**Fisher’s sons’ effect in sexual selection: absent, intermittent or just low experimental power?**

**Abstract**

The Fisherian sexual selection paradigm has been called the null model of sexual selection. At its heart is the expectation of a genetic correlation (rG) between female preference and male trait. However, recent meta-analysis has shown estimated correlations are often extremely weak and not statistically significant. We show here that systematic failure of studies to reject the null hypothesis that rG=0 is almost certainly due to the low power of most experimental designs used. We provide an easy way to assess experimental power *a priori* and suggest that current data make it difficult to definitively test a key component of the Fisher effect.

**Introduction**

Sexual selection theory has largely been dominated by two ideas about the indirect fitness benefits of mate choice: good genes and the Fisher effect (Kirkpatrick 1987; Andersson 1994; Wade 2014). The former is based on the idea that more attractive males sire offspring of higher quality (i.e. expected fitness), both sexual and non-sexual, so by mating with preferred males, females gain both naturally and sexually selected benefits via their offspring. The Fisher effect posits that attractive males sire more attractive sons and females enjoy indirect fitness benefits by mating with these males because of the increased mating success of their attractive sons (Fisher 1930).

Fisher (1930) realized that if initially, some male trait revealed a naturally selected fitness advantage, then any female paying attention to the trait, would also benefit. If both trait and preference were heritable, this would result in the genes for both increasing in frequency and becoming linked as sons and daughters inherit both the trait and preference alleles (Fisher 1930). As preference spread, this could in turn result in further trait exaggeration such that male trait values would be driven beyond their naturally selected optima, and the female fitness benefits of mate choice would now only be accrued via the attractiveness of sons. This was formally modeled by Lande (1981) who showed that one major determinant of the outcome of the Fisher process was the strength of the genetic correlation between trait and preference (also see Kirkpatrick 1982; Arnold 1983). If this genetic correlation was strong enough (or for example, natural selection on the male trait relatively weak), then runaway sexual selection could occur, as envisaged by Fisher (1930). The theoretical importance of the association between trait and preference genes explains in part the continued calls to document this genetic correlation (e.g. Bakker & Pomiankowski 1995; Bakker 1999).

For trait and preference to be genetically correlated, there needs to be genetic variation for both, and there is now ample evidence for this (e.g. Bakker & Pomiankowski 1995; Pomiankowski & Møller 1995; Alatalo et al. 1997; Houde 1997; Bakker 1999; Taylor *et al*. 2007; Sharma *et al*. 2010; reviewed in Prokop *et al.* 2012 and Prokuda & Roff 2014; also see Houle 1992). However, despite the importance of trait-preference associations for Fisherian sexual selection, a recent meta-analysis of studies providing estimates of these correlations and their strength found that genetic correlations are either very weak or non-existent (Greenfield *et al*. 2014; and see Bakker 1999). Of the 20 studies retained in the meta-analysis, few documented significant and strong genetic correlations, with 95% confidence intervals overlapping 0 in the vast majority of cases. This is problematic for the field as the Fisher model has been called the null model of sexual selection and should always be present (is inevitable) *if* females prefer certain male phenotypes and both phenotype and preference have genetic underpinnings (Prum 2010; Mead & Arnold 2004; Roff & Fairbairn 2014; Wade 2014).

While one reason for the lack of detectable correlation maybe the breakdown of linkage disequilibrium by experimenters assigning mates during the breeding element of parameter estimation (this causes a 50% reduction/generation: Bakker & Pomiankowski 1995; Bakker 1999), a major issue that has not been directly assessed is the power of the designs that have been employed in an attempt to detect genetic correlations by sexual selection researchers. Genetic covariances, and the correlations derived from them are hard to estimate with precision (Roff 1997). Furthermore, the sample sizes used in published studies have often been small (e.g. Simmons & Kotiaho 2007). This led us to wonder if the widespread failure to detect significant correlations in so many studies (Greenfield *et al.* 2014; and see Bakker 1999) was likely to be a power issue. If so, could we provide a simple way for researchers to estimate the power of commonly employed experimental designs and so assess the likelihood of Type II error when planning experiments to test the expected genetic correlation between preference and trait. We attempt both here. While power estimates have been previously published for different breeding designs (usually using an ANOVA approach: e.g. Roff 1997), our simulations are based on using the animal model, which provides the most flexible and powerful way to estimate genetic correlations from any pedigree structure. And while our conclusions mirror those from previous analytical and simulation-based studies of sampling variance for genetic correlations generally (e.g. Robertson 1959; Visscher 1998; Bijma & Bastiaansen 2014), the importance of the Fisher model and testing it fully means reiteration of the message to a sexual selection audience is probably helpful.

