Behavioral Ecology

The official journal of the
ISBE
International Society for Behavioral Ecology

Behavioral Ecology (2015), 26(2), 311–319. doi:10.1093/beheco/aru125

Invited Review

The effect of experimental design on the measurement of mate choice: a meta-analysis

Liam R. Dougherty,[a] and David M. Shuker[a]

[a]Centre for Biological Diversity, School of Biology, University of St Andrews, Harold Mitchell Building, St Andrews KY14 7AU, UK

Received 14 February 2014; revised 17 June 2014; accepted 19 June 2014; Advance Access publication 29 August 2014.

Quantifying the shape and strength of mating preferences is a vital component of the study of sexual selection and reproductive isolation, but the influence of experimental design on these estimates is unclear. Mating preferences may be tested using either no-choice or choice designs, and these tests may result in different estimates of preference strength. However, previous studies testing for this difference have given mixed results. To quantify the difference in the strength of mating preferences obtained using the 2 designs, we performed a meta-analysis of 38 studies on 40 species in which both experimental designs were used to test for preferences in a single species/trait/sex combination. We found that mating preferences were significantly stronger when tested using a choice design compared with a no-choice design. We suggest that this difference is due to the increased cost of rejecting partners in no-choice tests; if individuals perceive they are unlikely to remate in a no-choice situation they will be more likely to mate randomly. Importantly the use of choice tests in species in which mates are primarily encountered sequentially in the wild may lead to mating preferences being significantly overestimated. Furthermore, this pattern was seen for female mate choice but not for male mate choice, and for intraspecific choice but not for interspecies or interpopulation mate discrimination. Our study thus highlights the fact that the strength of mating preferences, and thus sexual selection, can vary significantly between experimental designs and across different social and ecological contexts.

Key words: **choice test, experimental design, interspecific choice, mate choice, mating preferences, meta-analysis.**

INTRODUCTION

Sexual selection arises via 2 main mechanisms: intersexual mate choice and intrasexual contest competition (Darwin 1871; Andersson 1994). As such, understanding mate choice and the underlying preferences that lead to choice are central to understanding the scope and action of sexual selection. Mate choice is an outcome and can be defined broadly as arising when a trait in one sex leads to nonrandom mating success in the other sex (Halliday 1983; Shuker 2010). Therefore, choice outcomes are influenced by underlying mating preferences (the sensory and behavioral properties that influence the propensity of individuals to mate with certain phenotypes; Jennions and Petrie 1997) and other factors that affect the expression of these preferences (Jennions and Petrie 1997; Wagner 1998). These include the degree of mate sampling (Janetos 1980; Gibson and Langen 1996; Wagner 1998), the condition of the choosing individual (Cotton et al. 2006; Beckers and Wagner 2013), and the costs and benefits associated with choice (e.g., Milinski and Bakker 1992). Measurement of mate choice may thus be influenced by experimental design, if different designs vary in any of these factors (Wagner 1998).

An important way in which experiments testing mate preferences can vary is in the number of options the subject is presented with during the test, which we refer to as the "choice paradigm" or "choice design." Tests can use either no-choice or choice designs (Wagner 1998). In a no-choice test each subject is presented with a single stimulus. Several no-choice trials may be performed using the same subject; these are referred to as sequential choice tests. In contrast, in a choice test each subject is given a choice between multiple (usually 2) stimuli presented simultaneously. The 2 paradigms differ most importantly in whether options can be directly compared or not. Because comparison is possible, choice tests detect relative, directional preferences between stimuli (Wagner 1998; MacLaren and Rowland 2006). As such, these tests may allow greater resolving power between options as even small differences in trait values may lead to large differences in choice outcomes (Doherty 1985; Wagner 1998). However, this effect may amplify the strength of preferences observed if a dichotomous yes or no response is recorded (Wagner et al. 1995; Wagner 1998). In contrast, no-choice experiments test for absolute preferences as no direct comparison is possible (Wagner 1998). No-choice tests also differ from choice tests in that the perceived mate encounter rate is lower: if a mate is rejected in a no-choice tests there may be no guarantee of a mating opportunity in the future (Werner and Lotem 2006; Barry and Kokko 2010; Booksmythe et al. 2011).

Address correspondence to L.R. Dougherty. E-mail: lrd5@st-andrews.ac.uk.

Thus, rejection of an option in a no-choice test may indicate a stronger or more robust preference than that seen in a choice test, because the subject has foregone mating despite this extra "cost of rejection." It seems likely that one or all of these factors may lead to differences in the strength of preferences observed in each paradigm.

There are many cases of both no-choice and choice paradigms being used to test for mating preferences in the same species in different studies. For example, male Pacific Blue-eye fish *Pseudomugil signifer* prefer larger females in both simultaneous (Wong and Jennions 2003) and sequential choice tests (Wong et al. 2004). Similarly female cockroaches *Nauphoeta cinerea* also prefer dominant males in both no-choice (Moore and Moore 1988) and choice tests (Moore and Breed 1986). A potentially more powerful comparison of paradigms is one in which preferences are tested on the same species in a single study. If experiments are carried out by the same experimenters in a similar way, this may potentially reduce the number of confounding variables that could lead to differences in observed preferences between tests. Several studies have found stronger mating preferences in choice tests compared with no-choice tests in this way (MacLaren and Rowland 2006; Barry et al. 2010; Booksmythe et al. 2011; Owen et al. 2012). However, several studies have also shown little effect of choice paradigm on the strength of mating preference (e.g., Gabor et al. 2000; Jang and Gerhardt 2006; Gershman and Sakaluk 2009; Jordan and Brooks 2012).

Here, we present a meta-analysis in which we quantify the effect of choice paradigm on the measurement of mate choice. We searched the literature for studies in which mating preferences were tested using both a no-choice and a choice paradigm, on the same species/trait/sex combination. Including the results of 2 experiments from the same study should reduce confounding factors such as effects associated with individual researchers, animal stocks, and so forth. The effect size used in the analysis can be most simply considered as the degree of nonrandom response with respect to a partner's trait presumed to be the target of mate choice. We included studies presenting both mate choice outcomes and also proxy measures of mating preference (see below). For clarity we refer to the mean effect sizes derived from our analysis as the "strength of preference" throughout. We included studies considering both male and female choice, as well as intraspecies, interpopulation, and interspecies choice (see Materials and Methods). We predict that, for the reasons mentioned above, mating preferences will be significantly stronger for choice tests compared with no-choice tests. We also predict that overall female choice will be stronger than male choice, as females generally invest more in each reproductive event and so should be more discriminating in their choice of mate (Andersson 1994). We also predict that interspecies choice will be stronger than intraspecies and interpopulation choice, as there are higher costs associated with making the wrong choice when choosing between a conspecific and a heterospecific individual (Andersson 1994).

MATERIALS AND METHODS

In presenting the methods we have attempted to follow as close as possible the PRISMA standards for reporting meta-analyses (Moher et al. 2009; see Nakagawa & Poulin 2012; see Figure 1 for diagram showing search results and the study selection process).

