The Empirical Study of Sexual Selection Group Report

G.S. Wilkinson, Rapporteur
T.H. Clutton-Brock   H.-U. Reyer
A. Grafen   W.J. Sutherland
P.H. Harvey   A.J. van Noordwijk
R.D. Howard   M.J. Wade
K.E. Linsenmair   P. Wirtz
H.J. Poethke

INTRODUCTION

The process of sexual selection occurs when some aspect of an organism's phenotype covaries with mating success or fecundity by influencing intrasexual competition or mate choice. When such selection operates differentially on the sexes and heritable variation exists for the selected characters, sexual dimorphism frequently results. As with other evolutionary phenomena, sexual selection may have been important in an organism's evolutionary past but its applicability to extant populations depends on the availability of phenotypic variation. In this report we critically review methodology and specific measures which have been proposed as descriptors of sexual selection in either or both contexts. To structure the presentation the report has been divided according to the three classical methods for the study of evolution: comparative cross-species studies, observational studies within a species, and experimental manipulations in the field or laboratory. For each method we point out whether the technique is appropriate for studying how selection has operated in the past or how selection is occurring at present and provide several suggestions for promising research which has yet to be pursued.

COMPARATIVE STUDIES

Comparative studies can often provide insight into the past role of sexual selection in producing cross-taxonomic variation. Furthermore, in some situations the microevolutionary forces which act within a species can be inferred and subsequently confirmed with either observational or experimenal studies.

There are three basic components in any comparative study: a) selection of a

Standing, left to right:
Paul Harvey, Eduard Linsenmair, Rick Howard, Peter Wirtz, Michael Wade

Seated (center), left to right:
Alan Grafen, Bill Sutherland, Arie van Noordwijk, Tim Clutton-Brock

Seated (front), left to right:
Uli Reyer, Achim Poethke, Jerry Wilkinson

phylogenetic group of species, b) choice of a measure suitable to the organisms which contains information about the importance of sexual selection, and c) identification of one or more ecological or social variables which the researcher may have prior reason to believe indicates the selective force. At each step several potential problems should be stressed.

Choosing an appropriate phylogenetic comparison can be difficult. Ideally, independent evolutionary events are identified so that problems of nonindependence do not bias the statistical analysis by nonrandomly inflating sample sizes. Ridley (1983) has suggested a technique, which unfortunately can only be used when the characters defining the phylogeny are dichotomous, for recognizing independent evolutionary events. Further statistical work is needed to overcome this problem. Ridley's method also necessitates knowledge of the true phylogeny. This is a problem for which systematists have no single, simple answer. Whether or not numerically intensive methods, such as Felsenstein's (1985) bootstrap technique, are required will depend on the amount of uncertainty in the proposed phylogeny.

In principle, any ecological or social character which indicates a selective agent can be used as a measure of the strength of selection. However, caution should be exercised when interpreting significant correlations because unmeasured selective forces may have produced the pattern in the data. For example, sexual dimorphism in canine length (the sexual selection measure) may correlate with harem size (the social variable implicating sexual selection) in some hypothetical group of primate species, thereby suggesting that longer canines evolved either in the context of male combat for access to females or female choice of males. It is possible, however, that the true function of canines, independent of any role in sexual selection, is to protect a male's own offspring from predators that occur commonly in the geographic range of some species more than others. Such a function might generate a spurious correlation between canine length and harem size that could only be detected by further comparative study or experimentation (see Harvey et al. 1978 for analysis of real data). Measures of sexual selection unrelated to any specific character, such as indices of the variance in mating success or the operational sex ratio, have additional problems in interpretation which will be discussed below under the section on *OBSERVATIONAL STUDIES*