**Methods.**

Using the R package pedantics (Morrissey *et al.* 2007; Morrissey & Wilson 2010) we simulated phenotypic values across different half-sib breeding designs in which each of *s* sires (*s* = 10, 20, 50, 100, 200) is crossed to *d* dams (*d* = 1, 3, 6) to produce *o* offspring/dam (*o* = 4, 6, 10). We assume offspring phenotypes are measured but parental ones are not and that the sex ratios of measured offspring are equal (i.e. sample size of observed phenotypes is (*s*\**d*\**o*)/2 for both male trait and female preference, with o/2 sons and o/2 daughters/dam). For each of the 45 breeding designs considered, we simulated 500 phenotypic datasets at each of 225 different quantitative genetic parameter sets. These parameter sets were defined by all possible combinations of male trait genetic variance *VAm* (from 0.1 to 0.5 at increments of 0.1), female preference genetic variance *VAf* (from 0.1 to 0.5 at increments of 0.1), and the intersexual genetic correlation *rG* (from 0 to 0.8 at increments of 0.1). We assume that trait and preference have unit (phenotypic) variance (i.e. are measured in standard deviation units) such that VA is equal to the narrow sense heritability (h2), and the residual variance *VR* is equal to 1-*VA*. Since individuals are either male or female, the total (within-trait) cross-sex phenotypic covariance is not identifiable and so was fixed to zero in the simulation (and subsequent model fitting step). While we assume no maternal or non-genetic effects, pedantics can be used to investigate these if they are a suspected feature of a system (Morrissey *et al.* 2007).

For each simulated data set we fitted and compared two bivariate animal models using ASREML-R. In the first, the genetic variance-covariance matrix (***G***) between male trait and female preference was estimated with no constraints. In the second the model was refitted with COVA (and hence *rG*) fixed to zero. Provided both reached convergence, likelihoods of the constrained and unconstrained models were then compared by likelihood ratio test to determine the significance of the genetic covariance. We make the standard assumption that twice the difference in model log likelihoods is distributed as χ2 on 1 degree of freedom. For each combination of breeding design and parameter set, statistical power was estimated as the proportion of simulated data sets (in which both models converged) for which *rG* was significantly different from zero at α=0.05.

The full R script used is provided in supplemental materials. It could be readily modified to cover other breeding designs, parameter sets, analytical models (ANOVA, sire models etc), methods of inference (e.g. model fitting by MCMCglmm or other packages), or null hypotheses (e.g. estimate power to detect cross-sex *rG* < 1 for planned studies of intra-locus sexual conflict).

**Results.**

Across all power estimates (for every heritability (=VA: see above), covariance and breeding design combination) convergence of both bivariate models was obtained for an average of 465 data sets (median 494, interquartile range 48) out of the possible 500 simulated. In total 11698 power estimates were generated, ranging from β = 0.004 to β = 1. Of these, only in 17.4% (2034 of 11698) of scenarios considered was power acceptable (β >0.8).

Inspection of the simulations shows that for experiments that measure a total of 300 animals or less, regardless of how these are allocated to sire, dam and offspring, power will never exceed 0.8 (irrespective of the true **G** matrix) unless rG is at least 0.8 and both heritabilities exceed 0.5, in which case two designs are just adequate (50 sires, 1 dam/sire, 6 offspring/dam and 75 sires, 1 dam/sire, 4 offspring/dam) (Figure 1). Doubling of this sample size means that power will be adequate for most designs only when heritabilities and genetic correlations are both very high – both h2 ≥ 0.5; rG ≥ 0.6 (Figure 1). As genetic variance and covariance strengths decrease, power declines rapidly even with a sample size of 600, such that power never exceeds 0.8 once the heritability of one trait is 0.5 or lower, or if the genetic correlation is lower than about 0.6 (Figure 1 & 2). Thus for most studies conducted to date (Greenfield *et al*. 2014) experimental designs used have certainly been underpowered for rejecting the null hypotheses of rG=0.

