



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

Journal:	<i>Behavioral Ecology</i>
Manuscript ID:	BEHECO-2014-0093.R2
Manuscript Type:	Original article
Keywords:	mate choice, mating preferences, meta-analysis, choice test, interspecific choice, experimental design

SCHOLARONE™
Manuscripts

Review Only

Dougherty & Shuker, 2014

Lay Summary

Mating preferences in animals are stronger when individuals have more than one option to choose from. We performed a meta-analysis of studies in which mating preferences were tested using two experimental designs: no-choice tests and choice tests. This difference in preferences between designs was only seen for female, intraspecific mate choice.

Individuals may be less choosy in no-choice tests because the likelihood of encountering another mate is perceived to be lower than in choice tests.

For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Dougherty & Shuker, 2014

1 *For Behavioral Ecology*

2

3 The effect of experimental design on the measurement of mate

4 choice: a meta-analysis

5

6

7 Liam R. Dougherty^{a*} and David M. Shuker^a

8

9

10 ^aSchool of Biology, University of St Andrews, Harold Mitchell Building, St Andrews, KY14

11 7AU, U.K.

12 *Correspondence: Email address: lrd5@st-andrews.ac.uk

13

14

15

16

17

18

19

20

21

22 Short title: Experimental design and mate choice: a meta-analysis

Dougherty & Shuker, 2014

23

24 Abstract

25

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

45

Dougherty & Shuker, 2014

46 Keywords: mate choice; mating preferences; meta-analysis; choice test; interspecific choice;
47 experimental design

48

49 Introduction

50

51 Sexual selection arises via two main mechanisms: intersexual mate-choice and intra-
52 sexual contest competition (Darwin 1871; Andersson 1994). As such, understanding mate
53 choice and the underlying preferences that lead to choice are central to understanding the
54 scope and action of sexual selection. Mate choice is an outcome, and can be defined broadly
55 as arising when a trait in one sex leads to non-random mating success in the other sex
56 (Halliday 1983; Shuker 2010). Therefore choice outcomes are influenced by underlying
57 mating preferences (the sensory and behavioral properties that influence the propensity of
58 individuals to mate with certain phenotypes; Jennions and Petrie 1997) and other factors
59 that affect the expression of these preferences (Jennions and Petrie 1997; Wagner 1998).
60 These include the degree of mate sampling (Janetos 1980; Gibson and Langen 1996; Wagner
61 1998), the condition of the choosing individual (Cotton et al. 2006; Beckers and Wagner
62 2013), and the costs and benefits associated with choice (e.g. Milinski and Bakker 1992).
63 Measurement of mate choice may thus be influenced by experimental design, if different
64 designs vary in any of these factors (Wagner 1998).

65

66 An important way in which experiments testing mate preferences can vary is in the
67 number of options the subject is presented with during the test, which we refer to as the
68 'choice paradigm' or 'choice design'. Tests can use either no-choice or choice designs

Dougherty & Shuker, 2014

1
2
3 69 (Wagner 1998). In a no-choice test each subject is presented with a single stimulus. Several
4
5 70 no-choice trials may be performed using the same subject; these are referred to as
6
7 71 sequential choice tests. In contrast, in a choice test each subject is given a choice between
8
9
10 72 multiple (usually two) stimuli presented simultaneously. The two paradigms differ most
11
12 73 importantly in whether options can be directly compared or not. Because comparison is
13
14 74 possible, choice tests detect relative, directional preferences between stimuli (Wagner
15
16 75 1998; MacLaren and Rowland 2006). As such, these tests may allow greater resolving power
17
18 76 between options as even small differences in trait values may lead to large differences in
19
20 77 choice outcomes (Doherty 1985; Wagner 1998). However, this effect may amplify the
21
22 78 strength of preferences observed if a dichotomous yes or no response is recorded (Wagner
23
24 79 et al. 1995; Wagner 1998). In contrast, no-choice experiments test for absolute preferences
25
26 80 as no direct comparison is possible (Wagner 1998). No-choice tests also differ from choice
27
28 81 tests in that the perceived mate encounter rate is lower: if a mate is rejected in a no-choice
29
30 82 tests there may be no guarantee of a mating opportunity in the future (Werner and Lotem
31
32 83 2006; Barry and Kokko 2010; Booksmythe et al. 2011). Thus rejection of an option in a no-
33
34 84 choice test may indicate a stronger or more robust preference than that seen in a choice
35
36 85 test, because the subject has foregone mating despite this extra 'cost of rejection'. It seems
37
38 86 likely that one or all of these factors may lead to differences in the strength of preferences
39
40 87 observed in each paradigm.
41
42
43
44
45
46
47
48
49

50 89 There are many cases of both no-choice and choice paradigms being used to test for
51
52 90 mating preferences in the same species in different studies. For example, male Pacific Blue-
53
54 91 eye fish *Pseudomugil signifier* prefer larger females in both simultaneous (Wong and
55
56 92 Jennions 2003) and sequential choice tests (Wong et al. 2004). Similarly female cockroaches
57
58
59
60

Dougherty & Shuker, 2014

93 *Nauphoeta cinerea* also prefer dominant males in both no-choice (Moore and Moore 1988)
94 and choice tests (Moore and Breed 1986). A potentially more powerful comparison of
95 paradigms is one in which preferences are tested on the same species in a single study. If
96 experiments are carried out by the same experimenters in a similar way, this may
97 potentially reduce the number of confounding variables that could lead to differences in
98 observed preferences between tests. Several studies have found stronger mating
99 preferences in choice tests compared to no-choice tests in this way (MacLaren and Rowland
100 2006; Barry et al. 2010; Booksmythe et al. 2011; Owen et al. 2012). However, several studies
101 have also shown little effect of choice paradigm on the strength of mating preference (e.g.
102 Gabor et al. 2000; Jang and Gerhardt 2006; Gershman and Sakaluk 2009; Jordan and Brooks
103 2011).

104
105 Here we present a meta-analysis in which we quantify the effect of choice paradigm
106 on the measurement of mate choice. We searched the literature for studies in which mating
107 preferences were tested using both a no-choice and a choice paradigm, on the same
108 species/trait/sex combination. Including the results of two experiments from the same
109 study should reduce confounding factors such as effects associated with individual
110 researchers, animal stocks, and so forth. The effect size used in the analysis can be most
111 simply considered as the degree of non-random response with respect to a partner's trait
112 presumed to be the target of mate choice. We included studies presenting both mate choice
113 outcomes and also proxy measures of mating preference (see below). For clarity we refer to
114 the mean effect sizes derived from our analysis as the 'strength of preference' throughout.
115 We included studies considering both male and female choice, as well as intra-species, inter-
116 population and inter-species choice (see methods). We predict that, for the reasons

Dougherty & Shuker, 2014

1
2
3 117 mentioned above, mating preferences will be significantly stronger for choice tests
4
5 118 compared to no-choice tests. We also predict that overall female choice will be stronger
6
7 119 than male choice, as females generally invest more in each reproductive event and so
8
9
10 120 should be more discriminating in their choice of mate (Andersson 1994). We also predict
11
12 121 that inter-species choice will be stronger than intra-species and inter-population choice, as
13
14 122 there are higher costs associated with making the wrong choice when choosing between a
15
16
17 123 conspecific and a heterospecific individual (Andersson 1994).
18
19
20 124

21 125 **Methods**

22
23
24 126
25
26
27 127 In presenting the methods we have attempted to follow as close as possible the
28
29 128 PRISMA standards for reporting meta-analyses (Moher et al. 2009; see Nakagawa & Poulin
30
31 129 2012; see Figure 1 for diagram showing search results and the study selection process).
32
33
34 130

35 36 131 **Search protocol**

37
38 132
39
40
41 133 We used three approaches to search the literature. First, after initial scoping
42
43 134 searches in September and October 2012, we performed keyword searches of several online
44
45 135 databases in June 2013. We took the first 100 results from the databases Google Scholar
46
47 136 (Google) and Scirus (Elsevier) for the search terms "*sequential simultaneous mate choice*",
48
49 137 on 17th June 2013. On 19th June we performed the following searches in both Web of
50
51 138 Knowledge (Thomson Reuters) (in the TOPIC field) and Scopus (Elsevier) (in the "Article Title,
52
53 139 Abstract, Keywords" field): "*no choice*" AND "*multiple choice*"; "*no choice*" AND "*two*
54
55 140 *choice*"; "*no choice*" AND "*simultaneous*"; "*sequential*" AND "*simultaneous*"; "*sexual*"

