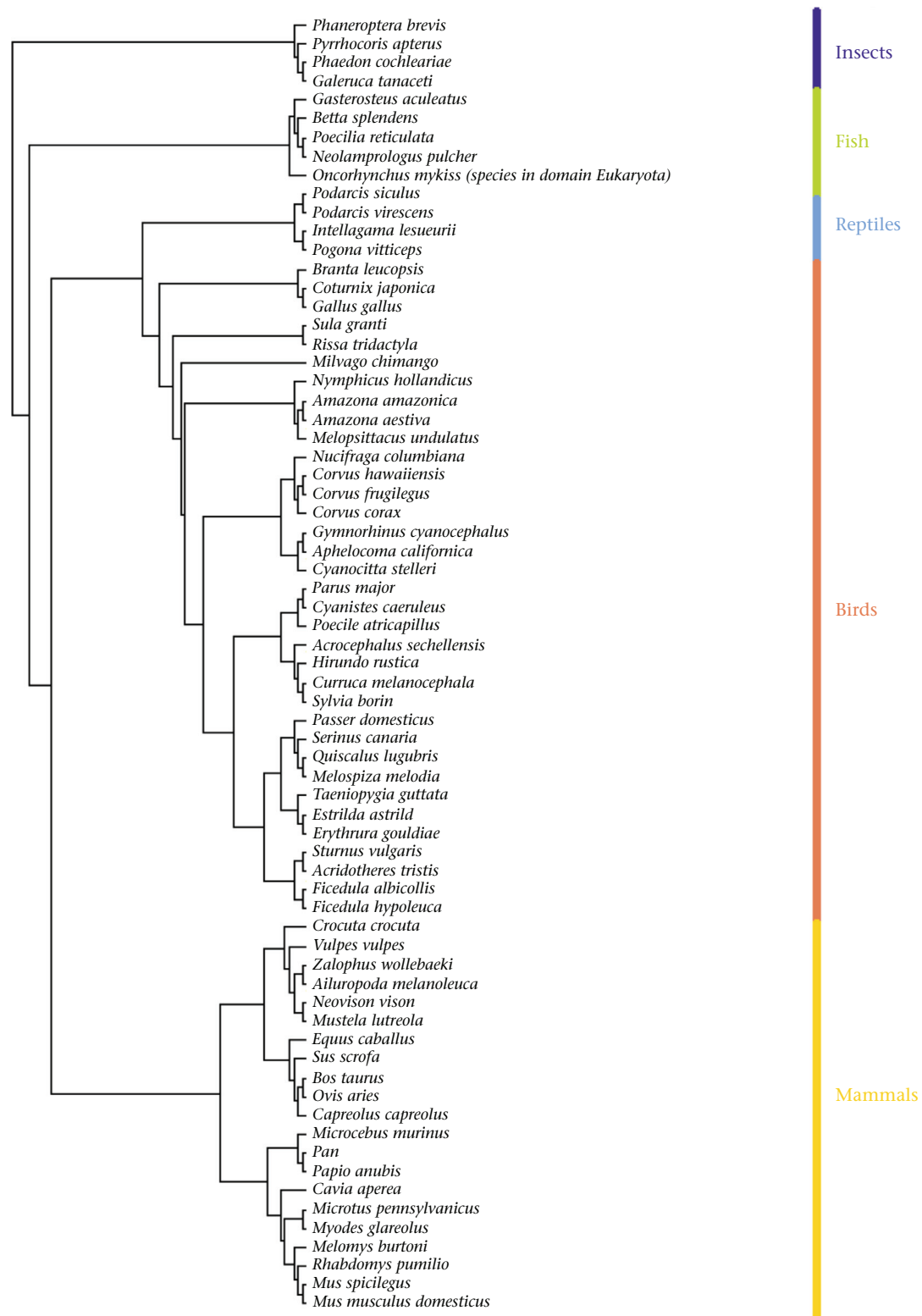
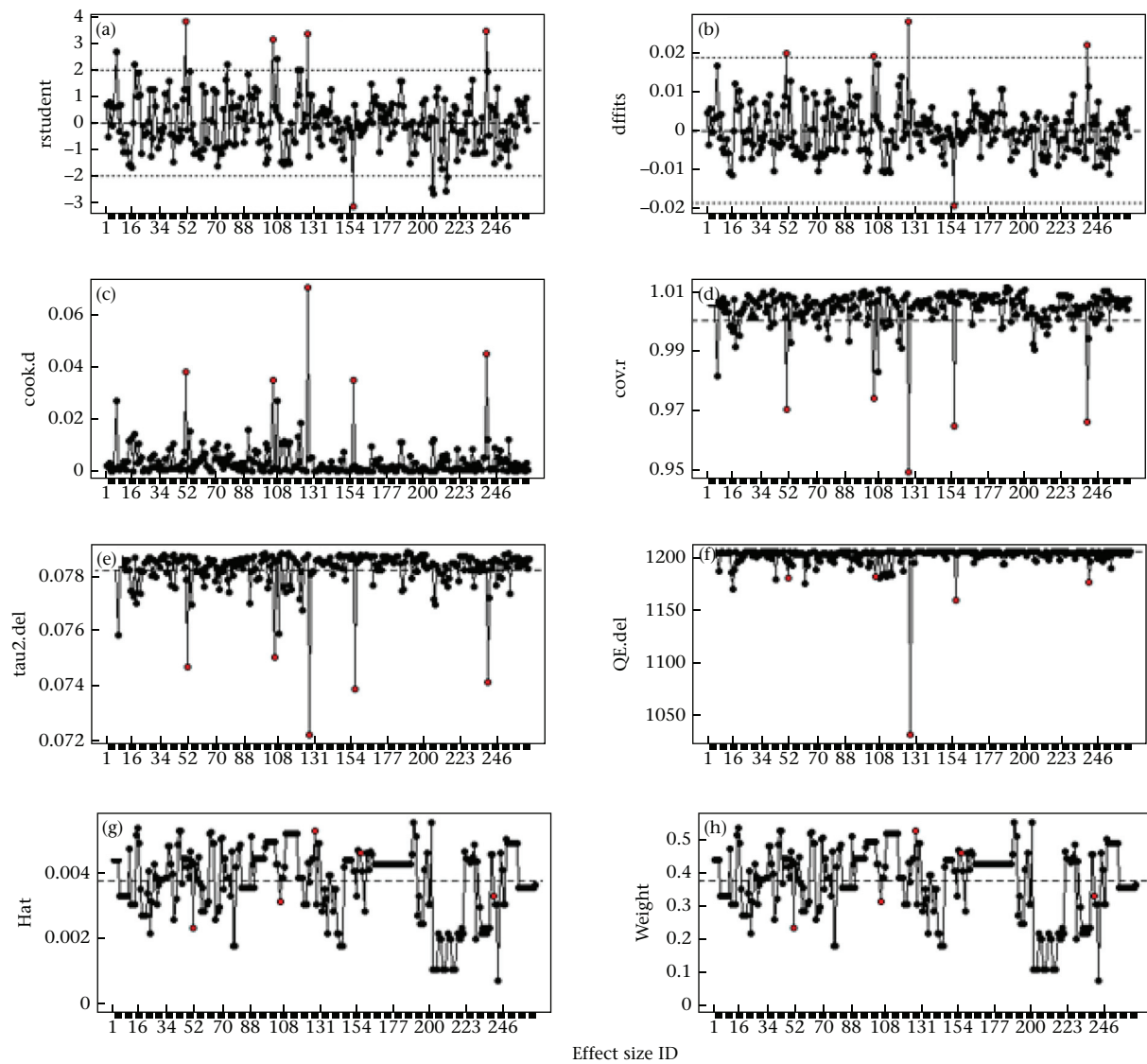**Figure A1.** Phylogenetic tree of species included in the meta-analysis.

Figure A2. Influence diagnostics of effect sizes. Red points indicate outliers. *X* axes show effect size ID. *Y* axes show the scores of the following leave-one-out diagnostic tests: (a) rstudent, (b) DFFITS value, (c) Cook's distance, (d) covariance ratio, (e) the leave-one-out amount of (residual) heterogeneity, (f) the leave-one-out test statistic for the test of (residual) heterogeneity, (g) hat values, (h) weights. Dashed lines represent the overall value of each test.
