Title

A new method for statistical detection of directional and stabilizing mating preference

Journal

Behavioral Ecology, 28(3)

ISSN

1045-2249

Authors

Roff, Derek A
Fairbairn, Daphne J
Prokuda, Alexandra

Publication Date

2017-05-01

DOI

10.1093/beheco/arx038

Peer reviewed

The published version of this paper is:

Roff, D. A., Fairbairn, D. J. and Prokuda, A., 2017. A new method for statistical detection of directional and stabilizing preference. *Behavioral Ecology* 28: 934-942. DOI: https://doi.org/10.1093/beheco/arx038.

LAY SUMMARY: The estimation of mating preference is essential to understand how sexual selection through mate choice shapes both mating systems and sexual dimorphisms. We present a new method for detecting and quantifying both stabilizing and directional mating preferences and demonstrate the experimental and statistical advantages of this method over previous approaches. We illustrate the method using data from a mate choice experiment in which female sand crickets were provided a choice of two males. Both directional and stabilizing preference are identified for components of the male call song.

A NEW METHOD FOR STATISTICAL DETECTION OF DIRECTIONAL AND STABILIZING MATING PREFERENCE

Derek A. Roff,[a] Daphne J. Fairbairn[b] and Alexandra Prokuda[c]

Department of Biology,

University of California,

Riverside, 92521,

California,

USA

Running title: Detection of directional and stabilizing mate preference

Emails:

a: derek.roff@ucr.edu

b: daphne.fairbairn@ucr.edu

c: ayp106@gmail.com

ABSTRACT

Estimation of mating preferences is a prerequisite for understanding how sexual selection through mate choice shapes both mating systems and sexual dimorphisms. Most studies of mating preferences assay mate choice using either a no choice or a binary choice design. Binary choice trials typically employ either an artificial signal or some fixed difference (e.g. color or size) between the signaling individuals. Although statistically more powerful than no choice designs, such experiments cannot be used to detect stabilizing preference. Further, the use of artificial signals is problematic because signal components tend to be varied in isolation, and hence do not reflect natural variation. Here we present a new method that uses natural variation among individuals in choice trials to determine if mating preference is absent, directional and/or stabilizing. The protocol is tested using simulation and shown to be robust to the preference function, to have the required statistical power, to be unbiased in almost all cases, and to give confidence regions that modestly overestimate the desired 95% criterion. We demonstrate the use of the method with data from mate choice trials of the sand cricket, *Gryllus firmus*. Software to apply this new approach is provided in Dryad.

keywords: mate choice, preference, directional preference, stabilizing preference

INTRODUCTION

Mating preferences act as significant selective agents on traits that influence the probability of being chosen (Jennions and Petrie 1997, Jennions et al. 2012). Such selection may be directional, stabilizing or a combination of both. Preferences for traits that function primarily in species recognition are likely to exert stabilizing selection because directional preference could potentially drive the preferred trait values into regions in which they might be confused with trait values found in other species (Ferreira and Ferguson 2002, Zuk et al. 2008). Directional preferences (i.e., preferences for increasing or for decreasing values of the preferred trait) are more likely to be found for traits such as the amount of time spent displaying, duration of individual displays, or the intensity of display (e.g., call loudness, color saturation or brightness: Jang and Greenfield 1998, Bentsen et al. 2006). Such traits are less likely to function in species recognition and more likely to vary with the health and condition of the signaler. Therefore, the form of selection acting on traits in the chosen sex is likely to vary among the target traits (hereafter, 'preferred traits'). The extent to which preferred traits are able to evolve independently in response to this selection depends upon the covariation among traits. For example, if a trait that is essential for species recognition is highly correlated with a trait that is initially subject to directional preference, the latter trait may be constrained in its response to directional selection because of the indirect effects of stabilizing selection on the correlated trait. Preference for this trait may ultimately shift to being stabilizing if potential mates with extreme trait values are confused with another species. Thus we might expect that directional selection through mate choice will primarily target traits that are not constrained by overlap with trait values in other species or by strong correlations with such traits.

Determining which traits are preferred and how selection through mate choice is acting on these traits is clearly a prerequisite for understanding of how both preferences and preferred traits evolve. For many species, traits likely to be assessed by the choosey sex (hereafter denoted as females for convenience) are readily apparent: for example, in orthopterans and anurans components of the call song would be a prime candidates. Call songs can differ along many dimensions, such as loudness, carrier frequency, repetition rate etc. Some of these characteristics may serve primarily as species recognition signals, others as indices of mate quality, and others are signals that females may prefer for no obvious reason (Bentsen et al. 2006).

Given that a number of traits or trait components may be preferred to a greater or lesser degree, the initial goal of most investigations of mate choice has been to identify which traits are preferred and if possible, measure the strength of the preference. The general approach has been to conduct experiments in which individual components are isolated and one or more measures of preference taken. This may be done using a no choice or a choice design. An excellent example of the no choice design is the study of female preference for the male calling song in the cricket *Gryllus integer* (Gray and Cade 1999). In this study artificially generated call songs that differed in pulses per trill were broadcast to a female and her response measured using a Kugel apparatus. Female preference was defined as the highest response elicited from the female. The experiment showed that females preferred an intermediate number of pulses per trill, clearly demonstrating stabilizing preference for this trait when varied independently of other calling traits. In a choice design, several potential mates or the signals from several potential mates are presented simultaneously to an individual of the choosing sex. The potential mates differ with respect to some character or characters that the choosing sex could use to discriminate amongst them. An example of this approach is the analysis of female preference for acoustic traits in the gray tree frog, *Hyla versicolor*. Each female was presented with two male calls from two speakers and her preference assessed through her orientation to the speakers (Gerhardt et al. 2000). Females were observed to prefer long duration calls with the strength of preference depending on whether the calls were below or above the average call duration (Gerhardt et al. 2000).

A meta-analysis of experiments using both choice and no choice designs showed that choice trials were more likely to detect significant preferences than no choice trials (Dougherty and Shuker 2015). We reviewed the designs of published choice experiments (supplemental material section 1) and discovered that in most (73%), the choosing sex (usually females) was presented with a choice of potential mates that differed on the basis of some *a priori* recognizable character such as condition or morphology. The remainder of choice studies used artificial signals (e.g. physical models or computer generated acoustical signals). In either case, the great majority of choice experiments (80%) parsed the trait distributions of the chosen sex into very few (usually only two) predetermined, discrete categories. In such cases statistical analysis was based on comparison of the mean or median trait values in chosen versus not-chosen individuals using categorical methods such as chi-square or Wilcoxon matched pairs. Use of such a limited number of trait categories may increase the power to detect significant preferences, but it precludes discrimination of directional from stabilizing preference as curvature cannot be established.

The challenge addressed in the present study is to develop a method based on choice experiments using natural signals that will allow researchers to determine not just which traits are preferred but also the form of the preference function (directional, stabilizing or both). Our method makes use of naturally occurring variation in the signal and thus does not suffer from problems of lack of variation in the signal, variation of only one component in isolation, or an unnatural mixture of components. Using our method, researchers will be able to analyze multiple traits within a single choice experiment and statistically test for the presence of directional and stabilizing selection. We first use a simulation model of a two choice experimental design to describe the new method and assess its statistical properties for hypothesis testing and parameter estimation. We then provide a ‘real world’ illustration of the method using data from a binary choice experiment using the sand cricket, *Gryllus firmus.* A “user-friendly” computer program running in R(2016) is available on the Dryad site (http://datadryad.org).

METHODS

Modeling a Mate Choice Experiment

We consider experiments that consist of two males presented to a female in which her choice is scored on a binary (0=rejected, 1=accepted) or continuous (e.g. number of approaches to a male) scale, and the distribution of female choice is used to determine if the female preference function is directional and/or stabilizing. The proposed method also estimates the population mean female preference.

Based on sample sizes typical of published choice experiments (see supplemental material, section 1) we set the sample size at 100 trials, each trial using a different female and a different pair of males. Male trait values were drawn at random from a normal distribution with a mean of 10 and a standard deviation of 1 or 3. In any trial one male was designated at random as the focal male and the other the non-focal male, their trait values being denoted as x_F and x_{NF}, respectively. Without loss of generality we assumed male trait values were always positive (this can always be made so by a suitable transformation). When choice is measured on a continuous scale, female preference is measured by either the relative preference, $R_P = n_F / (n_F + n_{NF})$, or the difference in preferences, $D_P = n_F - n_{NF}$, where n_F, n_{NF} are the measures of the female attraction, such as number of approaches, to the focal and non-focal males, respectively.

In trials with directional preference, we set the probability of the *i*th female selecting the focal male of the *i*th pair as

$$P_{F,i} = x_{F,i} / (x_{F,i} + x_{NF,i}) \quad (1)$$

Thus, females preferred the male with the larger trait value (i.e., the preference was directional and positive).