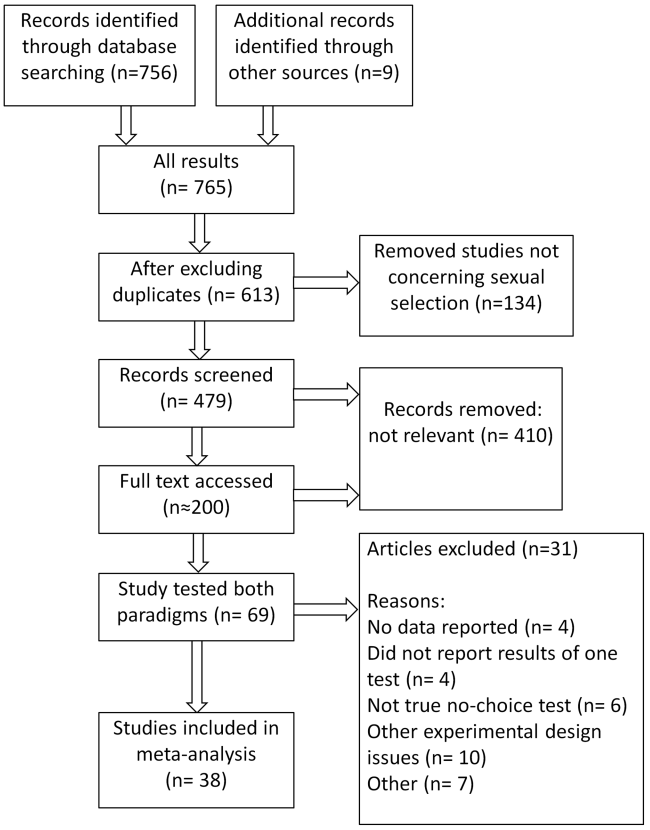

Figure 1
PRISMA flow chart of search results and the study selection process. See Supplementary Table S1 for list of papers excluded from the analysis.

Search protocol

We used 3 approaches to search the literature. First, after initial scoping searches in September and October 2012, we performed keyword searches of several online databases in June 2013. We took the first 100 results from the databases Google Scholar (Google) and Scirus (Elsevier) for the search terms "*sequential simultaneous mate choice*," on 17th June 2013. On 19th June we performed the following searches in both Web of Knowledge (Thomson Reuters) (in the TOPIC field) and Scopus (Elsevier) (in the "Article Title, Abstract, Keywords" field): "*no choice*" AND "*multiple choice*"; "*no choice*" AND "*two choice*"; "*no choice*" AND "*simultaneous*"; "*sequential*" AND "*simultaneous*"; "*sexual* isolat**" AND "*no choice*" AND "*multiple choice*". The number of results obtained for each search can be found in the Supplementary Material (and full endnote libraries are available on request).

Secondly, we used Web of Knowledge to search all studies citing 4 papers identified as being influential in this area: the review by Wagner (1998) on measuring mating preferences and experimental design; the highly cited study by Rowland (1982) on male choice in *Gasterosteus aculeatus*; and finally 2 more recent papers which explicitly tested for the effect of experimental design on mate preferences (Coyne et al. 2005; MacLaren and Rowland 2006). After our online searches, we then inspected the titles and abstracts of the results in order to remove papers that were obviously not relevant to our search. Papers that were deemed relevant were then read in detail in order to see whether the study could be included (see inclusion criteria below). Finally, we also followed papers cited in the text if our searches had not already located them.

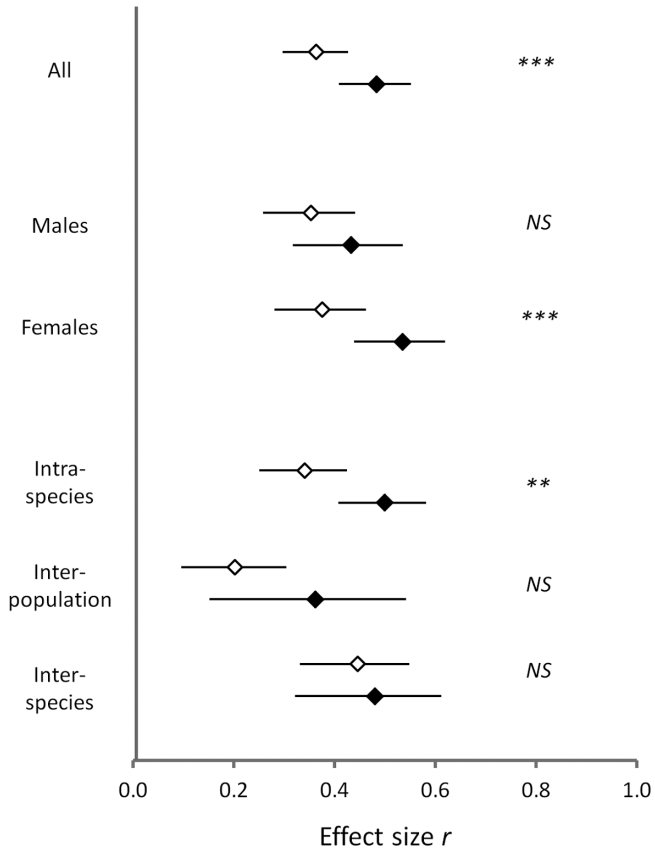

Figure 2

Mean strength of mating preferences (correlation coefficient r) for the 2 choice designs (white diamonds for no-choice tests and black diamonds for choice tests) for all effect sizes and split by sex and trait type. Bars show the bootstrapped 95% CIs around the mean effect size estimates derived from the meta-analytic models. See Table 1 for sample sizes (number of effect sizes) associated with each subgroup. See Materials and Methods for classification of trait types. P values show the results of a weighted least-squares regression testing the effect of choice paradigm on mean effect size for each subgroup of studies (see Results, $^{***}P < 0.0001$, $^{**}P < 0.001$). Statistical analyses were performed using Fisher's z transform of the correlation coefficient ($\tilde{z}r$), and then converted back to r for presentation.

Criteria for inclusion

We had several criteria for inclusion. Most importantly, each study needed to include at least one effect size corresponding to a no-choice test and one effect size corresponding to a choice test (for most studies multiple effect sizes were presented, see below). We included only studies in which each test was performed using the same species and sex, testing for a preference for the same trait. This is important as we found several cases where both no-choice and choice designs were performed but different traits were considered between tests (see Figure 1 for the most common reasons for excluding papers from our analysis, and Supplementary Table S1 for more detailed information). Importantly, our analysis includes measures of mate choice in the form of successful matings and also in the form of proxy behavioral measures (such as association time or courtship effort).

Both tests did not have to be performed using identical stimuli (indeed in most cases this would not be possible because individuals of the opposite sex were used as stimuli); however, stimuli did need to be comparable. One example of an excluded study should help to illustrate this point. Basolo (1995) tested for a female preference for males with (artificial) swords in the unsworded Platyfish *Priapella olmecae*. First the presence of a preference was tested using a choice test, in which a female chose between a normal, unsworded male and a male to which an artificial sword had been experimentally added. Second, no-choice tests were used to test for female preference for swords of differing sizes. However, there was no corresponding no-choice test using an unsworded male. Therefore, the choice design tests for a preference for swords, whereas the no-choice design tests for a preference for sword size. Therefore, we did not include this study in the analysis, as the stimuli used in each test were not directly comparable.