***Insert Fig. 1 near here***

If we use the most recently published mean estimated heritability values for trait and preference at a species level (ca. 0.3/0.5: Prokuda and Roff 2014), we find that for genetic correlations less than 0.1, none of the designs we simulate have power greater than 0.8, even though the largest designs simulated were much bigger than typically used empirically (6000 offspring), while at *rG* = 0.4, only 6 of 15 have satisfactory power (Figure S1). If we use study-wise empirical estimates of mean heritability values (ca. 0.2/0.3: Prokuda & Roff 2014), things are worse and we find that at *rG* = 0.4, only 4 of 15 designs we simulate are sufficiently powerful (Figure S1) and even at *rG* = 0.8, 40% of the designs we simulated were far from an acceptable beta (Figure S1). In fact, only designs assessing ca. 1000 offspring or more had β>0.8. At more differentiated heritability values similar outcomes were found, and at the lowest heritability values with genetic correlations less than 0.5, none of the designs simulated had sufficient power (data not shown: but see Figure 2). Finally, we estimated the precision of genetic correlation estimates as a function of sample size for various sampling designs. Designs have to be very large and/or genetic correlations have to be very strong for the 95% confidence intervals (CIs) of estimates not to overlap zero (Figure 3). This is true of some designs that have sample sizes as large as 1200 individuals when genetic correlations are relatively weak (rG ≤ 0.3), and even with this relatively large sample size, CIs are very large.

Again note that we assume all effects are additive, but the results suggest that using the animal model total sample size is more important than how effort is allocated to sire, dam, or offspring. We caution however that this outcome is almost certainly contingent on the absence of maternal and common environment effects (which can be simulated using pedantics by adapting the code we provide here).

***Insert Fig. 2 & 3 near here***

**Discussion**

Our simulations revealed that the sample sizes needed for powerful tests of the cross-sex trait-preference genetic correlations will have to be very large. Even with designs that employ 100 sires, 6 dams per sire, and 10 offspring per dam (= 6000 offspring), have male trait heritabilities of 0.5, which is slightly above the mean heritability of sexually selected traits (Bakker 1999; Prokuda & Roff 2014), and female preference heritability greater than 0.2 (which mirrors empirical estimates: Bakker 1999; Prokuda & Roff 2014), power only exceeds 0.8 when the genetic correlations are greater than 0.3. This is a very large design and exceeds by an order of magnitude the sample sizes used in many studies that have been deployed to measure trait-preference associations (Greenfield *et al*. 2014). At sample sizes closer to those more frequently employed (ca. 600 offspring), even when genetic correlations are a very sizable 0.6 and joint heritabilites are high (0.5: which exceeds mean estimates Prokuda & Roff 2014), power can be inadequate (e.g., β ≈ 0.75). The sample sizes used in the studies that made up the recent meta-analysis (Greenfield *et al*. 2014) were (when we could clearly determine sample size) frequently smaller than 600, and studies that employed full-sib/half-sib designs all had less that 50 sires. These are clearly low power. More dishearteningly, we find that if trait heritabilities are less than 0.5, only the largest of the designs we simulate (100-200 sires) generally have adequate power, and if the genetic correlations are less than 0.2, most of these also fail (even with high trait heritability). We note again that many previous authors have provided power estimators for quantitative genetic designs and it is clear that large designs are often needed (e.g. Roff 1997; Lynch & Walsh 1998), especially when estimating genetic correlations (e.g. Visscher 1998; Bijma & Bastiaansen 2014). Our power estimates are for an animal model analysis, the most flexible and powerful tool currently in use, but the message is unchanged (also see Meyer 1989). Our simulations are also consistent with analytical investigations of the sample sizes needed to accurately estimate genetic covariances (e.g. Robertson 1959; Visscher 1998).