Despite these difficulties, the comparative method offers a powerful tool for identifying selective agents responsible for sex differences in morphological or behavioral characters, both within and between radiations. Although such studies tend to be limited by availability of data, we can suggest a few areas in which sufficient data may be accessible; others have been suggested in the remaining group reports in this volume. All equilibrium models of sexual selection (Heisler et al. this volume) require that a sexually selected trait impose some viability cost when expressed. This assumption could be checked

by comparing age-specific patterns of survival across species that vary in the expression of particular sexually selected traits. In such a comparison one must be alert to the possibility that there are two ways for species to vary in this regard: when and how much the trait is expressed during ontogeny. Also, the degree of apparent convergence in sexually selected traits, e.g., bright coloration and feather elaboration in birds, is substantially less than that observed for some selected traits. For example, we can think of no sexual selection examples which parallel the tendency for large wading birds to be white. Other examples of plumage coloration convergence due to natural selection across radiations easily come to mind. We suggest that the degree of apparent arbitrariness in the coloration of some male birds may indicate species that have experienced female choice sexual selection since one would expect some convergence in color if signal intensity (either brightness or degree of contrast) was used in intrasexual displays. Comparative study on birds and perhaps reef fishes in conjunction with observations could verify this suggestion. We admit, though, that an appropriate measure to compare coloration patterns across species is not obvious.

OBSERVATIONAL STUDIES

Unquestionably, the majority of research on sexual selection falls within this category. A wide variety of measures have been proposed to describe the intensity of sexual selection within a species. Below we point out the strengths and weaknesses of several measures and suggest how they can be used most effectively. To facilitate this discussion we recognize two categories: noncharacter-based and character-based measures. Noncharacter-based measures simply quantify the degree of variation in reproductive success without a consideration of variation in phenotypic traits. Character-based measures, on the other hand, estimate the degree of covariance between one or more phenotypic traits and fitness (selection differentials) or estimate the regression of fitness on these traits (selection gradients).

Noncharacter-based measures

The relative variance in fitness or opportunity for selection, I, (Arnold and Wade 1984b); Wade this volume) sets an upper limit on the intensity of directional selection since the covariance between a character and fitness can be no greater than the covariance (i.e., variance) between fitness and itself. However, this measure, as well as all other variance-based measures, suffers two kinds of limitations. The first limitation is that variance in fitness is a necessary but not sufficient condition for selection. For selection to occur, one or more characters must also covary with fitness. Arnold and Wade (1984a)

illustrated this point using Triver's cross-sectional study of the lizard, *Anolis garmani*. In this example, the opportunity for sexual selection in females was nearly twice that in males. Nevertheless, the selection differential on male body size was five times that observed for females.

The second limitation concerns the random or "error" component in the fitness variance analysis. In order to estimate the magnitude of this environmentally induced component of fitness variance, repeated measures on individuals are needed. The "random" component of variance (see Sutherland this volume), is the within-individuals or error variance in a repeated measures study. Although some studies can estimate this error variance in the lab (e.g., Houck, Arnold, and Thisted 1985), in nature an individual has only one life history and repeated measures are not possible.

The search-time model of Sutherland (1985, this volume) is an attempt to show how the opportunity for sexual selection is dependent upon the fraction of time spent searching or competing for males. This algebraic model is similar to Triver's (1972) notion of parental investment but has the advantage that it can be related to other previously suggested measures, such as the operational sex ratio and relative variance in fitness. Sutherland's model, however, shares the same limitations as the variance-based measure discussed above. Although measuring search time may be more appealing for use in comparative studies than determining the variance in fitness, in actuality, search time as defined in Sutherland's model may be impractical to estimate in all but a few taxa.

In conclusion, it would be imprudent to initiate a study of sexual selection by measuring only variance in fitness, or some correlate of fitness such as search time, without simultaneously measuring the expression of a purported sexually selected trait. However, where variance in fitness can be estimated easily and partitioned into components which correspond to mating success and other life history stages, it might be informative to calculate some of the measures discussed above and compare them to simultaneous character-based measures of sexual selection to ascertain how much the two approaches do differ. So far, such comparisons have only confirmed the potential problems of variance measures. As mentioned above, differences in the opportunity for selection, I, between the sexes may show no relation in sign or magnitude to the differences in selection differentials. From measures of seasonal reproductive success for *Rana catesbeiana* and *R. sylvatica*, Howard (1987) calculated that $I_{males} > I_{females}$ in both species. However, no sexual dimorphism, either in body size or growth rate, occurs in *R. catesbeiana* while males are smaller and grow slower in *R. sylvatica*. In contrast to the predictions using I these patterns of dimorphism are consistent with the selection gradients estimated by regressing seasonal reproductive success on body size. The gradients are equal between male and female *R. catesbeiana* and the male gradient is less than the female gradient in *R. sylvatica*.