We define a no-choice test as one in which a subject is presented with a single stimulus or potential mate. This excludes designs commonly used in sexual isolation studies in which subjects are presented with several potential mates of a single type (e.g., Tomaru and Oguma 2000). This definition also includes sequential choice tests, in which several no-choice tests are performed concurrently using different stimuli. We define a choice test as one in which a subject is presented with more than one stimulus simultaneously. Most studies use a 2-choice test, but we also included those in which more than 2 options were given (e.g., 3-choice test: Beckers and Wagner 2011).

We included all stated measures of mate preference and rely on the authors' judgments on whether the measured behaviors accurately reflect mating preferences or not. We did not impose any limitations on the degree of randomization regarding the order of presentation of stimuli, or whether presented stimuli were controlled (e.g., synthetic calls) or not. We also did not impose limitations regarding whether the same individuals were used in both no-choice and choice tests, or whether the same stimuli were presented to all individuals. We included studies that tested both male and female mate choice. We also included studies considering both intraspecific traits ("intraspecies choice") as well as interspecific mate choice; that is, choice between a conspecific and a heterospecific individual ("interspecies choice"). We also included studies considering choice between different intraspecific populations and strains (due to different larval host plants), which we classified as "interpopulation choice." We refer to these 3 categories as "trait types."

Finally, we excluded studies in which we were unable to extract appropriate effect sizes (e.g., missing test statistics or sample sizes; Figure 1). For one study (Owen et al. 2012), we were provided with statistics not presented in the original paper after contacting the authors. We extracted data from text or tables, or indirectly from figures using the image analysis software Digitize It 2010 v4.0.2 (A. Carrascal). In several cases we reanalyzed data using reported data (e.g., means and standard deviations, frequency of successful and unsuccessful matings). See online Supplementary Material for our methodology in these cases.

Effect sizes

The studies included in our analysis used a very wide range of statistical tests when testing for mating preferences, which we converted to effect size r (analogous to the correlation coefficient). This effect size can thus be interpreted as the degree of nonrandom response by the chooser with respect to the trait in question (e.g., nonrandom mating or mate association): the larger the test statistic the greater the departure from a random response, and so the "stronger" the mating preference. For many tests the conversion

to r is simple (Koricheva et al. 2013), and it has the advantage of being an intuitive measure of the size of an effect. We used the effect size calculator in Metawin 2.0 (Rosenberg et al. 2000) to convert presented effect sizes into r. In several cases we had to repeat analyses in order to obtain useable test statistics (see Supplementary Material for details). We extracted all effect sizes presented in a study. For most studies multiple effect sizes were reported (e.g., effect sizes were presented for multiple measures of preference from the same individuals, or the same measures of preference for different groups of individuals or populations) and we controlled for this in our analysis by including study as a random factor (see Supplementary Material). In many cases there were different numbers of effect sizes reported for each choice design.

All effect sizes were considered positive except in 3 studies in which the direction of preference differed within a study between tests. In these cases we defined one preference as positive and the other as negative (9 negative effect sizes in the model). In the first case (Wood and Ringo 1980), significant mating preferences were detected for both con- and heterospecific individuals in different tests; here conspecific preference was considered as positive and heterospecific preference was considered as negative. In 2 cases (McNamara et al. 2004; King et al. 2005) significant preferences were detected for both virgin and mated females in different tests; in these cases preference for virgins was considered as positive and preference for mated females was considered as negative. We included the direction of preference in our analysis even when preferences were nonsignificant.

All statistical analyses were performed in R 3.0.1 (R Development Core Team 2012) using the Metafor package v1.9-2 (Viechtbauer 2010).

Meta-analysis

All meta-analyses were performed using Fishers' z transform of the correlation coefficient (Zr). Estimates of mean effect size estimates derived from the models were then converted back to r for presentation. Mean effect size was determined using a random-effects meta-analytic model using the rma.uni function in Metafor. We considered the mean effect size estimate to be significantly different from zero if the 95% confidence intervals (CIs) around the mean did not include zero. Though we have multiple effects sizes per study we did not include study as a factor in the model, as this did not change the results but did greatly increase the model Akaike Information Criteria (AIC) score, suggesting that the basic model was a better fit for the data. We present the results of multivariate meta-analysis models incorporating further random factors in the Supplementary Material (and see phylogenetic methods below). We used the I^2 statistic to determine the amount of heterogeneity in effect sizes across studies; this gives the percentage of variation in effect sizes due to heterogeneity rather than by chance (Higgins et al. 2003). I^2 is preferred over Cochran's Q as the relative amount of heterogeneity in the dataset can be determined (not just a significance value), and it is less affected by the number of effect sizes in the analysis (Higgins et al. 2003). We searched for potential moderators of effect size using meta-analytic mixed models using the rma.uni function (random-effects models with the addition of a categorical fixed-effect, see Koricheva et al. 2013) to test whether sex (male or female choice), trait type (intraspecies, interpopulation, or interspecies choice), taxonomic group (arachnid, crustacean, insect, fish, amphibian, reptile, or bird), or choice measure (matings or proxy measure) had a significant influence on effect size (using the Q_M statistic).

To test for the influence of experimental paradigm on the strength of mating preferences we first calculated mean effect sizes estimates separately for effect sizes from no-choice and choice tests. We then tested for a significant difference between effect sizes derived from the 2 experimental paradigms using a weighted least-squares regression model framework (in meta-analysis terminology this is a form of multilevel meta-regression, see Koricheva et al. 2013). This allows us to control for the non-independence of effect sizes taken from each study by including study as a random factor. Species was also fitted as a random factor, but without the addition of phylogenetic information as this had no effect on the meta-analysis models (see below). For these models effect size was weighted using the study weights derived from the overall random-effects meta-analysis model (for a random-effects model weights are calculated by taking into account the sample size of each study as well as the between-study variance of the dataset). We also obtained mean effect size estimates via random-effects models for no-choice and choice tests further split by our 3 main categorical variables (sex, trait type, and taxonomic group) and tested for a difference between paradigms within each of these subgroups using weighted least-squares regression.

Phylogenetic analysis

Recent studies have shown that the addition of phylogenetic information can have a significant impact on the effect size estimates from meta-analysis models (Chamberlain et al. 2012). We attempted to control for possible non-independence of effect sizes due to shared ancestry by performing a phylogenetically controlled meta-analysis. Briefly (see Supplementary Material for more detailed methods), we first constructed a supertree manually by combining trees (both genetic and taxonomic) from several different sources (see Supplementary Material). Branch lengths were arbitrarily set to one (Hadfield and Nakagawa 2010), and then made ultrametric using the cladogram option in FigTree v1.4 (Andrew Rambaut, 2012). This tree was then imported into the ape package v3.1.1 (Paradis et al. 2004) in Newick format, and a correlation matrix obtained using the vcv function. This correlation matrix could then be incorporated into a multivariate meta-analysis model as an additional random factor.