Dougherty & Shuker, 2014

1
2
3 141 *isolat** AND “no choice” AND “multiple choice”. The number of results obtained for each
4
5 142 search can be found in the supplementary material (and full endnote libraries are available
6
7 143 on request).
8
9

10 144

11
12 145 Secondly, we used Web of Knowledge to search all studies citing four papers
13
14 146 identified as being influential in this area: the review by Wagner (1998) on measuring
15
16 147 mating preferences and experimental design; the highly-cited study by Rowland (1982) on
17
18 148 male choice in *Gasterosteus aculeatus*; and finally two more recent papers which explicitly
19
20 149 tested for the effect of experimental design on mate preferences (Coyne et al. 2005;
21
22 150 MacLaren and Rowland 2006). After our online searches, we then inspected the titles and
23
24 151 abstracts of the results in order to remove papers that were obviously not relevant to our
25
26 152 search. Papers that were deemed relevant were then read in detail in order to see whether
27
28 153 the study could be included (see inclusion criteria below). Finally, we also followed papers
29
30 154 cited in the text if our searches had not already located them.
31
32
33
34
35

36 155

37 156 **Criteria for inclusion**

38 157

39
40
41 158 We had several criteria for inclusion. Most importantly, each study needed to
42
43 159 include at least one effect size corresponding to a no-choice test and one effect size
44
45 160 corresponding to a choice test (for most studies multiple effect sizes were presented, see
46
47 161 below). We included only studies in which each test was performed using the same species
48
49 162 and sex, testing for a preference for the same trait. This is important as we found several
50
51 163 cases where both no-choice and choice designs were performed but different traits were
52
53 164 considered between tests (see Figure 1 for the most common reasons for excluding papers
54
55
56
57
58
59
60

Dougherty & Shuker, 2014

1
2
3 165 from our analysis, and Table S1 for more detailed information). Importantly, our analysis
4
5 166 includes measures of mate choice in the form of successful matings and also in the form of
6
7
8 167 proxy behavioral measures (such as association time or courtship effort).
9

10 168

11
12 169 Both tests did not have to be performed using identical stimuli (indeed in most cases
13
14
15 170 this would not be possible because individuals of the opposite sex were used as stimuli),
16
17 171 however stimuli did need to be comparable. One example of an excluded study should help
18
19
20 172 to illustrate this point. Basolo (1995) tested for a female preference for males with
21
22 173 (artificial) swords in the unsworded Platyfish *Priapella olmecae*. First the presence of a
23
24 174 preference was tested using a choice test, in which a female chose between a normal,
25
26 175 unsworded male and a male to which an artificial sword had been experimentally added.
27
28 176 Second, no-choice tests were used to test for female preference for swords of differing
29
30
31 177 sizes. However, there was no corresponding no-choice test using an unsworded male.
32
33
34 178 Therefore the choice design tests for a preference for swords whereas the no-choice design
35
36 179 tests for a preference for sword *size*. Therefore we did not include this study in the analysis,
37
38 180 as the stimuli used in each test were not directly comparable.
39

40 181

41
42
43 182 We define a no-choice test as one in which a subject is presented with a single
44
45 183 stimulus or potential mate. This excludes designs commonly used in sexual isolation studies
46
47 184 in which subjects are presented with several potential mates of a single *type* (e.g. Tomaru
48
49 185 and Oguma 2000). This definition also includes sequential choice tests, in which several no-
50
51 186 choice tests are performed concurrently using different stimuli. We define a choice test as
52
53 187 one in which a subject is presented with more than one stimulus simultaneously. Most
54
55
56
57
58
59
60

Dougherty & Shuker, 2014

188 studies use a two-choice test, but we also included those in which more than two options
189 were given (e.g. three-choice test: Beckers & Wagner, 2011).

190

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

203

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

Dougherty & Shuker, 2014

212

213 **Effect sizes**

214

215 The studies included in our analysis used a very wide range of statistical tests when
216 testing for mating preferences, which we converted to effect size r (analogous to the
217 correlation coefficient). This effect size can thus be interpreted as the degree of non-
218 random response by the chooser with respect to the trait in question (e.g. non-random
219 mating or mate association): the larger the test statistic the greater the departure from a
220 random response, and so the 'stronger' the mating preference. For many tests the
221 conversion to r is simple (Koricheva et al, 2013), and it has the advantage of being an
222 intuitive measure of the size of an effect. We used the effect size calculator in Metawin 2.0
223 (Rosenberg et al. 2000) to convert presented effect sizes into r . In several cases we had to
224 repeat analyses in order to obtain useable test statistics (see supplementary material for
225 details). We extracted all effect sizes presented in a study. For most studies multiple effect
226 sizes were reported (for example, effect sizes were presented for multiple measures of
227 preference from the same individuals, or the same measures of preference for different
228 groups of individuals or populations) and we controlled for this in our analysis by including
229 study as a random factor (see supplementary material). In many cases there were different
230 numbers of effect sizes reported for each choice design.

231

232 All effect sizes were considered positive except in three studies in which the
233 direction of preference differed within a study between tests. In these cases we defined one
234 preference as positive and the other as negative (nine negative effect sizes in the model). In
235 the first case (Wood and Ringo 1980), significant mating preferences were detected for both

Dougherty & Shuker, 2014

236 con- and hetero-specific individuals in different tests; here conspecific preference was
237 considered as positive and heterospecific preference was considered as negative. In two
238 cases (McNamara 2004; King et al. 2005) significant preferences were detected for both
239 virgin and mated females in different tests; in these cases preference for virgins was
240 considered as positive and preference for mated females was considered as negative. We
241 included the direction of preference in our analysis even when preferences were non-
242 significant.

243

244 All statistical analyses were performed in R 3.0.1 (R Development Core Team 2013)
245 using the Metafor package v1.9-2 (Viechtbauer 2010).

246

247 **Meta-analysis**

248

249 All meta-analyses were performed using Fishers' z transform of the correlation
250 coefficient (Zr). Estimates of mean effect size estimates derived from the models were then
251 converted back to r for presentation. Mean effect size was determined using a random-
252 effects meta-analytic model using the `rma.uni` function in Metafor. We considered the mean
253 effect size estimate to be significantly different from zero if the 95% confidence intervals
254 around the mean did not include zero. Though we have multiple effects sizes per study we
255 did not include study as a factor in the model, as this did not change the results but did
256 greatly increase the model AIC score, suggesting that the basic model was a better fit for the
257 data. We present the results of multivariate meta-analysis models incorporating further
258 random factors in the supplementary material (and see phylogenetic methods below). We
259 used the I^2 statistic to determine the amount of heterogeneity in effect sizes across studies;

Dougherty & Shuker, 2014

1
2
3 260 this gives the percentage of variation in effect sizes due to heterogeneity rather than by
4
5 261 chance (Higgins et al. 2003). I^2 is preferred over Cochran's Q as the relative amount of
6
7 262 heterogeneity in the dataset can be determined (not just a significance value), and it is less
8
9 263 affected by the number of effect sizes in the analysis (Higgins et al. 2003). We searched for
10
11 264 potential moderators of effect size using meta-analytic mixed models using the rma.uni
12
13 265 function (random-effects models with the addition of a categorical fixed-effect, see
14
15 266 Koricheva et al. 2013) to test whether sex (male or female choice), trait type (intra-species,
16
17 267 inter-population or inter-species choice), taxonomic group (Arachnid, Crustacean, Insect,
18
19 268 Fish, Amphibian, Reptile or Bird) or choice measure (matings or proxy measure) had a
20
21 269 significant influence on effect size (using the Q_M statistic).
22
23
24
25
26
27
28

29 271 To test for the influence of experimental paradigm on the strength of mating
30
31 272 preferences we first calculated mean effect sizes estimates separately for effect sizes from
32
33 273 no-choice and choice tests. We then tested for a significant difference between effect sizes
34
35 274 derived from the two experimental paradigms using a weighted least-squares regression
36
37 275 model framework (in meta-analysis terminology this is a form of multi-level meta-
38
39 276 regression, see Koricheva et al. 2013). This allows us to control for the non-independence of
40
41 277 effect sizes taken from each study by including study as a random factor. Species was also
42
43 278 fitted as a random factor, but without the addition of phylogenetic information as this had
44
45 279 no effect on the meta-analysis models (see below). For these models effect size was
46
47 280 weighted using the study weights derived from the overall random effects meta-analysis
48
49 281 model (for a random-effects model weights are calculated by taking into account the sample
50
51 282 size of each study as well as the between-study variance of the dataset). We also obtained
52
53 283 mean effect size estimates via random-effects models for no-choice and choice tests further
54
55
56
57
58
59
60

Dougherty & Shuker, 2014

284 split by our three main categorical variables (sex, trait type and taxonomic group), and
285 tested for a difference between paradigms within each of these subgroups using weighted
286 least-squares regression.