For stabilizing preference the mean preference value for females was set either at 9 or 10. When the male and female means differed (i.e., when the female mean was 9) the preference function has both directional and stabilizing components. We used two different standard deviations (1 and 3) for both female preference and male trait values, because variances have been shown to influence the evolutionary trajectories of both preferences and preferred traits (Roff and Fairbairn 2014). In total, eight combinations of means and standard deviations were generated, including combinations in which the variance in female preference was less than that of the male trait and combinations in which the variance in female preference was substantially larger than the variance in male trait values.

Lande (1981) modelled stabilizing female preference as a Gaussian type function, which, under the present terminology can be written for the *i*th male (whether focal or non-focal) as,

$$P_i = e^{-\frac{1}{2}\left(\frac{x_i - y_i}{\nu}\right)^2} \tag{2}$$

where x_i is the male trait, y_i is the female trait and ν is the width of the tolerance function for female choice. Female choosiness decreases as ν increases. Lande (1981) assumed that the male and female traits (*x,y*) were normally distributed with means and variances that could be different. He assumed that ν was a constant. Under the latter assumption, ν and the constant ½ can be absorbed into the male female traits creating the simplified model

$$P_i = e^{-(x_i - y_i)^2} \tag{3}$$

which we shall refer to as the Exponential difference, or ED, function. This type of equation was chosen by Lande (1981) for mathematical convenience rather than the presence of empirical data indicating this as the appropriate function.

An alternative function that is equally plausible is that female preference declines as a function of the absolute difference between her preference and the male trait,

$$P_i = f\left(\left|x_i - y_i\right|\right) \tag{4}$$

For the ED model the probability of the focal male being chosen, $P_{F,i}$ was given by

$$P_{F,i} = \frac{e^{-\left(x_{F,i}-y_i\right)^2}}{e^{-\left(x_{F,i}-y_i\right)^2} + e^{-\left(x_{NF,i}-y_i\right)^2}} \tag{5}$$

For the second preference function, hereafter the AD function, we assumed that the probability of a female selecting the focal male was directly proportional to the distance from the preferred female value relative to the non-focal male,

$$P_{F,i} = 1 - \frac{\left|x_{F,i} - y_i\right|}{\left|x_{F,i} - y_i\right| + \left|x_{NF,i} - y_i\right|} \tag{6}$$

The above preference functions are probability functions. Simple algorithms based on these functions were used to generate the observed female choices in each of our simulation trials (see supplemental material, section 2). For the combinations of parameter values used in the present simulations the focal male is generally either strongly preferred or strongly rejected under the ED function, whereas, under the AD function the strength of preference is centered about 0.5 (see supplemental material 3). We included both of the preference functions to test the robustness of our estimation procedure. The method estimates the mean female preference, assuming a normal distribution of female preferences.

Quantifying Preference

Directional Preference

Under directional preference we would expect that as the difference between the trait values of the two potential mates increases so will the preference in one direction. We can therefore predict that the probability of a female choosing the focal male over the non-focal male, P_F, will be a monotonic function of the focal and non-focal male traits: we consider here both the relative trait value of the focal male, $x_{F,i}/\left(x_{F,i}+x_{NF,i}\right)$, and the difference in trait values, $x_{F,i}-x_{NF,i}$. A general test for directional preference is a regression of female preference on either of these male metrics. For binary data the estimated female preference for the focal male, $\hat{P}_{F,i}$ is measured as 1 for the chosen male and 0 for the other male. With continuous data the estimated female preference is measured as $\hat{P}_{F,i}=n_{F,i}/\left(n_{F,i}+n_{NF,i}\right)$, where $n_{F,i}$ is the index of female preference for the focal male and $n_{NF,i}$ her preference for the non-focal male: for example, $n_{F,i}$ could be the number of times the female approached the focal male. If female choice is binary (0,1) a logistic regression is appropriate. For the continuous case the relationship may not be linear and hence a non-linear regression or suitable transformation may be necessary.

In real world experiments, if the pair of males differ by some standard category (e.g. well-fed vs under-fed, long-winged vs short-winged) then one of these categories can be designated the focal male and the other the non-focal male. In the case where there is no distinguishing category and data are binary, males should be randomly assigned to be focal or non-focal males such that approximately 50% of focal males are preferred and 50% are not preferred by the female.

Stabilizing Preference

We first consider an experiment in which the data are 0,1. Suppose the mean female preference is μ. Consider the metric

$$d_i = \left|x_{F,i} - \mu\right| - \left|x_{NF,i} - \mu\right| \qquad (7)$$

Under the hypothesis of stabilizing preference, negative values of d_i should be associated with the focal male being preferred (because the absolute deviation of $x_{F,i}$ from μ is less than that of $x_{NF,i}$ from μ) and positive values with the non-focal male being preferred. For each value of d_i we compute the number of correct predictions made using the simple formula "focal male if $\left|x_{F,i} - \mu\right| < \left|x_{NF,i} - \mu\right|$, otherwise non-focal male". The best estimate of the mean female preference is the value that gives the greatest number of correct predictions, N_{max}.

The above procedure locates the best estimate of mean female preference, but it does not specify that the prediction is significantly better than chance. A suitable test is a logistic regression of female choice (0,1) on $\left|x_{F,i} - \hat{\mu}\right| - \left|x_{NF,i} - \hat{\mu}\right|$, where $\hat{\mu}$ is the best estimate of female preference. A statistically significant regression indicates significant stabilizing preference. This test is one-tailed as the slope must be negative under stabilizing preference.

When choice is a continuously varying measure we defined the independent variable as

$$X_i = 1 - \frac{\left|x_{F,i} - \mu\right|}{\left|x_{F,i} - \mu\right| + \left|x_{NF,i} - \mu\right|} \tag{8}$$

which has the useful property $0 \le X_i \le 1$ that avoids possible extreme values. The metric d_i (equation (7)) could also be used but we have not examined its statistical properties for the continuous case. Under the assumption of stabilizing preference the probability of a female choosing the focal male will be a monotonically increasing function of X_i. For the present analysis we assumed a linear function. To obtain an estimate of μ we proceed in three steps:

1) First we select a trial value of μ, say μ_T and then for each pair of males we use μ_T to estimate X_i.
2) Second, we regress the estimated female preference, $\hat{P}_{F,i}$, for the focal male on X. As noted above, in the present case we modeled a linear function and therefore we used a linear regression: in an empirical study the relationship might not be so and a transformation or alternate regression function may be required.
3) The final step is to vary μ_T and for each such value calculate the variance accounted for by the regression of μ_T on X: the best estimate of mean female preference, $\hat{\mu}$, is that value which gives the highest r^2, subject to the constraint that the slope of the regression of μ_T on X is positive. To be

statistically significant the probability associated with the regression must be less than 0.05. To distinguish negative from positive slopes, we retain the sign of the slope in the r^2 and designate this statistic as r_S^2 or signed r^2. . Graphically, we plot r_S^2 on μ_T.

The above procedures provide a statistical test of stabilizing female preference and an approximate estimate of the mean female preference, but not an estimate of the standard error. To obtain an estimate of the standard error we used the bootstrap: samples were drawn with replacement from the original data set and the estimate, $\hat{\mu}$, computed. A total of 1000 bootstraps were drawn and the estimate and its standard error computed as the mean and standard deviation of the 1000 bootstraps (Roff 2006).

Empirical Example

We provide an empirical example of the proposed method based on a two choice experiment with sand crickets (*Gryllus firmus*) in which each female was given the choice of two seven-day old males drawn at random from a recently established laboratory population. The data consist of the male song components and a measure of female attraction to these in 93 binary choice trials. These data were obtained using a modification of the T-maze design reported in Crnokrak and Roff (1995) by moving the tubes of the maze down so that they entered the buckets containing the males close to the bottom. This allowed the females to move freely into and out of the buckets. As in the original mazes, the males were held in containers within the buckets, preventing physical contact between the sexes. We used motion detectors set at the entrances to the buckets containing the males to monitor the activity of the females. Female preference for the focal male was estimated as the number of times the female tripped the motion detector on the focal male side divided by the total number of times both motion detectors were tripped. A computer monitoring system checked each male every second and recorded if the male was calling and the volume of the call. These data provided an estimate of time spent calling and the mean volume. We also obtained samples of the call songs using USB digital audio recorders. Song components were analyzed using Raven Pro (The Cornell Lab of Ornithology) with the following components being measured: pulses per chirp, pulse length, pulse rate, pulse period, chirp length, chirp rate and frequency. Pulse rate was defined as the rate of pulses within a chirp and chirp rate as the number of chirps per unit time within a singing bout, defined as a set of chirps less than 0.5 seconds apart. From this set we chose three components that illustrate the situations of "no preference", "directional preference" and "stabilizing preference". A complete analysis of the *G. firmus* will be presented elsewhere. We present data on three components that illustrate the situations of "no preference", "directional preference" and "stabilizing preference". The predicted preference for the focal male was set as $x_{F,i}/\left(x_{F,i}+x_{NF,i}\right)$ for directional preference and as given in equation (8) for stabilizing preference.