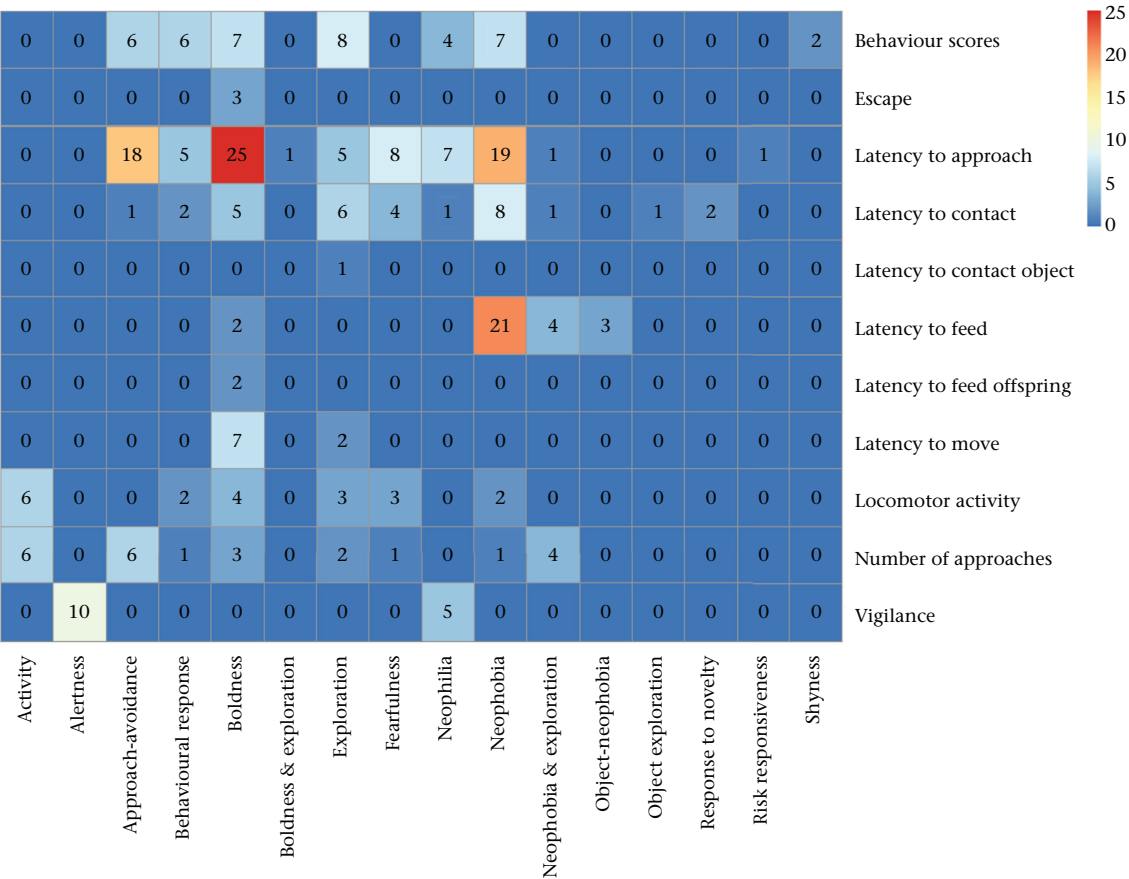
Figure A3. Heatmap showing the number of effect sizes by response behaviour and label for the behavioural phenotypes.
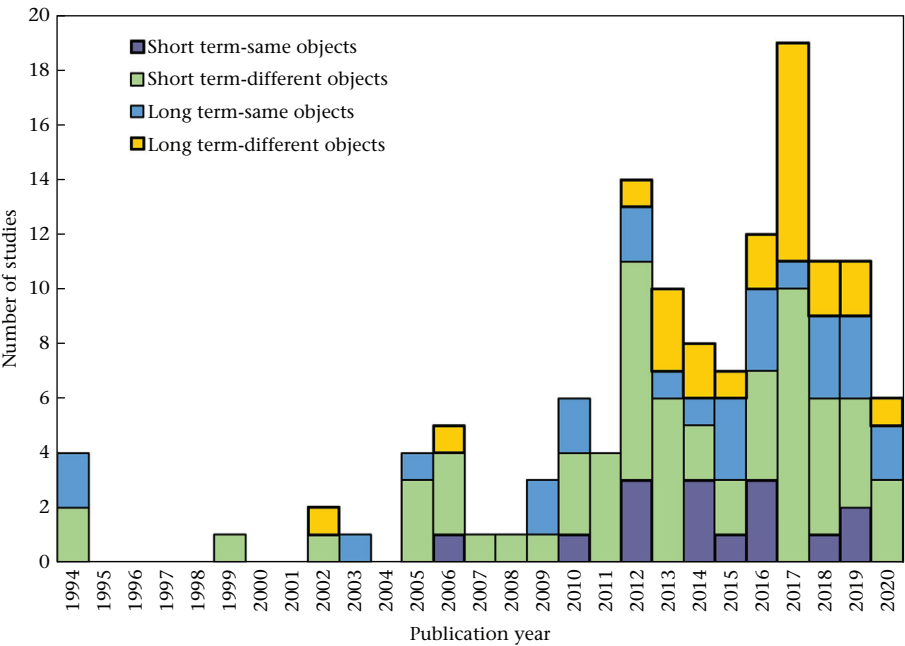

Figure A4. Number of studies per year. Different colours represent the use of same or different objects during repeated trials of the novel object test.