Character-based measures

Although a selection differential measures how much the mean value of a character has changed within a generation, it does not necessarily indicate that phenotypic variation in that character has a reproductive consequence. An observed selection differential could be solely due to selection acting on some other correlated character. To control for any phenotypic correlations a multiple regression of fitness on characters can be used to estimate the relative effect of selection on each character. The resulting partial regression coefficients are estimates of selection gradients which indicate the intensity of selection acting directly on each trait.

Such an analysis can be useful in ascertaining the mechanism by which selection is acting. For example, a selection gradient analysis could, in principle, determine that selection operates primarily on antler size rather than body size in a hypothetical ungulate. This method has also been recommended for estimating the shape of the adaptive landscape in order to predict change in phenotypic distributions over time (Lande and Arnold 1983). To infer how selection has acted in the distant past or how it will act over generations in the future requires not only that selection does not change over time but also that the genetic covariance matrix remains constant. While this latter assumption should be true as long as effects due to selection and drift are weak, there is empirical evidence (Lofsvold 1986) that it does not hold for time periods which correspond to the divergence of subspecies and species.

Selection gradient analysis requires fulfillment of at least six assumptions which are discussed briefly below. Further discussion of the problems and merits of this regression technique can be found in any regression methods text, such as Draper and Smith (1981), as well as in Manly (1985) and Endler (1986).

1) If a study neglects directly selected characters that are correlated with the included characters, the estimated gradients can be incorrect. Excluding relevant variables is, of course, a problem in any scientific study.

2) This technique makes the implicit assumption that directional selection acts on a linear combination of characters (stabilizing selection is estimated with a quadratic regression). Consequently, important variables must be included on a scale which preserves the purported relationship. If the appropriate transformation is included in the analysis, significant coefficients will emerge only on the transformed variable. For example, if sexual selection has acted directionally to increase body area and thus conspicuousness, in a reef fish then a significant gradient would appear only for the square of body length, not body length itself.

3) There may be insufficient phenotypic variation to detect the actual shape of the fitness landscape. For example, if the distribution of one character in the population is under the tail of a Gaussian fitness function, selection will appear

to be directional rather than stabilizing. The solution to this problem is either to re-estimate the gradients continually over time or artificially expand the phenotypic distribution. A rectangular distribution should maximize the power in estimating the selection surface.

4) If characters are not measured on every individual born into the population, then estimates of selection gradients can be biased. For example, if some individuals die before attaining the phenotype expressed at sexual maturity, they cannot be included in the distribution of adult phenotypes available to sexual selection. A partial solution to this problem is to divide selection into different episodes during an animal's lifetime (cf. Arnold and Wade 1984b).

5) If environmental variation causes spatial or temporal heterogeneity in selection or multiple nonadditive episodes of selection, estimates of selection gradients from pooled data sets can be misleading. This is especially relevant for sexual selection because of the inherent frequency dependence of the process. Furthermore, a nonstable age distribution necessarily creates temporal variation in selection in species with overlapping generations. Data sets therefore, must be screened carefully before pooling across sites or years. It should be noted, though, that the existence of heterogeneous selection can, in principle, be identified if contextural variables, such as indices of habitat heterogeneity or lek size, are included in the selection gradient anaylysis (Heisler and Damuth, unpublished manuscript).

6) The measured trait should not change with age otherwise effects of ontogeny and selection may be confused. This problem is particularly relevant for indeterminately growing species. Unfortunately, such species often have the advantage of revealing more morphological variation to selection. One solution to this problem is to estimate a time invariant character, such as the parameters of a function which describes the growth trajectory.

Despite these difficulties, selection gradient analysis is perhaps the most powerful method available for analysing observational data. Heisler *et al.* (this volume) discuss how this technique could be used to test assumptions and predictions of several sexual selection models. Of course, this technique requires an estimate of fitness for each individual. Lifetime reproductive success is usually recommended (Charlesworth 1980), although in age-structured populations with overlapping generations experiencing fluctuating episodes of selection it may not be the best measure of fitness. No recommendations for an alternative measure have yet been made for this complicated, but potentially frequent, case. If less than a full generation of data is collected, then the gradients will only apply to the components of fitness corresponding to the life history stage measured and should not be used to predict change across generations. In some situations, it may be possible to use surrogate measures of fitness components, such as mating speed, for seasonal or lifetime mating success if other evidence can substantiate a correlation between the surrogate measure and the fitness component. Identifying the appropriate interval over which a rate measure should be estimated can only be determined from familiarity with the reproductive schedule of the organism.