RESULTS

The Simulation model

General Patterns

The protocol for determining stabilizing preference can be graphically displayed by plotting r_S^2 (the signed r^2) for each trial value of μ, designated μ_T, on this trial value (left columns, Fig. 1). Directional preference is displayed by plotting preference for the focal male (P_F) on the relative value of the focal male (right columns, Fig. 1). Significant stabilizing preference shows two patterns of variation. First, if both stabilizing and directional preference are present, the function r_S^2 on μ_T shows both a peak and a rise above the critical significance value between the peak and one of the extrema of μ_T (top row, Fig. 1). In the case illustrated, the female mean preference was less than the mean male trait value (9 vs 10) and hence the highest value of r_S^2 lies above the significance value to the left of the peak. The second pattern of stabilizing preference is one in which there is a single peak and no significant directional preference (middle row, Fig. 1), which occurs when the mean female preference and mean male trait value coincide or the variance in female preference or male trait value is relatively large.

With only positive directional preference, the function of r_S^2 on μ_T shows an increasing S-shaped function (bottom row, Fig. 1), with r_S^2 plateauing and no single value of μ_T being "best".

Statistical Properties of Estimators

Three primary statistical properties that an estimator must satisfy are sufficient power, acceptably small bias and accurate standard errors (i.e. true mean is enclosed within ±2 standard errors 95% of the time). To determine the statistical power of the above procedures we ran 1,000 simulations per combination with 100 male pairs per simulation. In addition we ran simulations with directional preference only and also with random female preference. The analysis proceeds in two steps: the first is an hypothesis testing step that stabilizing preference exists and the second is parameter estimation. If the first step is passed (i.e. stabilizing preference is not rejected) then we move to the second step of parameter estimation. If the test for stabilizing preference is not significant we could still estimate the parameters of the stabilizing preference but this estimate is predicated on the unverified assumption that stabilizing preference is present. In the case of an experiment in which a treatment is applied it is permissible to estimate the effect of treatment even if it is non-significant because the difference between treatments is quantitative. The present case is fundamentally different as it requires the addition of an unverified assumption. Therefore, while we retained all simulations in the test for statistical power, to test for bias and the efficacy of the bootstrap, we considered only those simulations in which significant stabilizing preference was found. We continued running simulations until 1000 data sets were created in which significant stabilizing preference was found.

We present the statistical analyses for the continuous data here and the 0,1 data in the supplemental material (section 4). Tests were done using both $x_F / \left(x_F + x_{NF} \right)$ and $x_F - x_{NF}$: as they gave the same qualitative performance we report the results only for the former metric. Results for the ED and AD models were very similar with respect to both statistical power, bias and standard error estimates (see below).

Statistical Power

The first criterion for accepting any model is that it must be significantly different from the null model. A generally accepted desirable level for power (i.e. probability of rejecting the null hypothesis when the alternate hypothesis is true) is 80% (Cohen 1988). The method did an excellent job of statistically detecting pure directional preference. The statistical power to detect directional preference when females showed only directional preference was 100%. Type I error rates were also appropriate: directional preference was detected in only 5% of cases when female preference was random.

The method also did a good job of statistically detecting stabilizing preference in the absence of directional preference, although type I error rates were slightly higher and power was slightly lower than for pure directional preference. With random female preference apparent stabilizing preference was inferred in 6% of runs with SD of 1, and 10% of runs with SD of 3, which is slightly greater than the required 5% and indicates that marginally significant values in tests of stabilizing preference should be treated with circumspection.

The statistical power of the method for detecting stabilizing preference in the simulations with only stabilizing preference ranged from 40% to 100%, with 6 of the 8 combinations exceeding 90% and all but one exceeding 50% (lower right panel, Fig. 2). Power exceeds 90% if the variance in preference is equal or greater than the variance in male trait values, but is much lower (40% - 60%) when the male trait variance greatly exceeds the variance in female preference (ratio of 1:3 in lower right panel of Fig. 2).

With stabilizing preference and a female mean of 9 and a male mean of 10 there is both stabilizing and directional preference. The power to detect the stabilizing component of preference is high and similar to the power to detect pure stabilizing preference, being over 80% in 7 of 8 cases (bottom left panel, Fig. 2). When both male and female variances were 1, the power to also detect directional preference was 100% (top left panel, Fig. 2). However, for other combinations of variances, statistical power to detect the directional component of preference was lower. Therefore, when both stabilizing and directional preference occur the power to detect the latter is low unless both the male and female variances are small relative to the difference between the means: for example, in the present case directional preference was readily detected when the CV in females was 11% and in males 10% but not when the CV in females was 30% and in males 33%.

When female and male means are the same then a priori we might expect that there should be no directional preference. However, in the simulations directional preference was incorrectly detected in more than 5% of runs (top right panel, Fig. 2). Thus, the probability of a Type I statistical error is higher than ideal. Positive directional preference was more often detected than negative directional preference. We have not been able to demonstrate this effect analytically, but simulations using the simplified model of stabilizing female preference $P_F = |x_F - \mu| / \left(|x_F - \mu| + |x_{NF} - \mu| \right)$ on $x_F / \left(x_F + x_{NF} \right)$ have verified that the effect is real and depends on the relative variances, i.e. stabilizing preference in which the female mean preference matches the mean male trait may generate both stabilizing and directional preference. The presence of x_F, x_{NF} in both the predictor and response variables suggests that at least part of the correlation is spurious. If the significance level is set at 0.01 rather than 0.05, the probability of declaring significant directional preference falls close to 0.05 for the continuous data (mean = 0.056) and considerably less than 0.05 for 0,1 data (mean=0.017).

Bias and Standard error estimates

Bias was calculated as the estimated value minus the true value of the mean female preference divided by the actual value. There was no indication of a persistent bias in the estimate of the mean. Standard errors were given by the standard deviation of the bootstrap values. The two largest biases were -13% (AD model) and -17% (ED model) at combinations which occurred when the female mean was 9 and both variances equal to one. In all other cases the magnitude of percent bias was less than 5% and in all combinations the estimated confidence interval was 95% or greater. The standard error estimates were conservative in being generally larger than required.

Summary of Simulation Results

Overall, the simulation results indicate that the proposed methods for estimating stabilizing preference and pure directional preference are satisfactory and robust to the type of preference function. When preference is purely directional the statistical power is 100%. When selection is purely stabilizing, the statistical power is generally above 80% and false positives are less than 10%. In our simulations, the power was low only if the variance in male trait values greatly exceeded the variance in female preference, a circumstance that we shall argue is unlikely to occur in natural situations (see below). When preference is stabilizing but the female mean differs from the male mean, both stabilizing and directional preference should be detected. In these cases, the power to detect stabilizing preference remains high, but both type I and type II error rates for detecting the directional component of preference are higher than ideal. Therefore, when significant stabilizing preference is detected, conclusions about the concurrent presence of directional preference should be made with caution.

The method also does a good job of parameter estimation. When preference is stabilizing, mean female preference is estimated accurately and with negligible bias. The standard errors of mean female preference are only slightly overestimated and potentially troublesome bias only occurred in two combinations. Even in the latter cases the confidence limits enclosed the true value at the required 95% level.

Empirical Example

Single trait analysis

Figures 3, 4 and 5 illustrate the three general results from the analysis: no significant preference (Fig. 3), directional preference (Fig. 4), and stabilizing preference with the possibility of a directional component (Fig.5). Females showed no preference for any carrier frequency within the range observed (Fig.3). There was no evidence of any directional preference (P=0.480) and in no trial did mean female preference value cause the correlation between predicted and observed values to exceed the critical 5% level. The bootstrap shows the mean female preference close to the population mean, which is what we would expect if the trait was a species-recognition character under stabilizing preference. However in this case such a preference, if it exists, is too weak to be detected given the present sample size.

In contrast, there is strong evidence of female preference being correlated with time spent calling by the male (Fig. 4). There was a highly significant regression between observed and predicted preference values (*P*=0.0008) indicating directional preference (top panel, Fig. 4). This directionality is also indicated by the stabilizing preference test, which shows a continuous rise that exceeds the critical 5% level (compare this plot with the bottom left plot in Fig 1).

The third example shows significant statistical evidence of a trait, chirp rate, under stabilizing preference with a directional component (Fig.5). There was a highly significant directional component (*P*=0.005) and a clear unimodal stabilizing function with a probability of the signed r^2 at its peak of 0.0012 (compare to the top left panel in Fig. 1). The bootstrap estimate was 2.28 (SE=0.28).