This all suggests that it is probably not possible to currently draw strong conclusions about the average magnitude of trait-preference correlations or their significance, and hence the taxonomic distribution and importance of the Fisher effect. If we are to seriously try and estimate these associations, far more effort needs to be expended. This becomes all the more clear when the experimental effort (sample size) that needs to be expended to obtain powerful parameter estimation is so great. Be that as it may, the centrality of the Fisher effect to sexual selection theory means robust estimates of preference-trait associations are sorely needed, even if other requirements of the sons’ effect, like the heritability of male attractiveness, generally hold true (e.g. Prokop et al. 2012). However as we show here, considerable work will be needed to obtain reliable estimates, and despite this being a reiteration of an old theme (e.g. Robertson 1959), the message seems to have been forgotten to an extent (or so it appears based on the studies reviewed by Greenfield *et al*. (2014)).

An additional problem is that the minimum “biologically relevant” effect size is difficult to determine as the evolutionary consequences of any inter-sexual genetic correlations are partly determined by the strength of natural selection acting on the male trait (Lande 1981). However, if the estimated effect sizes obtained to date (Greenfield *et al*. 2014) are unbiased,then very large experiments will be needed to achieve statistical significance in any single study. The alternative is that we should be more sanguine and accept that estimates obtained with less powerful designs are nonetheless informative, particularly when used in meta-analyses. That way, at least we will have an estimate of the magnitude of preference-trait correlations in general, because meta-analyses can provide powerful conclusions even when individual studies are imprecise.

Absence of (statistically significant) evidence should not be conflated with evidence of absence, and while power limitation almost certainly contributes to the lack of significant rG estimates in the literature, it should not be a source of systematic bias towards small effect sizes. In general sources of bias in estimating quantitative genetic parameters are widely recognized (e.g., |rG| tending towards 1 if one or other VA estimate is very small, unknown influence of non-additive and sex-linked genetic effects on rG), although not necessarily easy to control for. However,a recent simulation study (Roff & Fairbairn 2014) points to another reason why trait- preference genetic correlation estimates may be low – females are usually only allowed to sample 1 or 2 males in experimental studies estimating mate preference. Roff and Fairbairn’s (2014) simulations varied a range of parameters including female and male genetic variances, mate sampling rules and numbers of males sampled by females. This latter parameter had a strong impact on the strength of the genetic correlations and the likelihood of runaway sexual selection, but in most empirical studies estimating preference heritability, only a few males are sampled. Similarly, the shape of preference functions had a major effect on simulation outcomes (Roff & Fairbairn 2014). So studies that want to estimate male-female trait-preference correlations could increase power by allowing females to sample a biologically realistic (for the species in question) number of males, and those that find very small correlations should then investigate whether females have absolute or relative mate preferences as with the latter, simulations could not generate strong genetic covariance (Roff & Fairbairn 2014). Furthermore, it is possible that the traits measured are poor proxies of total male attractiveness and as empiricists we must always be aware that outcomes and their accuracy will depend on how well the trait(s) measured capture attractiveness in its entirety (see e.g. Prokop & Drobniak 2016).

This leads us to reinforce previous recommendations so that new studies 1) avoid use of random mate assignment (Bakker & Pomiankowski 1995), 2) conduct *a priori* power analyses with the expectation of requiring large sample sizes (Green 1979) using R-script and figures we provide here, 3) allow females to sample realistic numbers of males in order to determine preference (Roff & Fairbairn 2014), and 4) ensure we are measuring attractiveness in a meaningful manner (Prokop & Drobniak 2016). There are of course practical challenges associated with each of these suggestions, but by not randomly allocating mates experimenters do not weaken trait-preference linkage, by using large sample sizes confidence intervals are reduced and sampling more males should strengthen correlations because preference estimates could be more accurate (likewise with measuring attractiveness). While our simulations have addressed the most common strategy of estimating rG from single trait observations in sibship data, we note that the power implications of departing from this may also be worth exploring. For instance, repeated measures designs in which preference is assayed multiple times per female may be beneficial, while iso-female lines could also be used because they capture linkage and repeated measures of preference can be obtained for each genotype (see David *et al*. 2005).