We ran multivariate meta-analytic models incorporating study, species, and phylogeny as additional random factors using the rma.mv function in Metafor. However, in comparison to these models the basic models gave a much better fit to the data: In all cases, adding these random factors increased the 95% CIs associated with the mean effect size estimates (see Supplementary Figure S2), as well as greatly increasing the model AIC scores, but did not change the significance of the results. Most importantly, in most cases the variance component associated with phylogenetic history was zero (with the exception of some of the smaller models), indicating that the effect sizes used in the analysis were not phylogenetically restricted, and that the increases in 95% CIs were entirely due to the addition of species and study as random factors. We thus present the simpler meta-analytic models here and present the results of the multivariate models in the Supplementary Material. Note that the weighted least-squares regression models presented here do include species and study as random factors.

Publication bias

We tested for 2 types of publication bias. To explore the potential for underreporting of nonsignificant results, we used 3 approaches. Firstly, we calculated fail-safe numbers using both Rosenberg's

method and Orwin's method. Rosenberg's method calculates the number of additional studies (or effect sizes in this case) with a value of zero that would need to be added to the analysis to result in a nonsignificant mean effect size. These additional effect sizes are also weighted by the average sample size of the dataset (Koricheva et al. 2013). Orwin's method calculates the number of additional effect sizes of a given value (set at 0.05) that would be needed to result in a designated "unimportant" mean effect size (again set at 0.05). We also performed a trim-and-fill analysis to test for funnel plot asymmetry, which allowed us to calculate a new effect size estimate after imputing missing studies (see Duval and Tweedie 2000). However, the main assumption of this analysis (that there is a single symmetric distribution of effect sizes) seems unlikely in this case (as there are several potential moderators and high heterogeneity: Koricheva et al. 2013). Finally, we tested for the nonparametric correlation between standardized effect size and study variance (Begg and Mazumdar 1994). We tested for a potential change in the strength of mating preference over time in 2 ways: firstly by testing for the rank correlation between effect size and publication year for each study, and secondly by performing a meta-regression using publication year as a covariate.

Dataset

In total we were able to extract data from 38 studies and 40 species, which gave a total of 214 effect sizes, of which 107 were derived from no-choice tests and 107 from choice tests. A total of 95 effect sizes measured female choice and 119 measured male choice. There were no studies on sex-role–reversed species, though 5 of the studies concerned male choice in fish with paternal care only (Rowland 1982; Jamieson and Colgan 1989; Belles-Isles et al. 1990; Itzkowitz et al. 1998; Wong and Svensson 2009). Totally, 133 effect sizes considered intraspecies choice, 18 considered interpopulation choice, and 63 considered interspecies choice. The analysis includes studies on seven species groups: arachnids (e.g., Parri et al. 1997), crustaceans (e.g., Booksmythe et al. 2011), insects (e.g., Boake and Poulsen 1997; Cook et al. 1994; Jennings et al. 2011; Lehmann and Lehmann 2008; Schöfl et al. 2011; Xu and Wang 2009), fish (e.g., Hurt et al. 2004; Kullmann and Klemme 2007; Suk and Choe 2002), amphibians (e.g., Phelps et al. 2006), reptiles (e.g., Rosenblum 2008), and birds (e.g., Gillingham et al. 2009; Rutstein et al. 2007). Insects and fish were the most common taxonomic groups studied (110 and 67 effect sizes, respectively); the remaining 5 groups all contributed less than 12 effect sizes each to the final analysis. Totally, 166 effect sizes were derived from proxy measures of preference, whereas 48 were derived from mating frequency data. In total, the dataset was based on data from 6322 individual subjects.

Of the 38 papers included in the final analysis, 29 were found using online searches. A further 8 studies were found by following references cited in other papers (Wood and Ringo 1980; Rowland 1982; Houde 1987; Hoikkala and Aspi 1993; Wagner et al. 1995; McNamara et al. 2004; Coyne et al. 2005; King et al. 2005). These studies were likely not detected either because the exact experimental design was not mentioned in the abstract and/or our search terms were not used to refer to the tests. We also included data from our own study which was unpublished at the time of analysis (Dougherty and Shuker 2014).

The raw data are provided as online Supplementary Material, as are details on how we calculated effect sizes (Supplementary Table S2) and the individual effect sizes extracted for all studies (Supplementary Table S3).

RESULTS

Overall, our meta-analysis revealed significant positive mating preferences (mean preference estimate derived from all 214 effect sizes: $r = 0.426$, 95% CI: 0.375–0.474). In fact, mean effect size estimates for all subgroup comparisons were significantly greater than zero, indicating significant mating preferences within all groups (Table 1). The strength of mate preference was significantly larger when tested using a choice paradigm ($r = 0.484$, 95% CI: 0.409–0.552) compared with a no-choice paradigm ($r = 0.364$, 95% CI: 0.297–0.427; weighted least-squares regression, main effect of paradigm: $F_{1, 168} = 12.42$, $P < 0.001$; Figure 2). The variation in effect sizes was large (suggested "high" I^2 values of greater than 75%: Higgins et al. 2003) across the whole dataset ($I^2 = 88.45\%$), as well as for

Table 1

Mean effect size estimates resulting from meta-analysis models performed separately using effect sizes derived from no-choice and choice tests from each subgroup

			No-choice tests					Choice tests				
Group	Studies	Species	Effect sizes	Mean r	Lower 95% CI	Upper 95% CI	I^2 (%)	Effect sizes	Mean r	Lower 95% CI	Upper 95% CI	I^2 (%)
All	38	40	107	0.364	0.297	0.427	85.6	107	0.484	0.409	0.552	89.55
Sex												
Males	20	21	61	0.353	0.259	0.441	86.75	58	0.433	0.318	0.536	90.43
Females	21	25	46	0.376	0.281	0.463	83.75	49	0.535	0.439	0.620	87.72
Trait type												
Intraspecies	29	29	68	0.341	0.251	0.425	82.24	65	0.500	0.408	0.582	86.03
Interpopulation	4	4	9	0.202	0.096	0.305	51.84	9	0.363	0.152	0.542	75.71
Interspecies	7	11	30	0.446	0.331	0.548	88.94	33	0.480	0.321	0.612	94.19
Taxonomic group												
Arachnid	1	1	1	0.500	—	—	—	1	0.744	—	—	—
Crustacean	2	1	5	0.390	−0.045	0.701	60.19	6	0.430	0.308	0.538	0
Insect	17	21	55	0.322	0.218	0.419	92.54	55	0.449	0.325	0.557	94.76
Fish	12	11	33	0.466	0.387	0.538	29.69	34	0.572	0.475	0.655	56.66
Amphibian	3	3	5	0.332	−0.016	0.608	82.01	4	0.595	0.225	0.815	80.55
Reptile	1	1	4	0.271	0.096	0.430	0	3	0.375	0.030	0.640	68.95
Bird	2	2	4	0.332	0.079	0.544	46.34	4	0.394	−0.086	0.725	83.21

All analyses were performed using Fisher's z transform of the correlation coefficient ($\tilde{z}r$), and then converted back to r for presentation. Mean effect size estimates, 95% CIs, and I^2 values were calculated using a random-effects meta-analytic model. CIs for estimates were calculated by bootstrapping 1000 times.

both no-choice tests (I^2 = 85.6%) and choice tests (I^2 = 89.55%), as would be expected for data deriving from multiple species and traits. I^2 values for subgroup models can be seen in Table 2. There was no significant difference in effect sizes derived from choice outcomes or proxy measures of preference (mixed-effects meta-analysis, $Q_{M\,1}$ = 0.4, P = 0.53).