DISCUSSION

The simulation models demonstrated that the proposed method can successfully detect both directional and stabilizing preference for sample sizes that are well within logistical limits. The statistical properties of the test are generally good, with high power to detect both stabilizing and pure directional preference. However, we did discover two areas of potential statistical error indicating that caution should be taken when drawing some statistical conclusions. The first problem area concerns the detection of directional preference when stabilizing preference is also present. In this case, the power of the method to detect the stabilizing component is good, but its power to also detect the directional component is relatively low unless the variances in both male trait values and female preferences are small. Thus, if the data follow a pattern similar to that illustrated in the top panels of Fig. 1 or in Fig. 5, but only the stabilizing component of preference is statistically significant, researchers should be cautious (i.e., tentative) in concluding the absence of directional preference unless $P>>0.05$. The opposite statistical problem also occurs: the type I error of incorrectly detecting of directional preference when preference is purely stabilizing occurs at a rate greater than 5%. To alleviate this problem, researchers should also be cautious in concluding that directional preference is present unless the probability value is ≤ 0.01.

The second problem area concerns the low power of the test to detect stabilizing selection when the variance in male trait values greatly exceeds the variance in female preference (SD ratio of 1:3 in our simulations). This combination of variances also led to the highest type I error rates for detecting directional selection in the presence of stabilizing selection. Fortunately, this circumstance seems unlikely to be relevant in nature. Although the variances of male traits and female preferences lie at the core of models of the evolution of mate choice (Lande 1981; Roff and Fairbairn 2014), very few empirical studies have estimated the variance in female preference. Nevertheless, the evidence that does exist indicates relatively little difference in the variances: a range from 0.8 to 1.6 in the ratio of the variance in female preference to the variance in the male trait (Roff and Fairbairn 2014). Further indication that the difference in variances is not great comes from the observation that the estimated stabilizing preference function generally seems to straddle and slightly exceed the distribution of the preferred trait (e.g. Ritchie et al. 2001; Brooks et al. 2005; Bentsen et al. 2006; Steele et al. 2011; Moreno-Gomez et al. 2015).

We applied the approach to empirical data on three song components of the sand cricket and with the approach identified a lack of preference, directional preference and stabilizing preference. For purposes of illustration, we have treated these traits as independent, but in many cases, different components of male displays are likely to be correlated with each other. Our method can be extended to sets of correlated traits in two possible ways. One method would be to initially analyze each trait independently, as we have done in our examples, and then enter all statistically significant traits into a single stepwise regression. A complementary approach in cases where the male traits are correlated would be to use principal components analysis (PCA) and analyze female preference for the resulting principal components.

The empirical results of the present analysis highlight the potential problems of restricting assays of female choice to single variables as is likely to be the case when artificial signals are used. Models for the evolution of preferential mating systems generally assume preferences for single traits. However, there is no theoretical bar to the evolution of multiple sexual preferences (Pomiankowski and Iwasa 1993) and evidence is accumulating that different females (or males) within the same population may weigh multiple traits differently in making their choice (e.g., Orthoptera: Hedrick and Weber 1998; Olvido and Wagner 2004; Bailey 2008; Verburgt et al. 2008; Stout et al. 2010;

Fowler-Finn and Rodriguez 2012; Hedrick and Kortet 2012; anurans: Murphy and Gerhardt 2000; fish: Brooks and Endler 2001; Brooks 2002; Morris et al. 2003; Pierotti et al. 2009; Labonne et al. 2009; Lehtonen et al. 2010; birds: Coleman et al. 2004; David and Cezilly 2011; humans: Zietsch et al. 2012). This is another reason for testing multiple males per female and for measuring multiple traits per male. Such variation among females would tend to eliminate evidence of overall directional or stabilizing preference on any single trait. By extension, finding strong directional preference for a particular trait is itself evidence that this is a trait that the majority of females find attractive (or repellant).

As a general recommendation, tests of female preference should be made relevant to the natural history of the species. Analysis of no choice trials is relatively simple (e.g. linear and quadratic regression comparisons), but the meta-analysis of Dougherty and Shuker (2015) showed that this protocol has less statistical power to detect either directional of stabilizing preference than choice experiments. Nevertheless, as noted by Dougherty and Shuker, no choice trials may be realistic for species that meet potential mates sequentially, and are hence preferable in these cases. We might be similarly critical of choice trials if they use a number of males that differs from the number typically encountered by a female in the wild. As most choice trials involve only two males it is legitimate to ask if this represents a normal situation. A survey of published estimates of the number of potential mates assessed by females in the wild showed that females assessed a mean of 4.5, and a median of 2.9, potential mates before making their choice (Roff and Fairbairn 2014). Thus, although the use of two males may not be unrealistic for many species, it could greatly underestimate the sampling strategies of females in others. We have developed the method for two-choice trials as this is the one most commonly used: however, the approach can be readily extended to multi-choice experiments.

The experimental and statistical protocols advanced and tested in this paper overcome the problem of limitations of choice experiments in that outcomes are not restricted to categorical designations (e.g. big vs small) nor to stimuli that may be warped versions of the natural stimulus (e.g. artificially constructed songs that differ only in frequency). In addition, both directional and stabilizing preferences can be statistically tested both with respect to individual components and to more complex stimuli using, for example, either multiple regression or principal components.

ACKNOWLEDGEMENTS

We are very grateful for the technical assistance of Jaime Guzman and Taylor Vanden Bossche, without whom this research would have been immensely more difficult. This research was supported by NSF grant IOS-1353463.

REFERENCES

Bailey NW, 2008. Love will tear you apart: different components of female choice exert contrasting selection pressures on male field crickets. Behav Ecol. 19:960-966.

Bentsen CL, Hunt J, Jennions MD, Brooks R, 2006. Complex multivariate sexual selection on male acoustic signaling in a wild population of *Teleogryllus commodus*. Am Nat. 167:E102-E116.

Brooks R, 2002. Variation in female mate choice within guppy populations: population divergence, multiple ornaments and the maintenance of polymorphism. Genetica 116:343-358.

Brooks R, Endler JA, 2001. Female guppies agree to differ: Phenotypic and genetic variation in mate-choice behavior and the consequences for sexual selection. Evolution 55:1644-1655.

Brooks R, Hunt J, Blows MW, Smith MJ, Bussiere LF, Jennions MD, 2005. Experimental evidence for multivariate stabilizing sexual selection. Evolution 59:871-880.

Cohen J, 1988. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Coleman SW, Patricelli GL, Borgia G, 2004. Variable female preferences drive complex male displays. Nature 428:742-745.

Crnokrak P, Roff DA, 1995. Fitness differences associated with calling behaviour in the two wing morphs of male sand crickets, *Gryllus firmus*. Anim. Behav. 50:1475-1481.

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

Dougherty LR, Shuker DM, 2015. The effect of experimental design on the measurement of mate choice: a meta-analysis. Behav Ecol. 26:311-319.

Ferreira M, Ferguson JWH, 2002. Geographic variation in the calling song of the field cricket *Gryllus bimaculatus* (Orthoptera : Gryllidae) and its relevance to mate recognition and mate choice. J Zool 257:163-170

Fowler-Finn KD, Rodriguez RL, 2012. The evolution of experience-mediated plasticity in mate preferences. J Evol Biol. 25:1855-1863.

Gerhardt HC, Tanner SD, Corrigan CM, Walton HC, 2000. Female preference functions based on call duration in the gray tree frog (*Hyla versicolor*). Behav Ecol. 11:663-669.

Gray DA, Cade WH, 1999. Quantitative genetics of sexual selection in the field cricket, *Gryllus integer*. Evolution 53:848-854.

Hedrick A, Weber T, 1998. Variance in female responses to the fine structure of male song in the field cricket, *Gryllus integer*. Behav Ecol. 9:582-591.

Hedrick AV, Kortet R, 2012. Effects of body size on selectivity for mating cues in different sensory modalities. Biol J Linn Soc. 105:160-168.

Jang YW, Greenfield MD, 1998. Absolute versus relative measurements of sexual selection: Assessing the contributions of ultrasonic signal characters to mate attraction in lesser wax moths, *Achroia grisella* (Lepidoptera : Pyralidae). Evolution 52:1383-1393.

Jennions MD, Petrie M, 1997. Variation in mate choice and mating preferences: A review of causes and consequences. Biol Rev. 72:283-327.

Jennions MD, Kahn AT, Kelly CD, Kokko H, 2012. Meta-analysis and sexual selection: past studies and future possibilities. Evol Ecol. 26:1119-1151.

Labonne J, Augery M, Parade M, Brinkert S, Prevost E, Heland M, Beall E, 2009. Female preference for male body size in brown trout, *Salmo trutta*: is big still fashionable? Anim Behav. 77:129-137

Lande R, 1981. Models of speciation by sexual selection on polygenic traits. PNAS 78:3721-3725.

Lehtonen TK, Wong BBM, Lindstrom K, 2010. Fluctuating mate preferences in a marine fish. Biol Lett 6:21-23.

Moreno-Gómez FN, Bacigalupe LD, Silva-Escobar AA, Soto-Gamboa M, 2015. Female and male phonotactic responses and the potential effect of sexual selection on the advertisement calls of a frog. Anim Behav. 104:79-86.