In spite of the sobering outcome of the Greenfield *et al.* (2014) meta-analysis and of our simulations, the logic of Fisher effect remains and we agree that indirect benefits via sons is an inescapable consequence of sexual selection when trait and preference have genetic underpinnings (Mead & Arnold 2004) (although runaway selection may not be: Arnold 1983). This inevitability is made clear by the various approximations that have been used to describe male-female trait-preference covariance. For example, the approximation rG = cVAmVAf (Barton & Turrelli 1991), states that male-female covariance depends on the additive genetic variance in trait and preference (VAmVAf) and the ability of preference and trait to create non-random mating (c) (also see Kirkpatrick and Barton 1997). Thus non-random mating or phenotypic associations between preference and trait, plus heritable preference and trait will generate trait-preference genetic covariance.

Female preference remains poorly understood in many regards despite repeated calls for more study (e.g. Bakker & Pomiankowski 1995; Hosken & House 2011) and equally there is a paucity of trait-preference covariance studies, despite this covariance being a key parameter in the most important model (Prum 2010) of indirect-benefit sexual selection. This limits full understanding of sexual selection, and as we show here, experiments to date simply have not been powerful enough to shed much light on the issue.

**References**

Alatalo, R. V., Mappes, J. & Elgar M. A. 1997. Heritabilites and paradigm shifts. *Nature* **385**: 402-403.

Andersson, M. 1994. *Sexual Selection*. Princeton University Press, Princeton.

Arnold, S. J. 1983. Sexual selection: the interface of theory and empiricism. In: *Mate Choice* (P. Bateson, ed.), pp 67-108. Cambridge University Press, Cambridge.

Bakker, T. C. M. 1999. The study of intersexual selection using quantitative genetics. *Behaviour* **136**: 1237-1266.

Bakker, T. C. M. & Pomiankowski, A. 1995. The genetic basis of female mate preference. *J. Evol. Biol.* **8**: 129-171.

Barton, N. & Turrelli, M. 1991. Natural and sexual selection on many loci. *Genetics* 127:229-255.

Bijma, P. & Bastiaansen, J. M. W. 2014. Standard errors of the genetic correlation: how much data do we need to estimate a purebred-crossbred genetic correlation? *Genet. Sel. Evol*. **46**: 79.

David, J. R., Gibert, P. Legout, H., Petavy, G., Capy, P. & Moreteau, B. 2005. Isofemale lines in *Drosophila*: an empirical approach to quantitative trait analysis in natural populations. *Heredity* **94**: 3-12.

Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. Claredon Press, Oxford.

Green, R. H. 1979. *Sampling Design and Statistical Methods for Environmental Biologists.* Wiley & Sons, New York.

Greenfield, M. D., Alem, S., Limousin, D. & Bailey, N. W. 2014. The dilemma of Fisherian sexual selection: mate choice for indirect benefits despite rarity and overall weakness of trait-preference genetic correlations. *Evolution* **68**: 3524-3536.

Hosken, D. J. & House, C. M. 2011. Sexual selection. *Curr. Biol*. **21**: R62-65.

Houde, A. E. 1997. *Sex, Color, and Mate Choice in Guppies*. Princeton University Press, Princeton.

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics* **130**: 195-204.

Kirkpatrick, M. 1982. Sexual selection and the evolution of female choice. *Evolution* **36**: 1-12.

Kirkpatrick, M. 1987. The evolutionary forces acting on female mating preference in polygynous animals. In: *Sexual Selection: Testing the Alternatives* (J. W. Bradbury and M. B. Anderson, eds), pp. 67-81. John Wiley and Sons, New York.

Kirkpatrick, M. & Barton, N. 1997. The strength of indirect selection on female mating preference. *Proc. Natl. Acad. Sci. USA* **78**: 3721-3725.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. *Proc. Natl. Acad. Sci. USA* **78**: 3721-3725.

Lynch, M. & Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, Mass.

Mead, L. S. & Arnold, S. J. 2004. Quantitative genetic models of sexual selection. *Trends Ecol. Evol.* **19**: 264-271.

Meyer, K. 1989. Estimation of genetic parameters. In: *Evolution and Animal Breeding: Reviews of Molecular and Quantitative Approaches in Honour of Alan Robertson* (W. G. Hill and T. F. C. Mackay, eds), pp 161-167. CAB International, Michigan.