287

288 **Phylogenetic analysis**

289

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

302

303 We ran multivariate meta-analytic models incorporating study, species and
304 phylogeny as additional random factors using the rma.mv function in Metafor. However in
305 comparison to these models the basic models gave a much better fit to the data: In all cases,
306 adding these random factors increased the 95% confidence intervals associated with the
307 mean effect size estimates (see Figure S2), as well as greatly increasing the model AIC

Dougherty & Shuker, 2014

1
2
3 308 scores, but did not change the significance of the results. Most importantly, in most cases
4
5 309 the variance component associated with phylogenetic history was zero (with the exception
6
7
8 310 of some of the smaller models), indicating that the effect sizes used in the analysis were not
9
10 311 phylogenetically restricted, and that the increases in 95% CI's were entirely due to the
11
12 312 addition of species and study as random factors. We thus present the simpler meta-analytic
13
14 313 models here and present the results of the multivariate models in the supplementary
15
16
17 314 materials. Note that the weighted least-squares regression models presented here do
18
19 315 include species and study as random factors.
20
21
22 316

23 24 317 **Publication bias**

25
26
27 318
28
29 319 We tested for two types of publication bias. To explore the potential for
30
31 320 underreporting of non-significant results, we used three approaches. Firstly, we calculated
32
33 321 fail-safe numbers using both Rosenberg's method and Orwin's method). Rosenberg's
34
35 322 method calculates the number of additional studies (or effect sizes in this case) with a value
36
37 323 of zero that would need to be added to the analysis to result in a non-significant mean
38
39 324 effect size. These additional effect sizes are also weighted by the average sample size of the
40
41 325 dataset (Koricheva et al. 2013). Orwin's method calculates the number of additional effect
42
43 326 sizes of a given value (set at 0.05) that would be needed to result in a designated
44
45 327 'unimportant' mean effect size (again set at 0.05). We also performed a trim-and-fill analysis
46
47 328 to test for funnel plot asymmetry, which allowed us to calculate a new effect size estimate
48
49 329 after imputing missing studies (see Duval and Tweedie 2000). However the main assumption
50
51 330 of this analysis (that there is a single symmetric distribution of effect sizes) seems unlikely in
52
53 331 this case (as there are several potential moderators and high heterogeneity: Koricheva et al,
54
55
56
57
58
59
60

Dougherty & Shuker, 2014

2013). Finally we tested for the non-parametric 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 two 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.

337

338 Dataset

339

340 In total we were able to extract data from 38 studies and 40 species, which gave a
341 total of 214 effect sizes, of which 107 were derived from no-choice tests and 107 from
342 choice tests. 95 effect sizes measured female choice and 119 measured male choice. There
343 were no studies on sex-role reversed species, though five of the studies concerned male
344 choice in fish with paternal care only (Rowland 1982; Jamieson and Colgan 1989; Belles-Isles
345 et al. 1990; Itzkowitz et al. 1998; Wong and Svensson 2009). 133 effect sizes considered
346 intra-species choice, 18 considered inter-population choice and 63 considered inter-species
347 choice. Insects and fish were the most common taxonomic groups studied (110 and 67
348 effect sizes respectively); the remaining five groups all contributed less than 12 effect sizes
349 each to the final analysis. 166 effect sizes were derived from proxy measures of preference
350 whereas 48 were derived from mating frequency data. In total, the dataset was based on
351 data from 6322 individual subjects.

352

353 Of the 38 papers included in the final analysis, 29 were found using online searches.
354 A further eight studies were found by following references cited in other papers (Wood and
355 Ringo 1980; Rowland 1982; Houde 1987; Hoikkala and Aspi 1993; Wagner et al. 1995;

Dougherty & Shuker, 2014

356 McNamara et al. 2004; Coyne et al. 2005; King et al. 2005). These studies were likely not
357 detected either because the exact experimental design was not mentioned in the abstract
358 and/or our search terms were not used to refer to the tests. We also included data from our
359 own study which was unpublished at the time of analysis (Dougherty and Shuker 2014).

360

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

364

365 Results

366

367 Overall, our meta-analysis revealed significant positive mating preferences (mean
368 preference estimate derived from all 214 effect sizes: $r = 0.426$, 95% CI: 0.375 to 0.474). In
369 fact, mean effect size estimates for all subgroup comparisons were significantly greater than
370 zero, indicating significant mating preferences within all groups (Table 1). The strength of
371 mate preference was significantly larger when tested using a choice paradigm ($r = 0.484$,
372 95% CI: 0.409 to 0.552) compared to a no-choice paradigm ($r = 0.364$, 95% CI: 0.297 to 0.427;
373 Weighted least-squares regression, main effect of paradigm: $F_{1,168} = 12.42$, $P < 0.001$; Figure
374 1). The variation in effect sizes was large (Suggested 'high' I^2 values of greater than 75%:
375 Milner et al. 2003) across the whole dataset ($I^2 = 88.45\%$), as well as for both no-choice tests
376 ($I^2 = 85.6\%$) and choice tests ($I^2 = 89.55\%$), as would be expected for data deriving from
377 multiple species and traits. I^2 values for subgroup models can be seen in Table 1. There was
378 no significant difference in effect sizes derived from choice outcomes or proxy measures of
379 preference (Mixed-effects meta-analysis, $Q_{M1} = 0.4$, $P = 0.53$).

Dougherty & Shuker, 2014

380

381 There was no difference in the strength of mating preferences between male and
382 female choice ($Q_{M1} = 1.83, P = 0.18$). Female mating preferences were stronger in choice
383 tests compared to no-choice tests however ($F_{1,68} = 18.46, P < 0.001$; Figure 1), but there was
384 no difference in male mating preferences between choice paradigms ($F_{1,95} = 1.66, P = 0.2$;
385 Figure 1).

386

387 Overall, there was no significant difference in the strength of mating preferences
388 between intra-species, inter-population and inter-species choice ($Q_{M2} = 2.51, P = 0.29$). In
389 terms of choice design though, intra-species mating preferences were stronger in choice
390 tests compared to no-choice tests ($F_{1,100} = 11.1, P = 0.001$; Figure 1), while there was no
391 difference between choice paradigms in terms of the strength of inter-population choice ($F_{1,13} = 1.64, P = 0.22$; Figure 1) or inter-species choice ($F_{1,51} = 0.96, P = 0.33$; Figure 1).

393

394 There was also no overall difference in the strength of mating preferences across the
395 seven taxonomic groups ($Q_{M6} = 6.49, P = 0.37$). Mating preferences were stronger in choice
396 tests compared to no-choice tests for insects ($F_{1,87} = 6.24, P = 0.014$), fish ($F_{1,52} = 4.1, P =$
397 0.048) and amphibians ($F_{1,5} = 11.8, P = 0.02$), but not for crustaceans ($F_{1,8} = 0.007, P = 0.94$),
398 reptiles ($F_{1,5} = 0.47, P = 0.52$) or birds ($F_{1,5} = 0.08, P = 0.78$), however the sample sizes for
399 these groups are small (Table 1).