Morris MR, Nicoletto PF, Hesselman E, 2003. A polymorphism in female preference for a polymorphic male trait in the swordtail fish *Xiphophorus cortezi*. Anim Behav. 65:45-52.

Murphy CG, Gerhardt HC, 2000. Mating preference functions of individual female barking treefrogs, *Hyla gratiosa*, for two properties of male advertisement calls. Evolution 54:660-669.

Olvido AE, Wagner WE, 2004. Signal components, acoustic preference functions and sexual selection in a cricket. Biol J Linn Soc. 83:461-472.

Pierotti MER, Martin-Fernandez JA, Seehausen O, 2009. Mapping individual variation in male mating preference space: multiple choice in a color polymorphic cichlid fish. Evolution 63:2372-2388.

Pomiankowski A, Iwasa Y, 1993. Evolution of multiple sexual preferences by fisher runaway process of sexual selection. Proc Roy Soc B-Biol Sci 253:173-181.

R 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ritchie MG, Saarikettu M, Livingstone S, Hoikkala A, 2001. Characterization of female preference functions for *Drosophila montana* courtship song and a test of the temperature coupling hypothesis. Evolution 55:721-727.

Roff DA, 2006. Introduction to Computer-Intensive Methods of Data Analysis in Biology. Cambridge: Cambridge University Press.

Roff DA, Fairbairn DJ, 2014. The evolution of phenotypes and genetic parameters under preferential mating. Ecol Evol 4:2759-2776

Steele DB, Siepielski AM, McPeek MA, 2011. Sexual selection and temporal phenotypic variation in a damselfly population. J Evol Biol 24:1517-1532.

Stout J, Navia B, Jeffery J, Samuel L, Hartwig L, Butlin A, Chung M, Wilson J, Dashner E, Atkins G, 2010. Plasticity of the phonotactic selectiveness of four species of chirping crickets (Gryllidae): Implications for call recognition. Physiol Ent. 35:99-116.

Verburgt L, Ferguson JWH, Weber T, 2008. Phonotactic response of female crickets on the Kramer treadmill: methodology, sensory and behavioural implications. J Comp Physiol a-Neuroethol Sens Neur and Behavl Physiol. 194:79-96.

Zietsch BP, Verweij KJH, Burri AV, 2012. Heritability of preferences for multiple cues of mate quality in humans. Evolution 66:1762-1772.

Zuk M, Rebar D, Scott SP, 2008. Courtship song is more variable than calling song in the field cricket *Teleogryllus oceanicus*. Anim Behav 76:1065-1071.

FIGURE CAPTIONS

Figure 1: Characteristic patterns for functions detecting stabilizing and directional preference. The left column shows the signed r^2 (r_S^2) as a function of the trial value of the mean female preference, μ_T. The dashed line shows the 5% probability value for the signed r^2: Values above this line have a probability of occurring by chance of less than 0.05. The right column shows the estimate of female preference for the focal male, $\hat{P}_F$, against the preference estimated as the relative value of the focal male trait. The solid line shows the fitted regression line.

Top row: Female preference is stabilizing with a mean of 9 and a standard deviation of 1, mean male value is at 10 and standard deviation of 1: this produces both significant stabilizing and directional preference. Middle row: Female preference is stabilizing with a mean at 9 and a standard deviation of 3, mean male value is at 10 with a standard deviation of 1: this results in significant stabilizing preference but not significant directional preference.

Bottom row: Significant directional preference but no stabilizing preference.

Figure 2: Proportion of times directional (top) and stabilizing (bottom) preference was detected when female preference was stabilizing with or without a directional component. Results for AD model shown in black, those for the ED model are in stippled white. The mean male trait value is 10 with left panels showing results for female mean=9, and right panels showing results for female mean=10. Note that when male and female means are the same then directional selection should be detected only 5% of times. Both directional and stabilizing selection should be detected in the left panels but only stabilizing in the right panels.

Figure 3. An example of a song trait, (Frequency) that shows no statistical indication of female preference.

Top panel: Regression of estimated preference for the focal male, P_F, on the preference predicted from the male trait values. Solid line=fitted regression.

Bottom panel: Signed r^2 on the trial mean female preference, μ_T. Solid red line indicates the fit at the trial female preference. Only values lying above the 5% line (dotted blue line) are significant. The dotted line shows the normal distribution based on the bootstrap values.

Figure 4. An example of a song trait, (Duration of time male called) that shows statistical indication of directional female preference.

Top panel: Regression of estimated preference for the focal male, P_F, on the preference predicted from the male trait values. Solid line=fitted regression.

Bottom panel: Signed r^2 on the trial mean female preference, μ_T. Solid red line indicates the fit at the trial female preference. Only values lying above the 5% line (dotted blue line) are significant.

Figure 5. An example of a song trait, (Chirp rate) that shows statistical indication of stabilizing female preference, with the possibility of a directional component.

Top panel: Regression of estimated preference for the focal male, P_F, on the preference predicted from the male trait values. Solid line=fitted regression.

Bottom panel: Signed r^2 on the trial mean female preference, μ_T. Solid red line indicates the fit at the trial female preference. Only values lying above the 5% line (dotted blue line) are significant. The dotted line shows the normal distribution based on the bootstrap values.

Figure 1: Characteristic patterns for functions detecting stabilizing and directional preference. The left column shows the signed r^2 (r_S^2) as a function of the trial value of the mean female preference, μ_T. The dashed line shows the 5% probability value for the signed r^2: Values above this line have a probability of occurring by chance of less than 0.05. The right column shows the estimate of female preference for the focal male, $\hat{P}_F$, against the preference estimated as the relative value of the focal male trait. The solid line shows the fitted regression line.

Top row: Female preference is stabilizing with a mean of 9 and a standard deviation of 1, mean male value is at 10 and standard deviation of 1: this produces both significant stabilizing and directional preference. Middle row: Female preference is stabilizing with a mean at 9 and a standard deviation of 3, mean male value is at 10 with a standard deviation of 1: this results in significant stabilizing preference but not significant directional preference. Bottom row: Significant directional preference but no stabilizing preference.

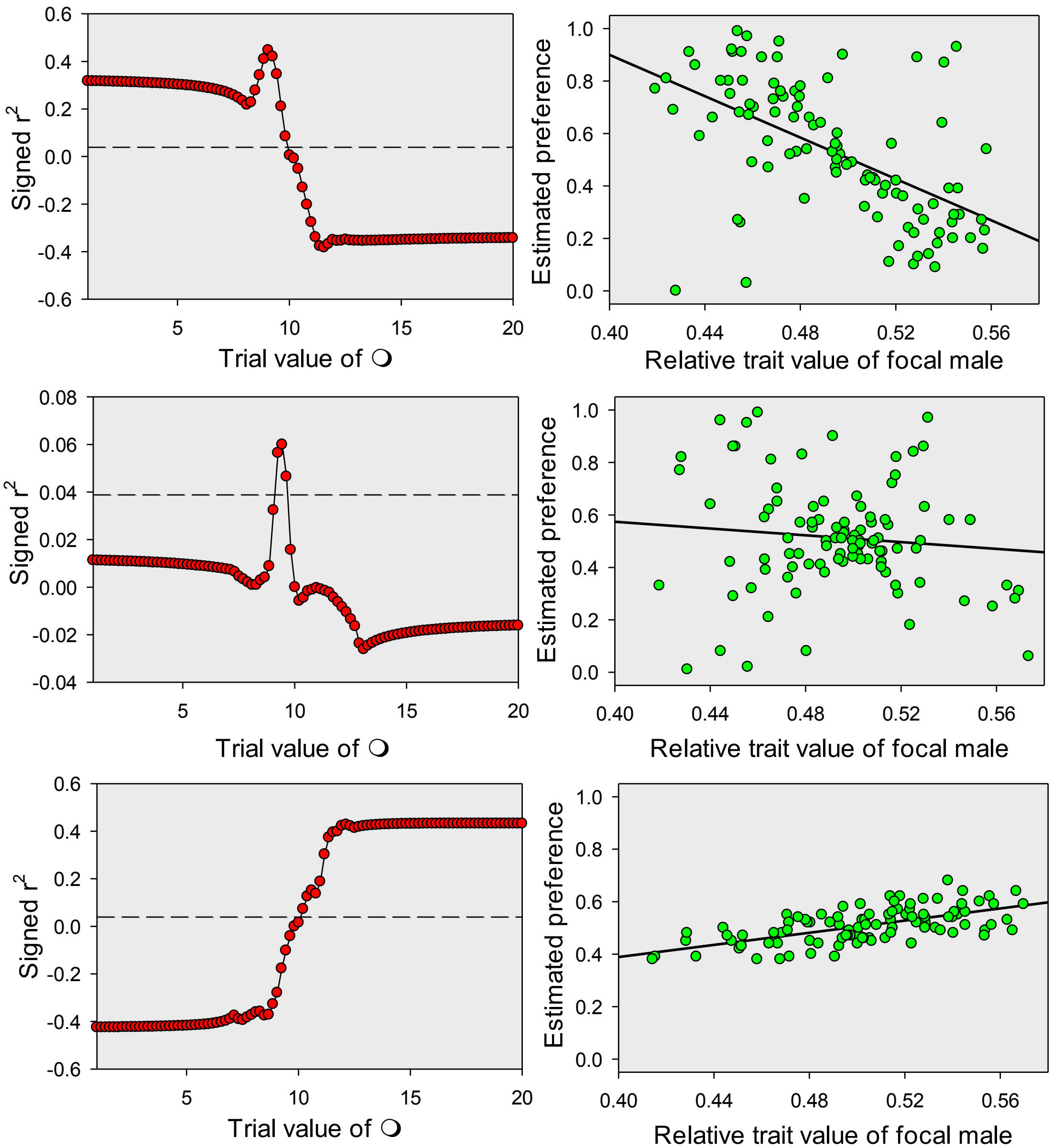

Signed r2
Trial value of ❍
Estimated preference
Relative trait value of focal male
Signed r2
Trial value of ❍
Estimated preference
Relative trait value of focal male
Signed r2
Trial value of ❍
Estimated preference
Relative trait value of focal male