There was no difference in the strength of mating preferences between male and female choice ($Q_{M\,1}$ = 1.83, P= 0.18). Female mating preferences were stronger in choice tests compared with no-choice tests however ($F_{1,\,68}$ = 18.46, P < 0.001; Figure 2), but there was no difference in male mating preferences between choice paradigms ($F_{1,\,95}$ = 1.66, P = 0.2; Figure 2).

Overall, there was no significant difference in the strength of mating preferences between intraspecies, interpopulation, and interspecies choice ($Q_{M\,2}$ = 2.51, P= 0.29). In terms of choice design though, intraspecies mating preferences were stronger in choice tests compared with no-choice tests ($F_{1,\,100}$ = 11.1, P = 0.001; Figure 2), while there was no difference between choice paradigms in terms of the strength of interpopulation choice ($F_{1,\,13}$ = 1.64, P = 0.22; Figure 2) or interspecies choice ($F_{1,\,51}$ = 0.96, P = 0.33; Figure 2).

There was also no overall difference in the strength of mating preferences across the 7 taxonomic groups ($Q_{M\,6}$ = 6.49, P= 0.37). Mating preferences were stronger in choice tests compared with no-choice tests for insects ($F_{1,\,87}$ = 6.24, P = 0.014), fish ($F_{1,\,52}$ = 4.1, P = 0.048), and amphibians ($F_{1,\,5}$ = 11.8, P = 0.02), but not for crustaceans ($F_{1,\,8}$ = 0.007, P = 0.94), reptiles ($F_{1,\,5}$ = 0.47, P = 0.52), or birds ($F_{1,\,5}$ = 0.08, P = 0.78); however, the sample sizes for these groups are small (Table 1).

Publication bias

We found a weak positive correlation between effect size and sample variance (Spearman's rank correlation, r_s = 0.14, P = 0.046. However, there was a much stronger correlation between standardized effect size and variance (τ = 0.16, P < 0.001). This was true for no-choice tests (τ = 0.18, P = 0.006) but not for choice tests (τ = 0.089, P = 0.18). The Rosenberg fail-safe number was 108 797, suggesting that an unrealistic number of studies with an effect size of zero would need to be added to our analysis to give a nonsignificant result. Orwin's fail-safe number was 1757, so that a large number of studies with effect size 0.05 would need to be added for the mean effect size to be reduced to 0.05. A regression test did not detect significant funnel plot asymmetry (Egger's test, t_{212} = 0.52, P = 0.6). However, trim-and-fill analysis detected 33 missing effect sizes on the right hand of the funnel plot (corresponding to large effect sizes, see Supplementary Figure S4). This is likely driven by the large number of effect sizes around Zr = 0, and it is unclear to what extent this represents a signal of publication bias given that these are studies with large effect sizes. Running the model after imputing these missing studies nevertheless leads to an increase in the overall mean effect size (r = 0.5, 95% CI: 0.45–0.54).

There was no significant correlation between effect size and year of publication (r_s = −0.0067, P = 0.92). However, meta-regression detected a weak negative relationship between effect size and publication year ($Q_{M\,1}$ = 4.82, P = 0.028). This can be seen from the cumulative meta-analysis forest plot in the Supplementary Figure S5.

DISCUSSION

Our meta-analysis of 38 studies shows that mating preferences are significantly stronger when tested using a choice test ("medium" effect size of 0.484, see Cohen 1992) compared with a no-choice test ("medium" effect size of 0.364, see Cohen 1992), with a difference in mean effect size of 0.12 between the 2 test designs ("small" effect, see Cohen 1992). Though small, this effect is highly significant and was very consistent across all studies used in the analysis. Our study therefore reiterates the fact that experimental design is an important factor in the measurement of mating preferences (Wagner 1998). This difference in the strength of preference between experimental paradigms was found for studies considering female choice but not those considering male choice, and for studies considering intraspecies choice but not those considering interspecies or interpopulation choice. We found little evidence for publication bias, though we did find a slight decrease in mean effect size with publication year, a common pattern in ecological meta-analyses (Jennions and Møller 2002).

We do not wish to suggest that one experimental design gives a more "accurate" measure of mating preferences than the other, but rather that our results show that the strength of mating preferences (and thus sexual selection) can vary greatly under different experimental designs. The use of different choice paradigms may in part depend on the question an experimenter wishes to ask, and a plurality of approaches may often be useful to tease apart mating preferences. However, we do suggest that the interpretation of our experiments takes this effect into account. Moreover, the 2 choice paradigms broadly correspond to the different forms of mate encounter in the wild (sequential vs. simultaneous encounter), and thus the strength of choice in natural populations may vary significantly between different social or ecological contexts (Jennions and Petrie 1997; Coyne et al. 2005; MacLaren and Rowland 2006; Miller and Svensson 2014). As such, if choice tests are used in the laboratory to test for preferences in species in which mates are mainly encountered sequentially in the wild, then in many cases the strength of mating preference measured may be an overestimate of what occurs in the wild (Barry and Kokko 2010). Indeed, choice tests appear to be the more common experimental design: Owen et al. (2012) estimated that 71% of studies citing Wagner (1998) included choice tests. Clearly the choice of experimental paradigm should depend on the patterns of mate encounter seen in the wild (Coyne et al. 2005; Mendelson and Shaw 2012). However, in many species we simply do not have the data to be able to assess which choice paradigm is the more ecologically realistic (apart from well-known examples such as lek or harem breeders; e.g., Gibson 1996). Two studies included in our meta-analysis illustrate how large the difference in mating preference can be between choice paradigms. The studies consider male mate choice in the mantid *Pseudomantis albofimbriata* (Barry et al. 2010) and in the fiddler crab *Uca mjoebergi* (Booksmythe et al. 2011). In both of these species, field data suggest that males are unlikely to encounter more than one female at a time in the wild, and so no-choice tests seem the most ecologically relevant design to use. However, in both cases significant mating preferences were detected in choice tests but not in the corresponding no-choice tests (Barry et al. 2010; Booksmythe et al. 2011). Therefore, in these cases mating preferences are unlikely to lead to sexual selection in the wild, except for on the rare occasions when males encounter females simultaneously.