An illustration of the utility of the gradient method over simple selection differentials is provided by Price et al. (1984). Selection differentials on beak depth and beak width in the finch *Geospiza fortis* recorded in 1976, 1980, and 1982 were all either positive or zero. However, the selection gradient on beak width was consistently negative while the gradient on depth was positive over those periods. The gradient on beak depth, not the differential, is more consistent with the identified selection pressure, i.e., the ability to feed efficiently on large seeds on Isle of Daphne Major in the Galapagos.

EXPERIMENTAL STUDIES

Experimental manipulation in which the outcome can either confirm or negate a stated hypothesis is often considered to be the most convincing method for excluding alternative hypotheses and inferring causation. This method, however, shares a major problem with comparative and observational studies: most evolutionary models do not provide mutually exclusive predictions which can be tested. However, experiments do provide a means of controlling potentially correlated factors and for expanding the phenotypic distribution to detect selection. Andersson's (1982) classic experiment on widowbirds illustrates both of these advantages. By randomizing experimental birds across territories he was able to control for the confounding effect of territory quality. Then, by shortening and lengthening the tail, he expanded the tail length distribution 5 standard deviations to each side of the mean to increase the likelihood of detecting selection.

Furthermore, in the laboratory one can isolate a single process and with a simple experiment, measure its influence. For example, Wilkinson (1987) estimated the response to sexual selection on wing size in *Drosophila melanogaster* by comparing the progeny of randomly paired flies to those which had been allowed to compete for mates in a large mating chamber. By also estimating the selection differential due to a large male mating advantage, qualitative support for an equilibrium model of sexual selection was obtained in which male-male competition (as well as other fitness components) favors large body size while larval viability is enhanced by small body size.

This experiment potentially shares a limitation common to all manipulations: unless the biology of the organism is well known, artifacts of the experiment can create misleading results which prevent extrapolation to natural populations. For example, when Wilkinson (1987) estimated the selection differential due to male-male competition he arbitrarily chose a sex-ratio and density of flies to place in a mating chamber. Either the sex-ratio,

the density, or confinement to a chamber could have altered mating behavior in the flies in unnatural ways. Fortunately, the sexual selection differential was independent of both sex-ratio and density. Furthermore, confinement did not apparently affect mating behavior because independent work by Partridge and her co-workers (personal communication) has demonstrated that larger males have a mating advantage when mating on rotting fruits in the wild, as well as in confinement. Such combinations of field and laboratory experimentation and observation can provide compelling arguments for how sexual selection operates.

As with estimates of selection differentials, correlations between a manipulated trait and an unmeasured trait can lead to misleading results if they are not recognized. A recent experiment on mate choice in guppies by Bischoff et al. (1985) illustrates the danger of correlated traits. In this experiment tail length was manipulated; however, tail length affects display rate which also influences mate choice. In this case, the correlation was detected and cleverly controlled by independently manipulating display rate through temperature change. From these experiments the authors concluded that mate choice by females depended twice as much on variation in tail length as on variation in display rate. Whether this experiment on artificially selected stocks of guppies is representative of how mate choice occurs in natural populations remains to be determined. A field experiment in which control over most variables is reduced but more realism is attained might add further insight and reveal the importance of other sexually selected characters. In this case, field and lab data suggest that male coloration, in addition to display rate and tail length, influences mate choice (Endler 1983; Kodric-Brown 1985; Houde personal communication).

We hope the reader recognizes that while some hypotheses can be discriminated by qualitative outcomes others require quantitative estimation of the relevant parameters. These two processes need not be exclusive. For example, some "good genes" models of sexual selection can be rejected if very little additive genetic variance for fitness exists. It should be noted, however, that the disease-oriented "good genes" model of Hamilton and Zuk (1982) makes the opposite prediction. These models can therefore be tested by estimating the magnitude of the additive genetic variance in total fitness. This subject is considered at greater length by Price et al. (this volume).