Figure 2: Proportion of times directional (top) and stabilizing (bottom) preference was detected when female preference was stabilizing with or without a directional component. Results for AD model shown in purple and blue, those for the ED model are in Cyan and Green. The mean male trait value is 10 with left panels showing results for female mean=9, and right panels showing results for female mean=10. Note that when male and female means are the same then directional selection should be detected only 5% of times. Both directional and stabilizing selection should be detected in the left panels but only stabilizing in the right panels.

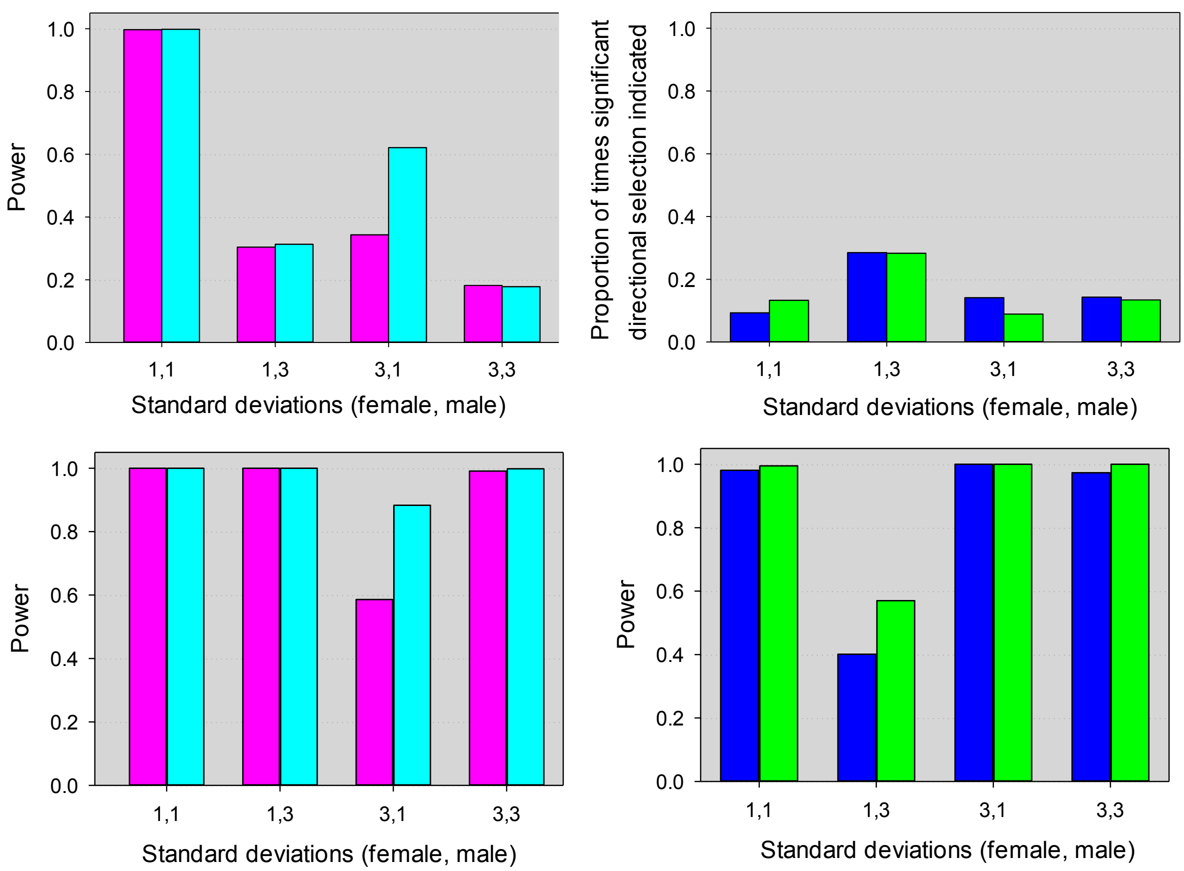


Figure 3. An example of a song trait, (Frequency) that shows no statistical indication of female preference.

Top panel: Regression of estimated preference for the focal male, P_F, on the preference predicted from the male trait values. Solid line=fitted regression.

Bottom panel: Signed r^2 on the trial mean female preference, μ_T. Solid red line indicates the fit at the trial female preference. Only values lying above the 5% line (dotted blue line) are significant. The dotted line shows the normal distribution based on the bootstrap values.

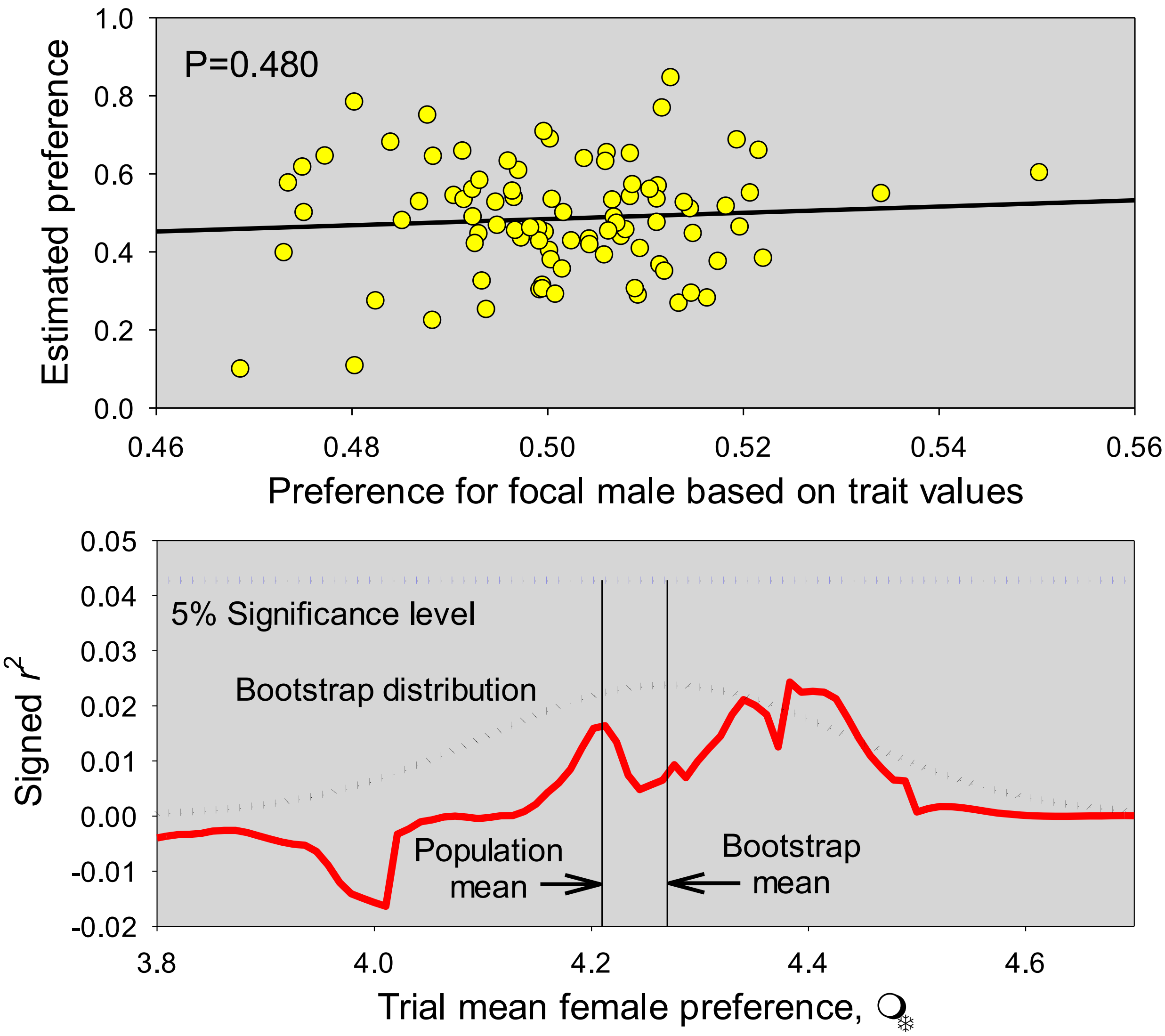


Figure 4. An example of a song trait, (Duration of time male called) that shows statistical indication of directional female preference.