We consider there to be 2 important factors that might lead to stronger mating preferences in choice tests. The first is cognition: a subject in a choice test may be better able to compare options comparatively when given a choice, either because the method of mate sampling has evolved under such conditions, or because being able to perceive differences between options becomes easier when

they can be compared simultaneously (Rowland 1982; Bateson and Healy 2005; Beatty and Franks 2012). This hypothesis assumes that the subject has the ability to actively compare options presented simultaneously, an assumption which may not apply to all species, especially if this requires more "complex" cognitive processes. However, the tactics and decision rules used to make mate choice decisions are unknown for most species, and distinguishing between hypotheses is difficult (Gibson and Langen 1996). Indeed it may be that in some species mates are assessed sequentially, perhaps using threshold-based decision rules, even when simultaneous comparisons are available (Gibson 1996; Kacelnik et al. 2011).

The second factor which may influence the strength of preference is the cost associated with rejecting an option in each test. This is because the perceived mate encounter rate is different under the 2 choice designs (Valone et al. 1996). In a choice test the cost of rejecting one of the options is zero, as there is always at least one other option available. Conversely, in a no-choice test the potential cost of rejection is higher due the fact that the likelihood of being presented with another option is unknown to the subject (and may depend on how often the subject has encountered mates before the test: in most cases this is never). If subjects in a no-choice test perceive that the risk of remaining unmated is high then they might be less likely to exhibit any mating preference and be more likely to mate randomly with respect to the stimulus being tested (Werner and Lotem 2006; Barry and Kokko 2010; Booksmythe et al. 2011). This explanation is more general than the one based on cognition: even if this cost of rejection varies between species it will generally always be higher in a no-choice test (compared with zero for choice tests). This leads to the prediction that we should not see any difference in the strength of preference between paradigms once this perceived mate encounter rate has been controlled for, for example, by giving subjects experience of the same number of mates before choice tests. We would also expect that varying the cost of rejection (e.g., by making the sex ratio more biased, or by varying the age of the subjects) should influence the strength of preference observed in no-choice tests (as is seen for example in sequential choice experiments: Milinski and Bakker 1992; Shelly and Bailey 1992; Lehmann 2007; Beckers and Wagner 2011) but should have no effect on the strength of preference in choice tests. Finally, we also predict that the difference in the strength of preference between designs should decrease as the costs of mating and/or reproduction increase (e.g., in species in which females are harmed during mating, or in which females invest heavily in offspring; Halliday 1983): if this cost is sufficiently high it will outweigh the cost of rejection and so subjects should remain choosy even in the no-choice situation.

We did not find stronger mating preferences overall for female choice compared with male choice as predicted. However, we did find that choice paradigm significantly influenced the strength of female choice, but not the strength of male choice. If the benefits of being choosy are higher for females (due to their larger investment in reproduction) then this may lead to stronger mating preferences in situations where the cost of choosing is small, namely in choice tests. Alternatively, males and females may differ in their mate assessment strategies. For example, if males have a threshold of mate quality above which they will accept all females, so that comparison is not important, then the number of options available will not change the patterns of mate choice observed. However, this explanation only holds if males are more likely to use threshold-based tactics for choosing mates, whereas females of the same species use comparative tactics. We also found a difference in the effect of choice paradigm depending on the type of choice, so that there was a significant difference between paradigms for studies considering intraspecific choice but not those considering interpopulation or interspecies choice. However, we are cautious to draw strong conclusions from this comparison due to the small sample sizes for the latter 2 groups. A theory based on the costs of choice would predict the opposite: if mating with the wrong species leads to zero fitness we should expect individuals to be more discriminating when choosing between conspecifics and heterospecifics than when choosing between conspecifics. However, if comparison is not important for species recognition, so that individuals have a threshold above which they accept a partner as a conspecific, the number of options available will not influence the strength of choice. The existence of such a threshold might be more persuasive in terms of con- and heterospecifics as opposed to some continuous measure of quality, for example, as individuals are either conspecifics (so you should consider mating with them) or they are not (so you should ignore them). However, there is still ongoing debate as to whether species recognition and mate choice are different processes or part of a continuum of mate choice (Ryan and Rand 1993; Mendelson and Shaw 2012; Phelps et al. 2006), but hopefully our data will contribute to that debate.

We found no influence of phylogenetic history on the strength of mating preferences across the 38 species included in our analysis. This is perhaps unsurprising given our dataset has several features which may make the detection of a phylogenetic signal unlikely. First, mate choice is predicted to be capable of evolving rapidly and thus is highly evolutionarily labile (Blomberg et al. 2003). Second, our analysis includes preference measures for a wide range of traits, and indeed in most cases the preferences tested are different even for closely related taxa. Finally, we obtained data from a range of species with a very wide taxonomic spread (with the exception of 9 species of *Drosophila*) so that most species are very distantly related. Indeed our method of constructing a phylogenetic tree greatly underestimates the branch lengths between distantly related species. This makes any potential phylogenetic signal very small (Björklund 1997).

Because of this taxonomic spread, our meta-analysis naturally includes a wide range of studies that vary in many aspects of experimental design, not least due to the specific logistic requirements of working with each study species. As few papers explicitly set out to test the effect of experimental paradigm on choice, in many cases confounding variables were not fully controlled for. The strength of meta-analysis is in detecting effects in such heterogeneous data (Koricheva et al. 2013). However, that is not to say that future experimenters should not attempt to control for such variables. We suggest that where possible experiments be fully randomized, and that the same response traits are used as measures of preference in both kinds of tests. A particularly powerful approach is to test the same subjects in both no-choice and choice tests. Only 3 studies in our analysis were able to do this (Rowland 1982; Verrell 1995; MacLaren and Rowland 2006). However, the order with which each individual is tested in each design must be fully randomized so as to avoid or standardize experience effects (e.g., see Reading and Backwell 2007; Wong and Svensson 2009). If individuals are allowed to interact during choice tests, and especially if choice outcomes are recorded, we may be unable to determine the interactions that lead to such outcomes (Martel and Boivin 2011). Similarly, it has been noted that in choice tests in which individuals can interact, intrasexual competition may occur between individuals of the chosen sex, and this may not reflect

the mating preferences of the choosing sex (e.g., Shackleton et al. 2005). There are undoubtedly many other aspects of experimental design that may influence the strength of mating preferences seen in the laboratory; for example, how animals are kept prior to testing (homosexual vs. heterosexual groups; see above), how preferences are scored (e.g., are subjects who do not respond to stimuli included in the analysis?), and even the personality (exploratory tendency) of subjects in tests that use association time as a preference measure (e.g., David and Cézilly 2011). The influence of these factors on the strength of mating preferences is outside the scope of this study, but we suggest that quantification of these effects will be possible.

In conclusion, our study finds that female, intraspecific mating preferences are significantly stronger when tested using a choice paradigm compared with a no-choice paradigm. We suggest that this is due to the increased cost of rejection in no-choice tests. This effect may not be limited to mate choice, but may indeed also be applicable to other areas of behavioral research in which these kinds of choice designs are used, such as studies of foraging (Kacelnik et al. 2011) or predation (Beatty and Franks 2012). We also show that the effect of experimental design on preferences depends on both the type of preference and the sex of the subject used in a test. This suggests that these groups may fundamentally differ in how they choose mates or in the costs of choosing. Importantly, choice tests in the laboratory may systematically inflate estimates of the strength of mating preferences in species in which this situation is demographically unrealistic in the wild. For this reason we recommend that studies of mate choice do not automatically start with choice tests. A plurality of approaches may be useful, but no-choice designs may be the most sensible starting point unless knowledge of the natural behavior of the study species suggests otherwise. Further, only by measuring mate choice in more natural social contexts will we fully understand its role in sexual selection and speciation.