400

401 **Publication bias**

402

Dougherty & Shuker, 2014

1
2
3 403 We found a weak positive correlation between effect size and sample variance
4
5 404 (Spearman's rank correlation, $r_s = 0.14$, $P = 0.046$. However, there was a much stronger
6
7
8 405 correlation between standardized effect size and variance ($\tau = 0.16$, $P < 0.001$). This was true
9
10 406 for no-choice tests ($\tau = 0.18$, $P = 0.006$) but not for choice tests ($\tau = 0.089$, $P = 0.18$). The
11
12 407 Rosenberg fail-safe number was 108797, suggesting that an unrealistic number of studies
13
14 408 with an effect size of zero would need to be added to our analysis to give a non-significant
15
16
17 409 result. Orwin's fail-safe number was 1757, so that a large number of studies with effect size
18
19 410 0.05 would need to be added for the mean effect size to be reduced to 0.05. A regression
20
21 411 test did not detect significant funnel plot asymmetry (Egger's test, $t_{212} = 0.52$, $P = 0.6$).
22
23 412 However trim and fill analysis detected 33 missing effect sizes on the right hand of the
24
25 413 funnel plot (corresponding to large effect sizes, see Figure S4). This is likely driven by the
26
27 414 large number of effect sizes around $Zr = 0$, and it is unclear to what extent this represents a
28
29 415 signal of publication bias given that these are studies with large effect sizes. Running the
30
31 416 model after imputing these missing studies nevertheless leads to an increase in the overall
32
33 417 mean effect size ($r = 0.5$, 95% CI: 0.45 to 0.54).
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

419 There was no significant correlation between effect size and year of publication ($r_s = -$
420 0.0067 , $P = 0.92$). However meta-regression detected a weak negative relationship between
421 effect size and publication year ($Q_{M1} = 4.82$, $P = 0.028$). This can be seen from the
422 cumulative meta-analysis forest plot in the supplementary material (Figure S5).

424 Discussion

425

Dougherty & Shuker, 2014

1
2
3 426 Our meta-analysis of 38 studies shows that mating preferences are significantly
4
5 427 stronger when tested using a choice test ('medium' effect size of 0.484, see Cohen, 1992)
6
7 428 compared to a no-choice test ('medium' effect size of 0.364, see Cohen, 1992), with a
8
9 429 difference in mean effect size of 0.12 between the two test designs ('small' effect, see
10
11 430 Cohen, 1992). Though small, this effect is highly significant and was very consistent across
12
13 431 all studies used in the analysis. Our study therefore re-iterates the fact that experimental
14
15 432 design is an important factor in the measurement of mating preferences (Wagner 1998).
16
17 433 This difference in the strength of preference between experimental paradigms was found
18
19 434 for studies considering female choice but not those considering male choice, and for studies
20
21 435 considering intra-species choice but not those considering inter-species or inter-population
22
23 436 choice. We found little evidence for publication bias, though we did find a slight decrease in
24
25 437 mean effect size with publication year, a common pattern in ecological meta-analyses
26
27 438 (Jennions & Moller, 2002).
28
29
30
31
32
33
34
35

439

36 440 We do not wish to suggest that one experimental design gives a more 'accurate'
37
38 441 measure of mating preferences than the other, but rather that our results show that the
39
40 442 strength of mating preferences (and thus sexual selection) can vary greatly under different
41
42 443 experimental designs. The use of different choice paradigms may in part depend on the
43
44 444 question an experimenter wishes to ask, and a plurality of approaches may often be useful
45
46 445 to tease apart mating preferences. However, we do suggest that the interpretation of our
47
48 446 experiments takes this effect into account. Moreover, the two choice paradigms broadly
49
50 447 correspond to the different forms of mate encounter in the wild (sequential versus
51
52 448 simultaneous encounter), and thus the strength of choice in natural populations may vary
53
54 449 significantly between different social or ecological contexts (Jennions and Petrie 1997;
55
56
57
58
59
60

Dougherty & Shuker, 2014

1
2
3 450 Coyne et al. 2005; MacLaren and Rowland 2006; Miller and Svensson 2014). As such, if
4
5 451 choice tests are used in the laboratory to test for preferences in species in which mates are
6
7
8 452 mainly encountered sequentially in the wild, then in many cases the strength of mating
9
10 453 preference measured may be an overestimate of what occurs in the wild (Barry and Kokko
11
12 454 2010). Indeed, choice tests appear to be the more common experimental design: Owen et
13
14
15 455 al. (2012) estimated that 71% of studies citing Wagner (1998) included choice tests. Clearly
16
17 456 the choice of experimental paradigm should depend on the patterns of mate encounter
18
19 457 seen in the wild (Coyne et al. 2005; Mendelson and Shaw 2012). However, in many species
20
21 458 we simply do not have the data to be able to assess which choice paradigm is the more
22
23 459 ecologically realistic (apart from well-known examples such as lek or harem breeders; e.g.
24
25
26 460 Gibson 1996). Two studies included in our meta-analysis illustrate how large the difference
27
28 461 in mating preference can be between choice paradigms. The studies consider male mate
29
30 462 choice in the mantid *Pseudomantis albobimbrata* (Barry et al. 2010) and in the fiddler crab
31
32 463 *Uca mjoebergi* (Booksmythe et al. 2010). In both of these species, field data suggest that
33
34 464 males are unlikely to encounter more than one female at a time in the wild, and so no-
35
36 465 choice tests seem the most ecologically relevant design to use. However in both cases
37
38 466 significant mating preferences were detected in choice tests but not in the corresponding
39
40 467 no-choice tests (Barry et al. 2010; Booksmythe et al. 2010). Therefore in these cases mating
41
42 468 preferences are unlikely to lead to sexual selection in the wild, except for on the rare
43
44 469 occasions when males encounter females simultaneously.
45
46
47
48
49
50
51
52

53 471 We consider there to be two important factors that might lead to stronger mating
54
55 472 preferences in choice tests. The first is cognition: a subject in a choice test may be better
56
57 473 able to compare options comparatively when given a choice, either because the method of
58
59
60

Dougherty & Shuker, 2014

1
2
3 474 mate sampling has evolved under such conditions, or because being able to perceive
4
5 475 differences between options becomes easier when they can be compared simultaneously
6
7 476 (Rowland 1982; Bateson and Healy 2005; Beatty and Franks 2012). This hypothesis assumes
8
9 477 that the subject has the ability to actively compare options presented simultaneously, an
10
11 478 assumption which may not apply to all species, especially if this requires more “complex”
12
13 479 cognitive processes. However, the tactics and decision rules used to make mate choice
14
15 480 decisions are unknown for most species, and distinguishing between hypotheses is difficult
16
17 481 (Gibson and Langen 1996). Indeed it may be that in some species mates are assessed
18
19 482 sequentially, perhaps using threshold-based decision rules, even when simultaneous
20
21 483 comparisons are available (Gibson 1996; Kacelnik et al. 2011).
22
23
24
25
26
27
28

29 485 The second factor which may influence the strength of preference is the cost
30
31 486 associated with rejecting an option in each test. This is because the perceived mate
32
33 487 encounter rate is different under the two choice designs (Valone et al. 1996). In a choice
34
35 488 test the cost of rejecting one of the options is zero, as there is always at least one other
36
37 489 option available. Conversely, in a no-choice test the potential cost of rejection is higher due
38
39 490 the fact that the likelihood of being presented with another option is unknown to the
40
41 491 subject (and may depend on how often the subject has encountered mates before the test:
42
43 492 in most cases this is never). If subjects in a no-choice test perceive that the risk of remaining
44
45 493 unmated is high then they might be less likely to exhibit any mating preference and be more
46
47 494 likely to mate randomly with respect to the stimulus being tested (Werner and Lotem 2006;
48
49 495 Barry and Kokko 2010; Booksmythe et al. 2011). This explanation is more general than the
50
51 496 one based on cognition: even if this cost of rejection varies between species it will generally
52
53 497 always be higher in a no-choice test (compared to zero for choice tests). This leads to the
54
55
56
57
58
59
60

Dougherty & Shuker, 2014

1
2
3 498 prediction that we should not see any difference in the strength of preference between
4
5 499 paradigms once this perceived mate encounter rate has been controlled for, for example by
6
7
8 500 giving subjects experience of the same number of mates before choice tests. We would also
9
10 501 expect that varying the cost of rejection (for example by making the sex ratio more biased,
11
12 502 or by varying the age of the subjects) should influence the strength of preference observed
13
14
15 503 in no-choice tests (as is seen for example in sequential choice experiments: Milinski and
16
17 504 Bakker 1992; Shelly and Bailey 1992; Lehmann 2007; Beckers and Wagner 2011) but should
18
19 505 have no effect on the strength of preference in choice tests. Finally, we also predict that the
20
21 506 difference in the strength of preference between designs should decrease as the costs of
22
23 507 mating and/or reproduction increase (for example in species in which females are harmed
24
25 508 during mating, or in which females invest heavily in offspring; Halliday 1983): if this cost is
26
27 509 sufficiently high it will outweigh the cost of rejection and so subjects should remain choosy
28
29 510 even in the no-choice situation.
30
31
32
33
34
35