Top panel: Regression of estimated preference for the focal male, P_F, on the preference predicted from the male trait values. Solid line=fitted regression.

Bottom panel: Signed r^2 on the trial mean female preference, μ_T. Solid red line indicates the fit at the trial female preference. Only values lying above the 5% line (dotted blue line) are significant.

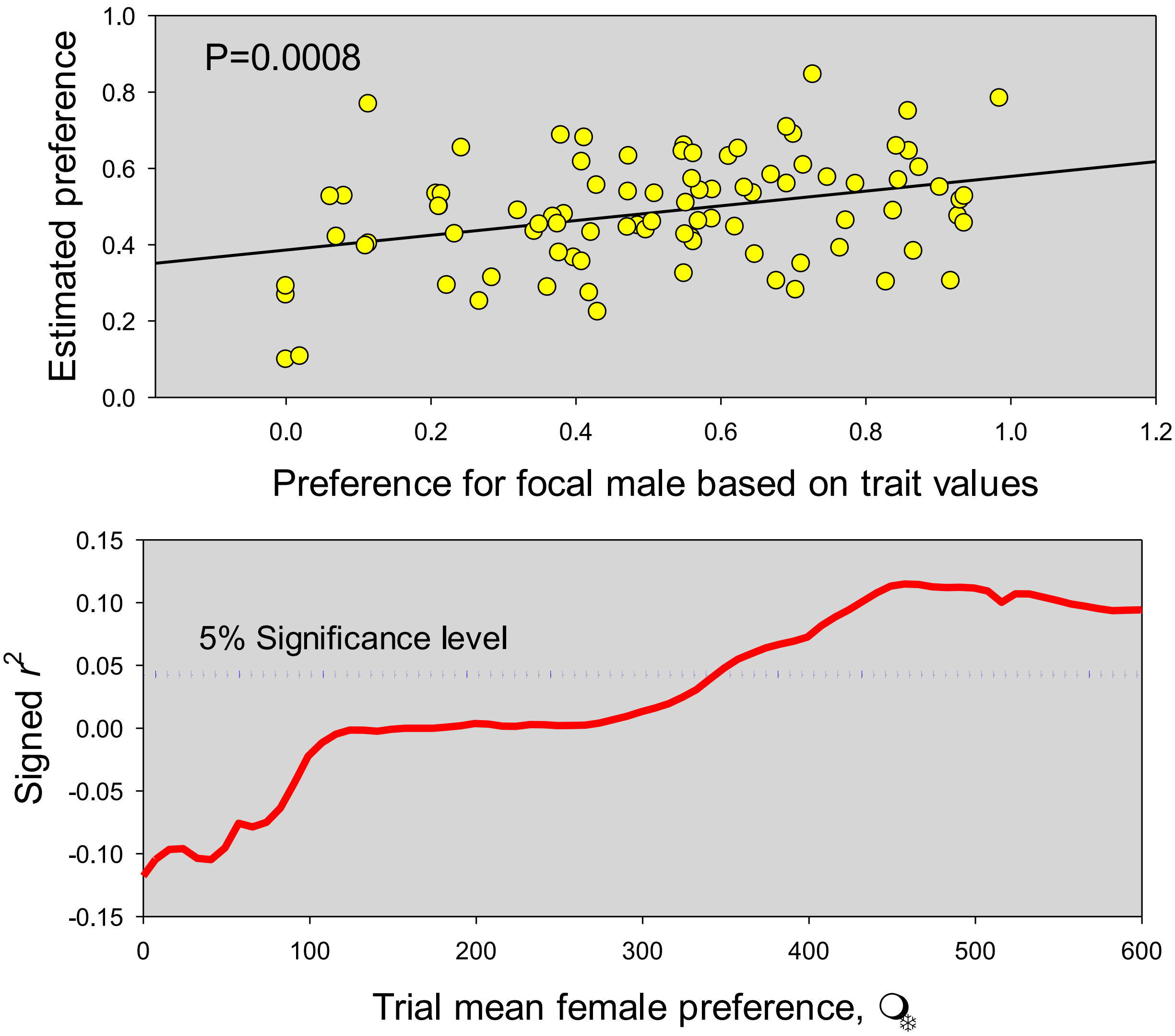
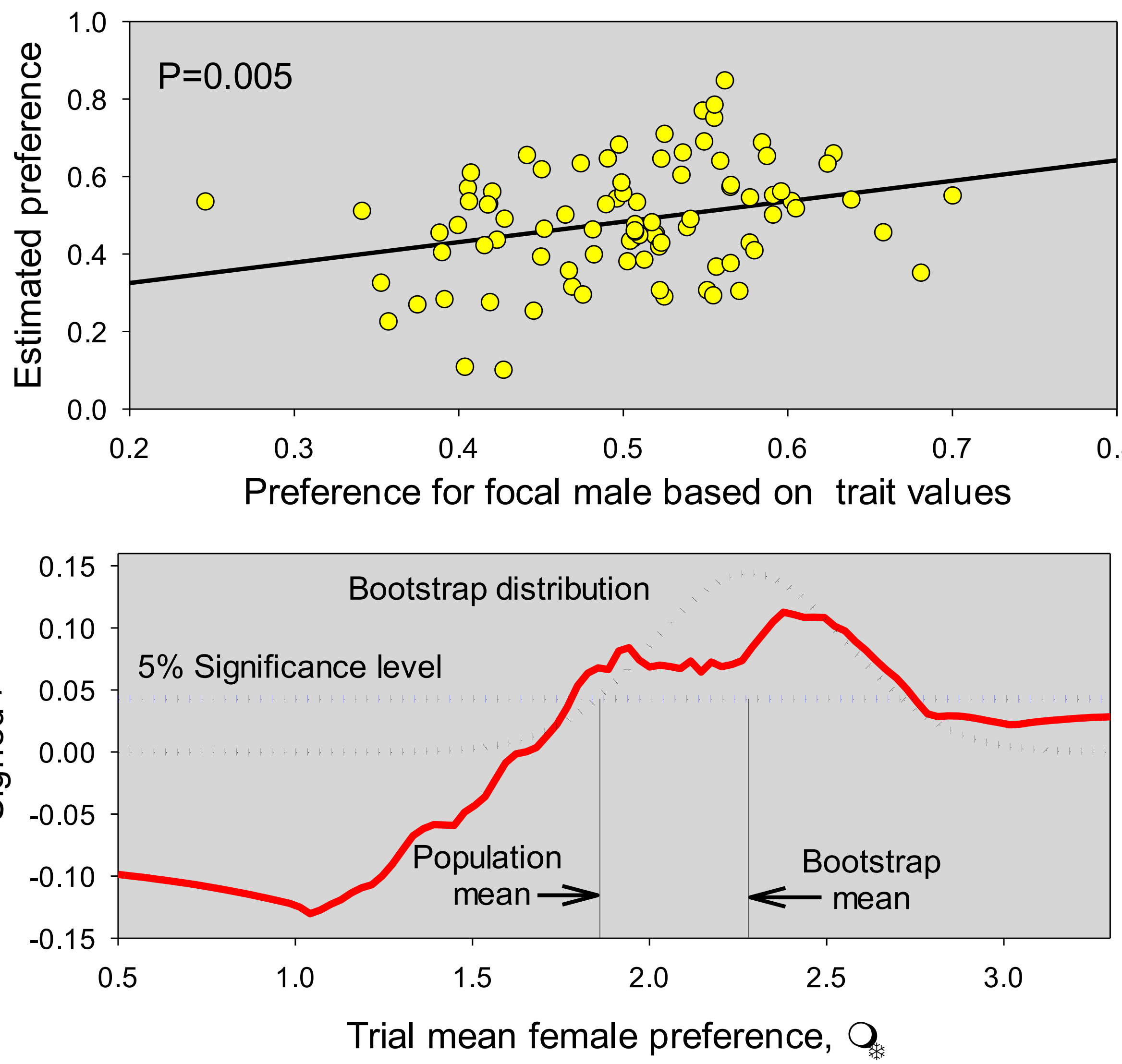


Figure 5. An example of a song trait, (Chirp rate) that shows statistical indication of stabilizing female preference, with the possibility of a directional component.

Top panel: Regression of estimated preference for the focal male, P_F, on the preference predicted from the male trait values. Solid line=fitted regression.