SUPPLEMENTARY MATERIAL

Supplementary material can be found at http://www.beheco.oxfordjournals.org/

FUNDING

This work was funded by the Natural Environment Research Council DTG Award to L.R.D (Grant code: ABE1-NERC11).

We would like to thank M. Ritchie and N. Bailey for extremely helpful comments and suggestions on an earlier version of the manuscript, especially concerning the costs of mating and interspecies choice, and R. Howard for providing summary statistics not available in his published work. We would also like to thank S. Nakagawa and 3 anonymous reviewers for their very helpful and supportive comments which improved the manuscript enormously.

Handling editor: Dr Shinichi Nakagawa

REFERENCES

Andersson MB. 1994. Sexual selection. Princeton (NJ): Princeton University Press.

*Barry KL, Holwell GI, Herberstein ME. 2010. Multimodal mate assessment by male praying mantids in a sexually cannibalistic mating system. Anim Behav. 79:1165–1172.

Barry KL, Kokko H. 2010. Male mate choice: why sequential choice can make its evolution difficult. Anim Behav. 80:163–169.

Basolo AL. 1995. Phylogenetic evidence for the role of a pre-existing bias in sexual selection. Proc Biol Sci. 259:307–311.

Bateson M, Healy SD. 2005. Comparative evaluation and its implications for mate choice. Trends Ecol Evol. 20:659–664.

Beatty C, Franks D. 2012. Discriminative predation: simultaneous and sequential encounter experiments. Curr Zool. 58:649–657.

Beckers OM, Wagner WE Jr. 2013. Parasitoid infestation changes female mating preferences. Anim Behav. 85:791–796.

*Beckers OM, Wagner WE Jr. 2011. Mate sampling strategy in a field cricket: evidence for a fixed threshold strategy with last chance option. Anim Behav. 81:519–527.

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

*Belles-Isles JC, Cloutier D, FitzGerald GJ. 1990. Female cannibalism and male courtship tactics in threespine sticklebacks. Behav Ecol Sociobiol. 26:363–368.

Björklund M. 1997. Are 'comparative methods' always necessary? Oikos. 80:607–612.

Blomberg SP, Garland T Jr, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution. 57:717–745.

*Boake CRB, Poulsen T. 1997. Correlates versus predictors of courtship success: courtship song in *Drosophila silvestris* and *D. heteroneura*. Anim Behav. 54:699–704.

*Booksmythe I, Jennions MD, Backwell PRY. 2011. Male fiddler crabs prefer conspecific females during simultaneous, but not sequential, mate choice. Anim Behav. 81:775–778.

Chamberlain SA, Hovick SM, Dibble CJ, Rasmussen NL, Van Allen BG, Maitner BS, Ahern JR, Bell-Dereske LP, Roy CL, Meza-Lopez M. 2012. Does phylogeny matter? Assessing the impact of phylogenetic information in ecological meta-analysis. Ecol Lett. 15:627–636.

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

*Cook SE, Vernon JG, Bateson M, Guilford T. 1994. Mate choice in the polymorphic African swallowtail butterfly, *Papilio dardanus*: male-like females may avoid sexual harassment. Anim Behav. 47:389–397.

Cotton S, Small J, Pomiankowski A. 2006. Sexual selection and condition-dependent mate preferences. Curr Biol. 16:R755–R765.

*Coyne JA, Elwyn S, Rolán-Alvarez E. 2005. Impact of experimental design on Drosophila sexual isolation studies: direct effects and comparison to field hybridization data. Evolution. 59:2588–2601.

Darwin CR. 1871. The descent of man, and selection in relation to sex. London, UK: John Murray.

David M, Cézilly F. 2011. Personality may confound common measures of mate-choice. PLoS One. 6:e24778.

Doherty JA. 1985. Phonotaxis in the cricket, *Gryllus bimaculatus* DeGeer: comparisons of choice and no-choice paradigms. J Comp Physiol A. 157:279–289.

*Dougherty LR, Shuker DM. 2014. Pre-copulatory sexual selection in the seed bug *Lygaeus equestris*: a comparison of choice and no-choice paradigms. Anim Behav. 89:207–214.

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

*Gabor CR, Krenz JD, Jaeger RG. 2000. Female choice, male interference, and sperm precedence in the red-spotted newt. Behav Ecol. 11:115–124.

*Gershman SN, Sakaluk SK. 2009. No Coolidge effect in decorated crickets. Ethology. 115:774–780.

Gibson RM. 1996. Female choice in sage grouse: the roles of attraction and active comparison. Behav Ecol Sociobiol. 39:55–59.

Gibson RM, Langen TA. 1996. How do animals choose their mates? Trends Ecol Evol. 11:468–470.

*Gillingham MA, Richardson DS, Løvlie H, Moynihan A, Worley K, Pizzari T. 2009. Cryptic preference for MHC-dissimilar females in male red junglefowl, Gallus gallus. Proc Biol Sci. 276:1083–1092.

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

Halliday TR. 1983. The study of mate choice. In: Bateson P, editor. Mate choice. Cambridge, UK: Cambridge University Press. p. 3–32.

Higgins JP, Thompson SG, Deeks JJ, Altman DG. 2003. Measuring inconsistency in meta-analyses. BMJ. 327:557–560.

*Hoikkala A, Aspi J. 1993. Criteria of female mate choice in *Drosophila littoralis, D. montana*, and *D. ezoana*. Evolution. 47:768–777.