511

36 512 We did not find stronger mating preferences overall for female choice compared to
37
38 513 male choice as predicted. However we did find that choice paradigm significantly influenced
39
40 514 the strength of female choice, but not the strength of male choice. If the benefits of being
41
42 515 choosy are higher for females (due to their larger investment in reproduction) then this may
43
44 516 lead to stronger mating preferences in situations where the cost of choosing is small,
45
46 517 namely in choice tests. Alternatively, males and females may differ in their mate assessment
47
48 518 strategies. For example, if males have a threshold of mate quality above which they will
49
50 519 accept all females, so that comparison is not important, then the number of options
51
52 520 available will not change the patterns of mate choice observed. However this explanation
53
54 521 only holds if males are more likely to use threshold-based tactics for choosing mates,
55
56
57
58
59
60

Dougherty & Shuker, 2014

1
2
3 522 whereas females of the same species use comparative tactics. We also found a difference in
4
5 523 the effect of choice paradigm depending on the type of choice, so that there was a
6
7
8 524 significant difference between paradigms for studies considering intra-specific choice but
9
10 525 not those considering inter-population or inter-species choice. However we are cautious to
11
12 526 draw strong conclusions from this comparison due to the small sample sizes for the latter
13
14
15 527 two groups. A theory based on the costs of choice would predict the opposite: if mating
16
17 528 with the wrong species leads to zero fitness we should expect individuals to be more
18
19 529 discriminating when choosing between conspecifics and heterospecifics than when choosing
20
21 530 between conspecifics. However, if comparison is not important for species recognition, so
22
23 531 that individuals have a threshold above which they accept a partner as a conspecific, the
24
25 532 number of options available will not influence the strength of choice. The existence of such
26
27 533 a threshold might be more persuasive in terms of con- and hetero-specifics as opposed to
28
29 534 some continuous measure of quality for example, as individuals are either conspecifics (so
30
31 535 you should consider mating with them) or they are not (so you should ignore them).
32
33
34 536 However, there is still ongoing debate as to whether species recognition and mate choice
35
36 537 are different processes or part of a continuum of mate choice (Ryan and Rand 1993;
37
38 538 Mendelson and Shaw 2004; Phelps et al. 2006), but hopefully our data will contribute to
39
40 539 that debate.
41
42
43
44
45
46
47

48 541 We found no influence of phylogenetic history on the strength of mating preferences
49
50 542 across the 38 species included in our analysis. This is perhaps unsurprising given our dataset
51
52 543 has several features which may make the detection of a phylogenetic signal unlikely. First,
53
54 544 mate choice is predicted to be capable of evolving rapidly and thus is highly evolutionarily
55
56 545 labile (Blomberg et al. 2003). Second, our analysis includes preference measures for a wide
57
58
59
60

Dougherty & Shuker, 2014

1
2
3 546 range of traits, and indeed in most cases the preferences tested are different even for
4
5 547 closely related taxa. Finally, we obtained data from a range of species with a very wide
6
7
8 548 taxonomic spread (with the exception of nine species of *Drosophila*) so that most species
9
10 549 are very distantly related. Indeed our method of constructing a phylogenetic tree greatly
11
12 550 underestimates the branch lengths between distantly related species. This makes any
13
14
15 551 potential phylogenetic signal very small (Björklund 1997).
16

17 552

19 553 Because of this taxonomic spread, our meta-analysis naturally includes a wide range
20
21 554 of studies that vary in many aspects of experimental design, not least due to the specific
22
23 555 logistic requirements of working with each study species. As few papers explicitly set out to
24
25
26 556 test the effect of experimental paradigm on choice, in many cases confounding variables
27
28 557 were not fully controlled for. The strength of meta-analysis is in detecting effects in such
29
30
31 558 heterogeneous data (Koricheva et al. 2013). However, that is not to say that future
32
33
34 559 experimenters should not attempt to control for such variables. We suggest that where
35
36 560 possible experiments be fully randomized, and that the same response traits are used as
37
38 561 measures of preference in both kinds of tests. A particularly powerful approach is to test the
39
40
41 562 same subjects in both no-choice and choice tests. Only three studies in our analysis were
42
43 563 able to do this (Rowland 1982; Verrell 1995; MacLaren and Rowland 2006). However the
44
45 564 order with which each individual is tested in each design must be fully randomized so as to
46
47 565 avoid or standardize experience effects (e.g. see Reading and Backwell 2007; Wong and
48
49 566 Svensson 2009). If individuals are allowed to interact during choice tests, and especially if
50
51
52 567 choice outcomes are recorded, we may be unable to determine the interactions that lead to
53
54
55 568 such outcomes (Martel and Boivin 2011). Similarly, it has been noted that in choice tests in
56
57 569 which individuals can interact, intrasexual competition may occur between individuals of
58
59
60

Dougherty & Shuker, 2014

1
2
3 570 the chosen sex, and this may not reflect the mating preferences of the choosing sex (e.g.
4
5 571 Shackleton et al. 2005). There are undoubtedly many other aspects of experimental design
6
7 572 that may influence the strength of mating preferences seen in the laboratory; for example
8
9 573 how animals are kept prior to testing (homosexual vs heterosexual groups; see above), how
10
11 574 preferences are scored (for example: are subjects who do not respond to stimuli included in
12
13 575 the analysis?), and even the personality (exploratory tendency) of subjects in tests that use
14
15 576 association time as a preference measure (e.g. David & Cezily 2011). The influence of these
16
17 577 factors on the strength of mating preferences is outside the scope of this study, but we
18
19 578 suggest that quantification of these effects will be possible.
20
21
22
23
24
25
26

27 580 In conclusion, our study finds that female, intra-specific mating preferences are
28
29 581 significantly stronger when tested using a choice paradigm compared to a no-choice
30
31 582 paradigm. We suggest that this is due to the increased cost of rejection in no-choice tests.
32
33 583 This effect may not be limited to mate choice, but may indeed also be applicable to other
34
35 584 areas of behavioral research in which these kinds of choice designs are used, such as studies
36
37 585 of foraging (Kacelnik et al. 2011) or predation (Beatty and Franks 2012). We also show that
38
39 586 the effect of experimental design on preferences depends on both the type of preference
40
41 587 and the sex of the subject used in a test. This suggests that these groups may fundamentally
42
43 588 differ in how they choose mates or in the costs of choosing. Importantly, choice tests in the
44
45 589 laboratory may systematically inflate estimates of the strength of mating preferences in
46
47 590 species in which this situation is demographically unrealistic in the wild. For this reason we
48
49 591 recommend that studies of mate choice do not automatically start with choice tests. A
50
51 592 plurality of approaches may be useful, but no-choice designs may be the most sensible
52
53 593 starting point unless knowledge of the natural behavior of the study species suggests
54
55
56
57
58
59
60

Dougherty & Shuker, 2014

594 otherwise. Further, only by measuring mate choice in more natural social contexts will we
595 fully understand its role in sexual selection and speciation.

596

597

598 Acknowledgements

599

600 We would like to thank Mike Ritchie and Nathan Bailey for extremely helpful
601 comments and suggestions on an earlier version of the manuscript, especially concerning
602 the costs of mating and inter-species choice, and Rick Howard for providing summary
603 statistics not available in his published work. We would also like to thank Shinichi Nakagawa
604 and three anonymous reviewers for their very helpful and supportive comments which
605 improved the manuscript enormously. This work was funded by the Natural Environment
606 Research Council DTG Award to LRD.

607

608

609 References

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

611

612 Andersson MB, 1994. Sexual selection. Princeton, NJ: Princeton University Press.

613 *Barry KL, Holwell GI, Herberstein ME, 2010. Multimodal mate assessment by male praying

614 mantids in a sexually cannibalistic mating system. Anim Behav. 79:1165-1172.

615 Barry KL, Kokko H, 2010. Male mate choice: why sequential choice can make its evolution

616 difficult. Anim Behav. 80:163-169.

Dougherty & Shuker, 2014

617 Basolo AL, 1995. Phylogenetic evidence for the role of a pre-existing bias in sexual selection.

618 Proc Biol Sci. 259:307-311.

619 Bateson M, Healy SD, 2005. Comparative evaluation and its implications for mate choice.

620 Trends Ecol Evol. 20:659-664.

621 Beatty C, Franks D, 2012. Discriminative predation: simultaneous and sequential encounter

622 experiments. Curr Zool. 58:649-657.