Bottom panel: Signed r^2 on the trial mean female preference, μ_T. Solid red line indicates the fit at the trial female preference. Only values lying above the 5% line (dotted blue line) are significant. The dotted line shows the normal distribution based on the bootstrap values.



SUPPLEMENTAL MATERIAL

1. A Review of current practice in choice experiments

To establish a representative sample of studies using choice designs to evaluate mate choice, we conducted a literature search using Web of Science. We first searched using the keywords "stabilizing, mate choice" and then searched again using the keywords "directional, mate choice". All studies that tested for stabilizing preference also tested for directional preference, but some studies only evaluated directional preference. For our sample, we retained only studies that included statistical tests for stabilizing or directional selection. Our search retrieved 365 papers, of which 38 described no choice experiments and 41 described choice experiments (Table S1). While the resulting sample of studies is not exhaustive, the data are sufficient to deduce general patterns. Choice experiments typically involved discrimination between two potential mates (mean= 2.4, median=2, SD=1.66, SE=0.04), and the mean number of trials (i.e., number of individuals of the choosy sex tested) ranged from 12 to 2400 (mean=147, median=28, SD=411). As indicated by the differences between the mean and median, the distribution of sample sizes was strongly skewed to the right, with only 9 studies (22%) having more than 100 trials.

We scored signal type as artificial, manipulated or natural. Artificial signals such as physical models or computer generated acoustic signals were used in 27% of choice experiments in our sample. The majority of experiments (73%) used manipulated signals, in which the potential mates to be compared were selected by the experimenter so that they differed measurably on the basis of an *a priori* recognizable character such as condition or morphology. Almost all (26 of 30) of these experiments compared males assigned to only two categories, and so were unable to distinguish directional from stabilizing preference as curvature cannot be established. In these studies, statistical analysis was based on the comparison of means or medians (C. of. M) and used categorical methods such as chi-square, Wilcoxon matched pairs or Mann-Whitney.

The third signal type used, natural signals, consisted of experiments in which the potential mates were selected at random from a population and not selected on the basis of some particular character. This type of variation was hardly ever used (no examples were found in choice experiments and only two examples in no choice experiments) though it more closely represents a natural situation than the other two.

2. Determination of female choice in the simulations

Our preference function (equation 1) is a probability function. To determine which of the two males was actually chosen in each trial of the simulation, we used the following algorithms. For the binary measurement experiment we generated a uniform random number, r_i , between 0 and 1and determined that the female selected the focal male whenever $P_{F,i} \geq r_i$. In the continuous case we modeled female preference as 100 separate trials with the probability for any single trial being given by $P_{F,i}$. The number of times the female selected the focal male, $n_{F,i}$, was generated and the relative preference, R_P, then computed as $n_{F,i}/100$ and the difference preference, D_P, as $2n_{F,i} - 100$.

3. The probability distribution of a focal male being chosen

To determine the probability distribution of a focal male being chosen we generated 100,000 triplets and calculated this probability from equations 5 or 6 (in main text), as required by the assumed female preference model. The distributions were almost entirely determined by the type of model and the standard deviations but not the means: hence, we have plotted only the results using means of 10 for both female preferences and the male trait (Fig. S1). Under the ED function the focal male is generally either strongly preferred or strongly rejected. In contrast, under the AD function the probability distributions take their highest values around 0.5 and hence the strength of preference is much lower than in the ED model.

4. Statistical analysis of the simulation model: 0,1 choice experiments

Not surprisingly, for the same sample size tests using 0,1 data were less satisfactory than using continuous data. Results for the ED and AD models were similar but the values for the AP model are consistently less than those of the ED model (Fig. S2). The statistical power of the protocols for detecting stabilizing preference in the simulations with stabilizing preference where data was coded as 0,1 exceeded 80% in 10 of the 16 combinations (Lower panels, Fig. S2). Power to detect directional preference was also low in most combinations (upper left panel, Fig. S2). The detection of apparent directional selection when male and female means were the same was typically about 5% (Upper right panel, Fig.S2)

Statistical power to detect directional preference when females showed only directional preference was less than 80% for 0,1 data (11.6% and 62.9% for male combinations 10:1 and 10:3, respectively) but 100% for the continuous data. For both 0,1 and continuous data directional preference was detected in only 5% of cases when female preference was random The bias in parameter estimates was less than 5% and negative in 13 of the 16 combinations for the 0,1 data, indicating that the estimates tended to underestimate the true mean.

REFERENCES

Ahuja A, Singh RS, 2008. Variation and evolution of male sex combs in *Drosophila*: Nature of selection response and theories of genetic variation for sexual traits. Genetics 179:503-509.

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.

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

Bentsen CL, Hunt J, Jennions MD, Brooks R, 2006. Complex multivariate sexual selection on male acoustic signaling in a wild population of *Teleogryllus commodus*. Am Nat. 167:E102-E116.

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

Brooks R, Hunt J, Blows MW, Smith MJ, Bussiere LF, Jennions MD, 2005. Experimental evidence for multivariate stabilizing sexual selection. Evolution 59:871-880

Chenoweth SF, Blows MW, 2005. Contrasting mutual sexual selection on homologous signal traits in *Drosophila serrata*. Am Nat. 165:281-289.

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

Crnokrak P, Roff DA, 1995. Fitness differences associated with calling behaviour in the two wing morphs of male sand crickets, *Gryllus firmus*. Anim. Behav. 50:1475-1481

Crnokrak P, Roff DA, 1998. The genetic basis of the trade-off between calling and wing morph in males of the cricket *Gryllus firmus*. Evolution 52:1111-1118.

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

Gerhardt HC, Tanner SD, Corrigan CM, Walton HC, 2000. Female preference functions based on call duration in the gray tree frog (*Hyla versicolor*). Behav Ecol. 11:663-669.

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

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

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.

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.

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 and Sociobiol. 60:150-158.

Jang YW, Greenfield MD, 1998. Absolute versus relative measurements of sexual selection: Assessing the contributions of ultrasonic signal characters to mate attraction in lesser wax moths, *Achroia grisella* (Lepidoptera : Pyralidae). Evolution 52:1383-1393.

King BH, Saporito KB, Ellison JH, Bratzke RM, 2004. Unattractiveness of mated females to males in the parasitoid wasp Spalangia endius. Behav Ecol and Sociobiol. 57:350-356.

Lehmann GUC, Lehmann AW, 2007. Bushcricket song as a clue for spermatophore size? Behav Ecol and Sociobiol. 62:569-578.

MacLaren RD, Rowland WJ, 2006. Differences in female preference for male body size in *Poecilia latipinnausing* simultaneous versus sequential stimulus presentation designs. Behav. 143:273-292.

McNamara KB, Jones TeM, Elgar MA, 2004. Female reproductive status and mate choice in the hide beetle, *Dermestes maculatus*. J Ins Behav. 17:337-352.

Oh KP, Shaw KL, 2013. Multivariate sexual selection in a rapidly evolving speciation phenotype. Proc Roy Soc Biol Sci. 280:20130482.

Owen MA, Rohrer K, Howard RD, 2012. Mate choice for a novel male phenotype in zebrafish, *Danio rerio*. Anim Behav. 83:811-820.

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.

Reading KL, Backwell PRY, 2007. Can beggars be choosers? Male mate choice in a fiddler crab. Anim Behav. 74:867-872.

Roff DA, Crnokrak P, Fairbairn DJ, 2003. The evolution of trade-offs: geographic variation in call duration and flight ability in the sand cricket, *Gryllus firmus*. J Evol Biol. 16:744-753.

Rosenblum EB, 2008. Preference for local mates in a recently diverged population of the lesser earless lizard (*Holbrookia maculata*) at white sands. J Herp. 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.

Schofl 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.

Steiger S, Capodeanu-Nagler A, Gershman SN, Weddle CB, Rapkin J, Sakaluk SK, Hunt J, 2015. Female choice for male cuticular hydrocarbon profile in decorated crickets is not based on similarity to their own profile. J Evol Biol 28:2175-2186.

Suk HY, Choe JC, 2002. Females prefer males with larger first dorsal fins in the common freshwater goby. J Fish Biol. 61:899-914.

Sztepanacz JL, Rundle HD, 2012. Reduced genetic variance among high fitness individuals: Inferring stabilizing selection on male sexual displays in *Drosophila serrata*. Evolution 66:3101-3110.

Verrell PA, 1995. Males choose larger females as mates in the salamander *Desmognathus santeetlab*. Ethology 99:162-171.

Wagner Jr WE, Murray AM, Cade WH, 1995. Phenotypic variation in the mating preferences of female field crickets, *Gryllus integer*. Anim Behav. 49:1269-1281.

Wong BBM, Svensson PA, 2009. Strategic male signalling effort in a desert-dwelling fish. Behav Ecol and Sociobiol. 63:543-549.

Xu J, Wang Q, 2009. A polyandrous female moth discriminates against previous mates to gain genetic diversity. Anim Behav. 78:1309-1315.