*Houde AE. 1987. Mate choice based upon naturally occurring color-pattern variation in a guppy population. Evolution. 41:1–10.
*Hurt CR, Stears-Ellis S, Hughes KA, Hedrick PW. 2004. Mating behaviour in the endangered Sonoran topminnow: speciation in action. Anim Behav. 67:343–351.
*Itzkowitz M, Draud MJ, Barnes JL, Haley M. 1998. Does it matter that male beaugregory damselfish have a mate preference? Behav Ecol Sociobiol. 42:149–155.
*Jamieson IG, Colgan PW. 1989. Eggs in the nests of males and their effect on mate choice in the three-spined stickleback. Anim Behav. 38:859–865.
Janetos AC. 1980. Strategies of female mate choice: a theoretical analysis. Behav Ecol Sociobiol. 7:107–112.
*Jang Y, Gerhardt HC. 2006. Divergence in female calling song discrimination between sympatric and allopatric populations of the southern wood cricket *Gryllus fultoni* (Orthoptera: Gryllidae). Behav Ecol Sociobiol. 60:150–158.
*Jennings JH, Mazzi D, Ritchie MG, Hoikkala A. 2011. Sexual and postmating reproductive isolation between allopatric Drosophila montana populations suggest speciation potential. BMC Evol Biol. 11:68.
Jennions MD, Møller AP. 2002. Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. Proc Biol Sci. 269:43–48.
Jennions MD, Petrie M. 1997. Variation in mate choice and mating preferences: a review of causes and consequences. Biol Rev Camb Philos Soc. 72:283–327.
*Jordan LA, Brooks RC. 2012. Recent social history alters male courtship preferences. Evolution. 66:280–287.
Kacelnik A, Vasconcelos M, Monteiro T, Aw J. 2011. Darwin's 'tug-of-war' vs. starlings' horse-racing': how adaptations for sequential encounters drive simultaneous choice. Behav Ecol Sociobiol. 65:547–558.
*King B, Saporito K, Ellison J, Bratzke R. 2005. Unattractiveness of mated females to males in the parasitoid wasp *Spalangia endius*. Behav Ecol Sociobiol. 57:350–356.
Koricheva J, Gurevitch J, Mengeresen K. 2013. Handbook of meta-analysis in ecology and evolution. Princeton (NJ): Princeton University Press.
*Kullmann H, Klemme B. 2007. Female mating preference for own males on species and population level in Chromaphyosemion killifishes (Cyprinodontiformes, Nothobranchiidae). Zoology (Jena). 110:377–386.
Lehmann GUC. 2007. Density-dependent plasticity of sequential mate choice in a bushcricket (Orthoptera: Tettigoniidae). Aust J Zool. 55:123–130.
*Lehmann GUC, Lehmann AW. 2008. Bushcricket song as a clue for spermatophore size? Behav Ecol Sociobiol. 62:569–578.
*MacLaren RD, Rowland WJ. 2006. Differences in female preference for male body size in *Poecilia latipinna* using simultaneous versus sequential stimulus presentation designs. Behaviour. 143:273–292.
Martel V, Boivin G. 2011. Do choice tests really test choice? J Insect Behav. 24:329–336.
*McNamara KB, Jones TM, Elgar MA. 2004. Female reproductive status and mate choice in the hide beetle, *Dermestes maculatus*. J Insect Behav. 17:337–352.
Mendelson TC, Shaw KL. 2012. The (mis)concept of species recognition. Trends Ecol Evol. 27:421–427.
Milinski M, Bakker TCM. 1992. Costs influence sequential mate choice in sticklebacks, *Gasterosteus aculeatus*. Proc Biol Sci. 250:229–233.
Miller CW, Svensson EI. 2014. Sexual selection in complex environments. Annu Rev Entomol. 59:427–445.
Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 151:264–9, W64.
Moore AJ, Breed MD. 1986. Mate assessment in a cockroach, *Nauphoeta cinerea*. Anim Behav. 34:1160–1165.
Moore AJ, Moore PJ. 1988. Female strategy during mate choice: threshold assessment. Evolution. 42:387–391.
Nakagawa S, Poulin R. 2012. Meta-analytic insights into evolutionary ecology: an introduction and synthesis. Evol Ecol. 26:1085–1099.
*Owen MA, Rohrer K, Howard RD. 2012. Mate choice for a novel male phenotype in zebrafish, *Danio rerio*. Anim Behav. 83:811–820.
Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.
*Parri S, Alatalo RV, Kotiaho J, Mappes J. 1997. Female choice for male drumming in the wolf spider *Hygrolycosa rubrofasciata*. Anim Behav. 53:305–312.
*Phelps SM, Rand AS, Ryan MJ. 2006. A cognitive framework for mate choice and species recognition. Am Nat. 167:28–42.
R Development Core Team. 2012. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
*Reading KL, Backwell PR. 2007. Can beggars be choosers? Male mate choice in a fiddler crab. Anim Behav. 74:867–872.
Rosenberg MS, Adams DC, Gurevitch J. 2000. MetaWin: statistical software for meta-analysis. Sunderland (MA): Sinauer Associates.
*Rosenblum EB. 2008. Preference for local mates in a recently diverged population of the lesser earless lizard (*Holbrookia maculata*) at White Sands. J Herpetol. 42:572–583.
*Rowland WJ. 1982. Mate choice by male sticklebacks, *Gasterosteus aculeatus*. Anim Behav. 30:1093–1098.
*Rutstein AN, Brazill-Boast J, Griffith SC. 2007. Evaluating mate choice in the zebra finch. Anim Behav. 74:1277–1284.
Ryan MJ, Rand AS. 1993. Species recognition and sexual selection as a unitary problem in animal communication. Evolution. 47:647–657.
*Schöfl G, Dill A, Heckel DG, Groot AT. 2011. Allochronic separation versus mate choice: nonrandom patterns of mating between fall armyworm host strains. Am Nat. 177:470–485.
Shackleton MA, Jennions MD, Hunt J. 2005. Fighting success and attractiveness as predictors of male mating success in the black field cricket, *Teleogryllus commodus*: the effectiveness of no-choice tests. Behav Ecol Sociobiol. 58:1–8.
Shelly TE, Bailey WJ. 1992. Experimental manipulation of mate choice by male katydids: the effect of female encounter rate. Behav Ecol Sociobiol. 30:277–282.
Shuker DM. 2010. Sexual selection: endless forms or tangled bank? Anim Behav. 79:e11–e17.
*Suk HY, Choe JC. 2002. Females prefer males with larger first dorsal fins in the common freshwater goby. J Fish Biol. 61:899–914.
Tomaru M, Oguma Y. 2000. Mate choice in Drosophila melanogaster and D. sechellia: criteria and their variation depending on courtship song. Anim Behav. 60:797–804.
Valone TJ, Nordell SE, Giraldeau L-A, Templeton JJ. 1996. The empirical question of thresholds and mechanisms of mate choice. Evol Ecol. 10:447–455.
*Verrell PA. 1995. Males choose larger females as mates in the salamander *Desmognathus santeelah*. Ethology. 99:162–171.
Viechtbauer W. 2010. Conducting meta-analyses in R with the metafor package. J Stat Softw. 36:1–48.
Wagner WE. 1998. Measuring female mating preferences. Anim Behav. 55:1029–1042.
*Wagner WE Jr, Murray A, Cade W, 1995. Phenotypic variation in the mating preferences of female field crickets, *Gryllus integer*. Anim Behav. 49:1269–1281.
Werner NY, Lotem A. 2006. Experimental evidence for male sequential mate preference in a lekking species. Ethology. 112:657–663.
Wong BB, Jennions MD. 2003. Costs influence male mate choice in a freshwater fish. Proc Biol Sci. 270(1 Suppl):S36–S38.
Wong BB, Jennions MD, Keogh JS. 2004. Sequential male mate choice in a fish, the Pacific blue-eye *Pseudomugil signifer*. Behav Ecol Sociobiol. 56:253–256.
*Wong BB, Svensson PA. 2009. Strategic male signalling effort in a desert-dwelling fish. Behav Ecol Sociobiol. 63:543–549.
*Wood D, Ringo JM. 1980. Male mating discrimination in *Drosophila melanogaster, D. simulans* and their hybrids. Evolution. 34:320–329.
*Xu J, Wang Q. 2009. A polyandrous female moth discriminates against previous mates to gain genetic diversity. Anim Behav. 78:1309–1315.

References marked with * were used in the meta-analysis.