623 Beckers OM, Wagner Jr WE, 2013. Parasitoid infestation changes female mating

624 preferences. Anim Behav. 85:791-796.

625 *Beckers OM, Wagner WE, Jr., 2011. Mate sampling strategy in a field cricket: evidence for a

626 fixed threshold strategy with last chance option. Anim Behav. 81:519-527.

627 Begg CB, Mazumdar M, 1994. Operating characteristics of a rank correlation test for

628 publication bias. Biometrics. 50:1088-1101.

629 *Belles-Isles JC, Cloutier D, FitzGerald GJ, 1990. Female cannibalism and male courtship

630 tactics in threespine sticklebacks. Behav Ecol Sociobiol. 26:363-368.

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

632 Blomberg SP, Garland T, Ives AR, 2003. Testing for phylogenetic signal in comparative data:

633 behavioral traits are more labile. Evolution 57:717-745.

634 *Boake CRB, Poulsen T, 1997. Correlates versus predictors of courtship success: courtship

635 song in *Drosophila silvestris* and *D.heteroneura*. Anim Behav. 54:699-704.

636 *Booksmythe I, Jennions MD, Backwell PRY, 2011. Male fiddler crabs prefer conspecific

637 females during simultaneous, but not sequential, mate choice. Anim Behav. 81:775-

638 778.

639 Chamberlain SA, Hovick SM, Dibble CJ, Rasmussen NL, Van Allen BG, Maitner BS, Ahern JR,

640 Bell-Dereske LP, Roy CL, Meza-Lopez M, 2012. Does phylogeny matter? Assessing the

Dougherty & Shuker, 2014

- 1
2
3 641 impact of phylogenetic information in ecological meta-analysis. Ecology Letters
4
5 642 15:627-636.
6
7
8 643 Cohen J, 1992. A power primer. Psychological Bulletin 112:155.
9
10 644 *Cook SE, Vernon JG, Bateson M, Guilford T, 1994. Mate choice in the polymorphic African
11
12 645 swallowtail butterfly, *Papilio dardanus*: male-like females may avoid sexual
13
14 646 harassment. Anim Behav. 47:389-397.
15
16
17 647 Cotton S, Small J, Pomiankowski A, 2006. Sexual selection and condition-dependent mate
18
19 648 preferences. Curr Biol. 16:R755-R765.
20
21
22 649 *Coyne JA, Elwyn S, Rolán-Alvarez E, 2005. Impact of experimental design on *Drosophila*
23
24 650 sexual isolation studies: direct effects and comparison to field hybridization data.
25
26 651 Evolution. 59:2588-2601.
27
28
29 652 Darwin CR, 1871. The descent of man, and selection in relation to sex. London, UK: John
30
31 653 Murray.
32
33
34 654 David M, Cézilly F, 2011. Personality may confound common measures of mate-choice. Plos
35
36 655 One 6:e24778.
37
38 656 Doherty JA, 1985. Phonotaxis in the cricket, *Gryllus bimaculatus* DeGeer: comparisons of
39
40 657 choice and no-choice paradigms. J Comp Physiol A. 157:279-289.
41
42
43 658 *Dougherty LR, Shuker DM, 2014. Pre-copulatory sexual selection in the seed bug *Lygaeus*
44
45 659 *equestris*: a comparison of choice and no-choice paradigms. Anim Behav. 89:207-
46
47 660 214.
48
49
50 661 Duval S, Tweedie R, 2000. Trim and fill: a simple funnel-plot-based method of testing and
51
52 662 adjusting for publication bias in meta-analysis. Biometrics. 56:455-463.
53
54
55 663 *Gabor CR, Krenz JD, Jaeger RG, 2000. Female choice, male interference, and sperm
56
57 664 precedence in the red-spotted newt. Behav Ecol. 11:115-124.
58
59
60

Dougherty & Shuker, 2014

- 1
2
3 665 *Gershman SN, Sakaluk SK, 2009. No Coolidge Effect in decorated crickets. *Ethology*.
4
5 666 115:774-780.
6
7 667 Gibson RM, 1996. Female choice in sage grouse: the roles of attraction and active
8
9 668 comparison. *Behav Ecol Sociobiol*. 39:55-59.
10
11 669 Gibson RM, Langen TA, 1996. How do animals choose their mates? *Trends Ecol Evol*. 11:468-
12
13 670 470.
14
15 671 *Gillingham MAF, Richardson DS, Løvlie H, Moynihan A, Worley K, Pizzari T, 2009. Cryptic
16
17 672 preference for MHC-dissimilar females in male red junglefowl, *Gallus gallus*. *Proc*
18
19 673 *Biol Sci*. 276:1083-1092.
20
21 674 Hadfield J, Nakagawa S, 2010. General quantitative genetic methods for comparative
22
23 675 biology: phylogenies, taxonomies and multi-trait models for continuous and
24
25 676 categorical characters. *Journal of Evolutionary Biology* 23:494-508.
26
27 677 Halliday TR, 1983. The study of mate choice. In: Bateson P, editor. *Mate choice*. Cambridge,
28
29 678 UK: Cambridge University Press. p. 3-32.
30
31 679 Higgins J, Thompson SG, Deeks JJ, Altman DG, 2003. Measuring inconsistency in meta-
32
33 680 analyses. *Br Med J*. 327:557-560.
34
35 681 *Hoikkala A, Aspi J, 1993. Criteria of female mate choice in *Drosophila littoralis*, *D. montana*,
36
37 682 and *D. ezoana*. *Evolution*. 47:768-777.
38
39 683 *Houde AE, 1987. Mate choice based upon naturally occurring color-pattern variation in a
40
41 684 guppy population. *Evolution*. 41:1-10.
42
43 685 *Hurt CR, Stears-Ellis S, Hughes KA, Hedrick PW, 2004. Mating behaviour in the endangered
44
45 686 Sonoran topminnow: speciation in action. *Anim Behav*. 67:343-351.
46
47 687 *Itzkowitz M, Draud MJ, Barnes JL, Haley M, 1998. Does it matter that male beaugregory
48
49 688 damselfish have a mate preference? *Behav Ecol Sociobiol*. 42:149-155.
50
51
52
53
54
55
56
57
58
59
60

Dougherty & Shuker, 2014

689 *Jamieson IG, Colgan PW, 1989. Eggs in the nests of males and their effect on mate choice

690 in the three-spined stickleback. *Anim Behav.* 38:859-865.

691 Janetos AC, 1980. Strategies of female mate choice: a theoretical analysis. *Behav Ecol*

692 *Sociobiol.* 7:107-112.

693 *Jang Y, Gerhardt HC, 2006. Divergence in female calling song discrimination between

694 sympatric and allopatric populations of the southern wood cricket *Gryllus fultoni*

695 (Orthoptera: Gryllidae). *Behav Ecol Sociobiol.* 60:150-158.

696 *Jennings JH, Mazzi D, Ritchie MG, Hoikkala A, 2011. Sexual and postmating reproductive

697 isolation between allopatric *Drosophila montana* populations suggest speciation

698 potential. *BMC Evol Biol.* 11.

699 Jennions MD, Møller AP, 2002. Relationships fade with time: a meta-analysis of temporal

700 trends in publication in ecology and evolution. *Proceedings of the Royal Society of*

701 *London Series B: Biological Sciences* 269:43-48.

702 Jennions MD, Petrie M, 1997. Variation in mate choice and mating preferences: a review of

703 causes and consequences. *Biol Rev.* 72:283-327.

704 *Jordan LA, Brooks RC, 2012. Recent social history alters male courtship preferences.

705 *Evolution.* 66:280-287.

706 Kacelnik A, Vasconcelos M, Monteiro T, Aw J, 2011. Darwin's 'tug-of-war' vs.

707 starlings' 'horse-racing': how adaptations for sequential encounters drive

708 simultaneous choice. *Behav Ecol Sociobiol.* 65:547-558.

709 *King B, Saporito K, Ellison J, Bratzke R, 2005. Unattractiveness of mated females to males in

710 the parasitoid wasp *Spalangia endius*. *Behav Ecol Sociobiol.* 57:350-356.

711 Koricheva J, Gurevitch J, Mengersen K, 2013. *Handbook of meta-analysis in ecology and*

712 *evolution.* Princeton, NJ: Princeton University Press.