**Morrissey, M. B., Wilson,** A. J., **Pemberton, J. M. & Ferguson M. M.** 2007. A framework for power and sensitivity analyses for quantitative genetic studies of natural populations, and case studies in Soay sheep (*Ovis aries*). J. Evol. Biol*.* **20**: 2309-2321.

**Morrissey, M. B. & Wilson, A. J.** 2010. *pedantics*, an R package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing. *Mol. Ecol. Res.* **10**: 711-719.

Pomiankowski, A. & Moller, A. P. 1995. A resolution to the lek paradox. *Proc. R. Soc. Lond., B* **253**: 173-181.

Prokop, Z. M., Michalczyk, L., Drobniak, S. M., Herdegen, M., & Radwan, J. 2012. Meta-analysis suggests choosy females get sexy sons more than “good genes”. *Evolution* **66**: 2665-2673.

Prokuda, A. Y. & Roff, D. A. 2014. The quantitative genetics of sexually selected traits, preferred traits and preference: a review and analysis of the data. *J. Evol. Biol.* **27**: 2283-2296.

Prum, R. O. 2010. The Lande-Kirkpatrick mechanism is the null model of evolution by intersexual selection: implications for meaning, honesty, and design in intersexual signals. *Evolution* **64**: 3085-3100.

Robertson, A. 1959. The sampling variance of the genetic correlation coefficient. *Biometrics* **15**: 469-485.

Roff, D. A. 1997. *Evolutionary Quantitative Genetics*. Chapman and Hall, New York.

Roff, D. A. & Fairbairn, D. J. 2014.The evolution of phenotypes and genetic parameters under preferential mating. *Ecol. Evol*. **4**: 2759-2776.

Sharma, M.D., Tregenza, T. & Hosken, D. J. 2010. Female mate preference in *Drosophila simulans*: evolution and costs. *J. Evol. Biol*. **23**: 1672-1679.

Simmons, L. W. & Kotiaho, J. S. 2007. Quantitative genetic correlation between trait and preference supports a sexually selected sperm process. *Proc. Natl. Acad. Sci. USA* **104**: 16604-16608.

Taylor, M. L., Wedell, N. & Hosken, D. J. 2007. The heritability of attractiveness. *Curr. Biol*. **17**: R959-R960.

Visscher, P. M. 1998. On the sampling variance of intraclass correlations and genetic correlations. *Genetics* **149**: 1605-1614.

Wade, M. J. 2014. Genotype-by-environment interactions and sexual selection: female choice in a complex world. In: *Genotype-by-Environment Interactions and Sexual Selection* (J. Hunt and D. J. Hosken, eds), pp. 3-18. Wiley Blackwell, UK.

**Figure 1**. Power plots of simulations that employed the sample sizes most frequently used in empirical studies that estimate intersexual preference-trait genetic correlations. Shown here are power estimates when heritabilities for trait and preference are both high, 0.5 which exceeds average joint estimates (Prokuda and Roff 2014). The top panel shows various breeding designs fixed at 300 total animals, and the lower panel 600 animals. Note totals are fixed regardless of how animals are allocated to sire, dam or offspring. So for example a design of 2 sires with 5 dams each and 30 offspring/dam and 100 sires each with one dam all of whom have 3 offspring both sample 300 animals. At the smaller sample size, β > 0.8 only for two designs and genetic parameters, while with the larger sample, β > 0.8 only when rG ≥ 0.6. Shaded area on both panels represents the 95% CI of the mean genetic correlation (0.05-0.23) reported in Greenfield *et al.* (2014).

**Figure 2.** Plots showingestimated power to detect cross-sex genetic correlations with varying sample sizes, trait heritabilities (0.1 to 0.5)and genetic correlations (0.2 or 0.8). Open circle, closed circle and open triangle represent model outputs where genetic correlations were fixed at 0.8 and preference and trait heritabilities were 0.5, 0.3 and 0.1 respectively. Closed triangle, open square and closed square represent data with genetic correlation of 0.2 and preference and trait heritabilities were 0.5, 0.3 and 0.1 respectively. This again shows that many designs have weak power (especially those that sample less than 3000 offspring). With preference and trait heritabilities of 0.3 and genetic correlation of 0.2 (which approximates current estimates of these values) a sample size of 12000 individuals is required to achieve acceptable power.