Dougherty & Shuker, 2014

- 1
2
3 713 *Kullmann H, Klemme B, 2007. Female mating preference for own males on species and
4
5 714 population level in Chromaphyosemion killifishes (Cyprinodontiformes,
6
7 715 Nothobranchiidae). *Zoology*. 110:377-386.
8
9
10 716 Lehmann GUC, 2007. Density-dependent plasticity of sequential mate choice in a
11
12 717 bushcricket (Orthoptera: Tettigoniidae). *Aust J Zool*. 55:123-130.
13
14
15 718 *Lehmann GUC, Lehmann AW, 2008. Bushcricket song as a clue for spermatophore size?
16
17 719 *Behav Ecol Sociobiol*. 62:569-578.
18
19
20 720 *MacLaren RD, Rowland WJ, 2006. Differences in female preference for male body size in
21
22 721 *Poecilia latipinna* using simultaneous versus sequential stimulus presentation
23
24 722 designs. *Behaviour*. 143:273-292.
25
26
27 723 Martel V, Boivin G, 2011. Do choice tests really test choice? *J Insect Behav*. 24:329-336.
28
29 724 *McNamara KB, Jones TM, Elgar MA, 2004. Female reproductive status and mate choice in
30
31 725 the hide beetle, *Dermestes maculatus*. *J Insect Behav*. 17:337-352.
32
33
34 726 Mendelson TC, Shaw KL, 2012. The (mis)concept of species recognition. *Trends Ecol Evol*.
35
36 727 27:421-427.
37
38 728 Milinski M, Bakker TCM, 1992. Costs influence sequential mate choice in sticklebacks,
39
40 729 *Gasterosteus aculeatus*. *Proc Biol Sci*. 250:229-233.
41
42
43 730 Miller CW, Svensson EI, 2014. Sexual selection in complex environments. *Annu Rev Entomol*.
44
45 731 59:427-445.
46
47
48 732 Moher D, Liberati A, Tetzlaff J, Altman DG, 2009. Preferred reporting items for systematic
49
50 733 reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*
51
52 734 151:264-269.
53
54
55 735 Moore AJ, Breed MD, 1986. Mate assessment in a cockroach, *Nauphoeta cinerea*. *Anim*
56
57 736 *Behav*. 34:1160-1165.
58
59
60

Dougherty & Shuker, 2014

- 1
2
3 737 Moore AJ, Moore PJ, 1988. Female strategy during mate choice: threshold assessment.
4
5 738 Evolution. 42:387-391.
6
7 739 Nakagawa S, Poulin R, 2012. Meta-analytic insights into evolutionary ecology: an
8
9 740 introduction and synthesis. *Evolutionary Ecology* 26:1085-1099.
10
11 741 *Owen MA, Rohrer K, Howard RD, 2012. Mate choice for a novel male phenotype in
12
13 742 zebrafish, *Danio rerio*. *Anim Behav.* 83:811-820.
14
15 743 Paradis E, Claude J, Strimmer K, 2004. APE: analyses of phylogenetics and evolution in R
16
17 744 language. *Bioinformatics* 20:289-290.
18
19 745 *Parri S, Alatalo RV, Kotiaho J, Mappes J, 1997. Female choice for male drumming in the
20
21 746 wolf spider *Hygrolycosa rubrofasciata*. *Anim Behav.* 53:305-312.
22
23 747 *Phelps SM, Rand AS, Ryan MJ, 2006. A cognitive framework for mate choice and species
24
25 748 recognition. *Am Nat.* 167:28-42.
26
27 749 R Development Core Team, 2012. R: A language and environment for statistical computing.
28
29 750 Vienna, Austria: R Foundation for Statistical Computing.
30
31 751 *Reading KL, Backwell PR, 2007. Can beggars be choosers? Male mate choice in a fiddler
32
33 752 crab. *Anim Behav.* 74:867-872.
34
35 753 Rosenberg MS, Adams DC, Gurevitch J, 2000. MetaWin: statistical software for meta-
36
37 754 analysis. Sunderland, MA: Sinauer Associates.
38
39 755 *Rosenblum EB, 2008. Preference for local mates in a recently diverged population of the
40
41 756 lesser earless lizard (*Holbrookia maculata*) at White Sands. *J Herpetol.* 42:572-583.
42
43 757 *Rowland WJ, 1982. Mate choice by male sticklebacks, *Gasterosteus aculeatus*. *Anim Behav.*
44
45 758 30:1093-1098.
46
47 759 *Rutstein AN, Brazill-Boast J, Griffith SC, 2007. Evaluating mate choice in the zebra finch.
48
49 760 *Anim Behav.* 74:1277-1284.
50
51
52
53
54
55
56
57
58
59
60

Dougherty & Shuker, 2014

- 1
2
3 761 Ryan MJ, Rand AS, 1993. Species recognition and sexual selection as a unitary problem in
4
5 762 animal communication. *Evolution*. 47:647-657.
6
7
8 763 *Schofl G, Dill A, Heckel DG, Groot AT, 2011. Allochronic separation versus mate choice:
9
10 764 nonrandom patterns of mating between fall armyworm host strains. *Am Nat*.
11
12 765 177:470-485.
13
14
15 766 Shackleton MA, Jennions MD, Hunt J, 2005. Fighting success and attractiveness as predictors
16
17 767 of male mating success in the black field cricket, *Teleogryllus commodus*: the
18
19 768 effectiveness of no-choice tests. *Behav Ecol Sociobiol*. 58:1-8.
20
21
22 769 Shelly TE, Bailey WJ, 1992. Experimental manipulation of mate choice by male katydids: the
23
24 770 effect of female encounter rate. *Behav Ecol Sociobiol*. 30:277-282.
25
26
27 771 Shuker DM, 2010. Sexual selection: endless forms or tangled bank? *Anim Behav*. 79:e11-
28
29 772 e17.
30
31 773 *Suk HY, Choe JC, 2002. Females prefer males with larger first dorsal fins in the common
32
33 774 freshwater goby. *J Fish Biol*. 61:899-914.
34
35
36 775 Tomaru M, Oguma Y, 2000. Mate choice in *Drosophila melanogaster* and *D. sechellia*:
37
38 776 criteria and their variation depending on courtship song. *Anim Behav*. 60:797-804.
39
40
41 777 Valone TJ, Nordell SE, Giraldeau L-A, Templeton JJ, 1996. The empirical question of
42
43 778 thresholds and mechanisms of mate choice. *Evol Ecol*. 10:447-455.
44
45
46 779 *Verrell PA, 1995. Males choose larger females as mates in the salamander *Desmognathus*
47
48 780 *santeelah*. *Ethology*. 99:162-171.
49
50
51 781 Viechtbauer W, 2010. Conducting meta-analyses in R with the metafor package. *J Stat*
52
53 782 *Softw*. 36:1-48.
54
55 783 Wagner Jr WE, 1998. Measuring female mating preferences. *Anim Behav*. 55:1029-1042.
56
57
58
59
60

Dougherty & Shuker, 2014

- 1
2
3 784 *Wagner Jr WE, Murray A, Cade W, 1995. Phenotypic variation in the mating preferences of
4
5 785 female field crickets, *Gryllus integer*. Anim Behav. 49:1269-1281.
6
7
8 786 Werner NY, Lotem A, 2006. Experimental evidence for male sequential mate preference in a
9
10 787 lekking species. Ethology. 112:657-663.
11
12 788 Wong BB, Jennions MD, 2003. Costs influence male mate choice in a freshwater fish. Proc
13
14 789 Biol Sci. 270:S36-S38.
15
16
17 790 Wong BB, Jennions MD, Keogh JS, 2004. Sequential male mate choice in a fish, the Pacific
18
19 791 blue-eye *Pseudomugil signifer*. Behav Ecol Sociobiol. 56:253-256.
20
21
22 792 *Wong BB, Svensson PA, 2009. Strategic male signalling effort in a desert-dwelling fish.
23
24 793 Behav Ecol Sociobiol. 63:543-549.
25
26
27 794 *Wood D, Ringo JM, 1980. Male mating discrimination in *Drosophila melanogaster*, *D.*
28
29 795 *simulans* and their hybrids. Evolution. 34:320-329.
30
31 796 *Xu J, Wang Q, 2009. A polyandrous female moth discriminates against previous mates to
32
33 797 gain genetic diversity. Anim Behav. 78:1309-1315.
34
35
36 798
37
38 799
39
40 800
41
42
43 801
44
45 802
46
47
48 803
49
50 804
51
52 805
53
54
55 806
56
57
58
59
60

Dougherty & Shuker, 2014

807 **Figure 1.** PRISMA flow chart of search results and the study selection process. See table S1
808 for list of papers excluded from the analysis.

809

810 **Figure 2.** Mean strength of mating preferences (correlation coefficient r) for the two choice
811 designs (white diamonds for no-choice tests and black diamonds for choice tests) for all
812 effect sizes and split by sex and trait type. Bars show the bootstrapped 95% confidence
813 intervals around the mean effect size estimates derived from the mate-analytic models. See
814 Table 1 for sample sizes (number of effect sizes) associated with each subgroup. See
815 methods for classification of trait types. P values show the results of a weighted least-
816 squares regression testing the effect of choice paradigm on mean effect size for each
817 subgroup of studies (see results, *** $P < 0.0001$, ** $P < 0.001$). Statistical analyses were
818 performed using Fisher's z transform of the correlation coefficient (Z_r), and then converted
819 back to r for presentation.

Dougherty & Shuker, 2014

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

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												
Intra-species	29	29	68	0.341	0.251	0.425	82.24	65	0.500	0.408	0.582	86.03
Inter-population	4	4	9	0.202	0.096	0.305	51.84	9	0.363	0.152	0.542	75.71
Inter-species	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

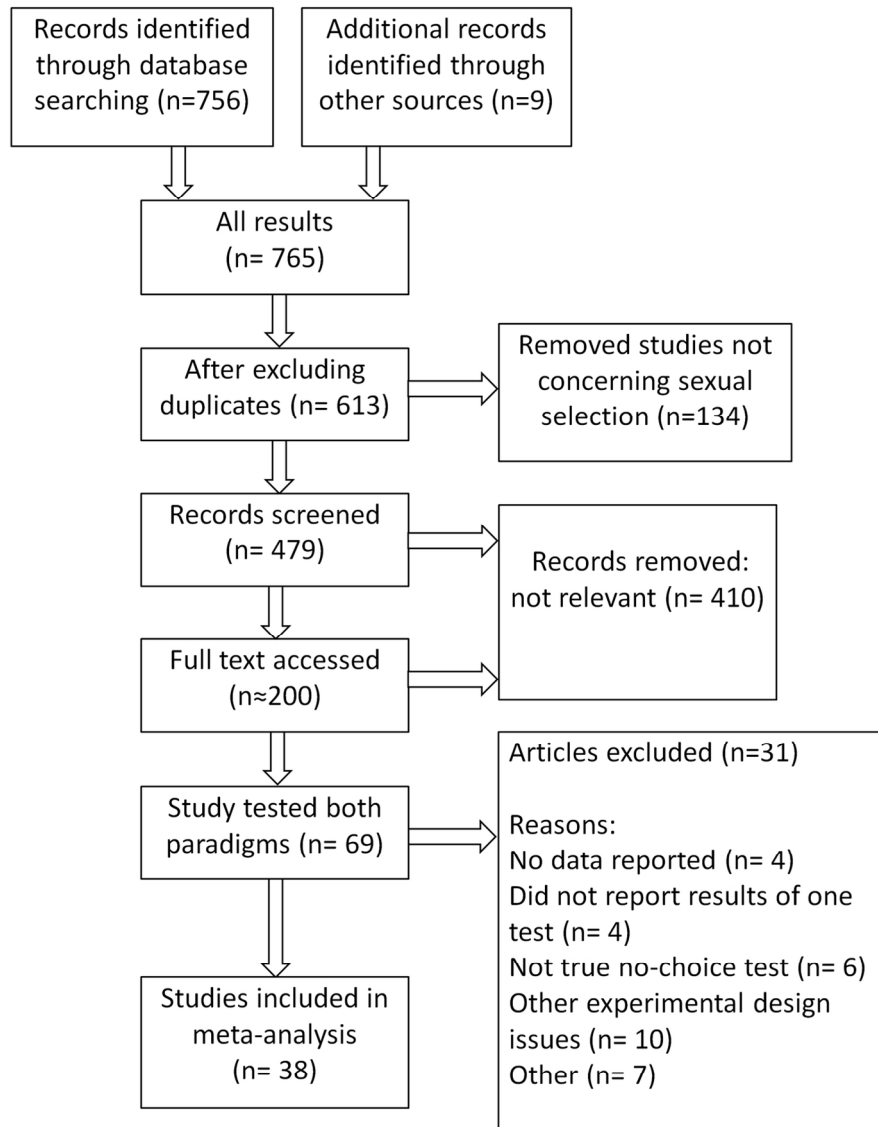


Figure 1. PRISMA flow chart of search results and the study selection process. See table S1 for list of papers excluded from the analysis.

62x79mm (600 x 600 DPI)

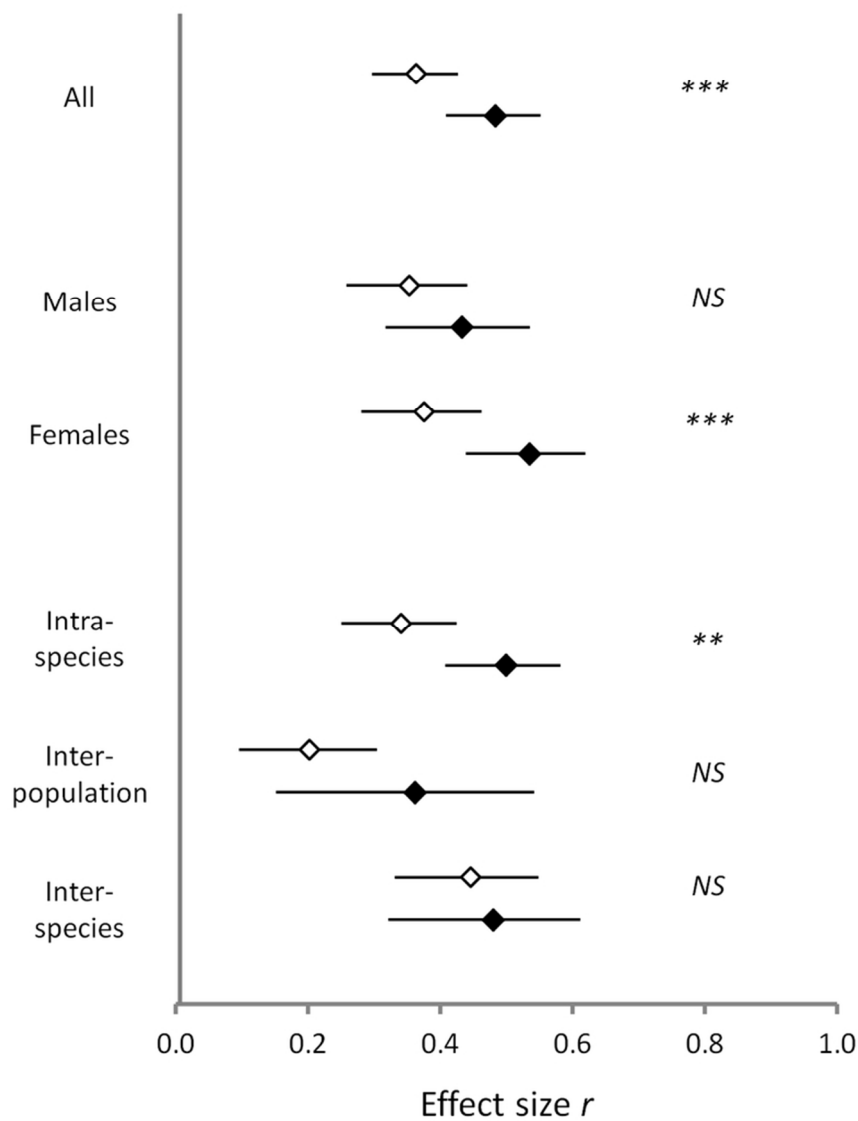


Figure 2. Mean strength of mating preferences (correlation coefficient r) for the two choice designs (white diamonds for no-choice tests and black diamonds for choice tests) for all effect sizes and split by sex and trait type. Bars show the bootstrapped 95% confidence intervals around the mean effect size estimates derived from the mate-analytic models. See Table 1 for sample sizes (number of effect sizes) associated with each subgroup. See 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 (Z_r), and then converted back to r for presentation.

35x46mm (600 x 600 DPI)