In addition to testing for additive genetic variance in fitness, we can suggest several additional areas in which experimentation is sorely needed. One problem is determining the rules for female choice. Whether females have absolute or relative preferences for a specific trait has different consequences for the evolution of the male trait than if females choose the best-of-n males (Seger 1985). Likewise, determining the relative intensity of male-male competition versus female choice of mates can strongly affect the evolutionary outcome. Models of male-male competition predict either a stable equilibrium

point (Parker 1983, Charlesworth 1984) or continual evolution for increasing expression of the male trait (Maynard Smith and Brown 1986). In contrast, some female choice models predict a line of equilibria which can be either neutrally stable or unstable (Lande 1981, Kirkpatrick 1982). If there is a cost to preferences, this line becomes an equilibrium point with the cost minimized. As with the equilibrium line, this point also can be either stable or unstable. Thus, experiments measuring costs associated with female preferences are needed. We emphasize that one cost of mating, such as traveling to a lek to mate, is not relevant to these models; only the cost associated with choosing a particular male is critical.

PROBLEMS FOR FUTURE STUDY

A number of mechanisms other than transmission of additive genetic factors can affect the similarity between parent and offspring. These alternative forms of phenotypic transmission may have nonintuitive and unpredictable effects on both the dynamics and outcomes of a sexual selection process. In addition to maternal effects — such as the correlation between male adult body size and birth weight in red deer (birth weight is determined primarily by maternal condition), genotype environment correlations (i.e., habitat selection), and genotype environment interactions (Via and Lande 1985) — we can envision direct phenotypic transmission across and within generations in females exerting more unanimous preferences through copying or attempts by some males to mimic the behavioral traits of more successful males. In females, this process could increase the intensity of selection on males and could alter the covariances in the transmission matrix (genetic and cultural) when multiple cues are involved. Copying among males might create selection for traits which cannot be copied. The apparent prevalence of exaggerated behavioral displays which appear to be related to vigor in males of lekking species (e.g., Gibson and Bradbury 1985) suggests that these traits may be used by males (and also perhaps preferred by females) because they make copying difficult. This area needs further theoretical work before predictions can be made. Boyd and Richerson (1985) have recently suggested a framework for analyzing phenotypic transmission of quantative characters given specific learning models that might be applied profitably to the theoretical side of the problem. Also, multivariate path models which have been used to detect genotypic and phenotypic transmistion of diseases (Vogler 1985) could be used, in principle, to detect if these effects occur in animals.

REFERENCES

Andersson, M. 1982. Female choice selects for extreme tail length in a widowbird. Nature 299: 818–820.

Arnold, S.J. Wade, M.J. 1984a. On the measurement of natural and sexual selection: applications. Evolution 38: 720–734

Arnold, S.J. and Wade, M.J. 1984b. On the measurement of natural sexual selection: theory. Evolution 38: 709–719.

Bischoff, R.J.; Gould, J.L. and Rubenstein D.I. 1985. Tail size and female choice in the guppy (*Poecilia reticulata*). Behav. Ecol. Sociobiol. 17: 253–255.

Boyd, R., and P.R. Richerson. 1985. Culture and the Evolutionary Process. Chicago: Univ. of Chicago Press.

Charlesworth, B. 1980. Evolution in Age-Structured Populations. Cambridge: Cambridge Univ. Press.

Charlesworth, B. 1984. The cost of phenotypic evolution. Paleobiology 10: 319–327.

Draper, N., and Smith, H. 1981. Applied Regression Analysis, 2nd ed. New York: Wiley.

Endler, J.A. 1983. Natural and sexual selection on color patterns in poeciliid fishes. Environ. Biol. Fish. 9: 173–190.

Endler, J.A. 1986. Natural Selection in the Wild. Princeton: Princeton Univ. Press.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791.

Gibson, R.M. and Bradbury, J.W. 1985. Sexual selection in lekking sage grouse: phenotypic correlates of male mating success. Behav. Ecol. Sociobiol. 18: 117–123.

Hamilton, W.D., and Zuk, M. 1982. Heritable true fitness and bright birds: a role for parasites. Science 218: 386–387.

Harvey, P.H.; Kavanagh, M.; and Clutton-Brock, T.H. 1978. Sexual dimorphism in primate teeth. J. Zool. Lon. 186: 475–486.

Houck, L.D.; Arnold, S.J.; and Thisted, R.A. 1985. A statistical study of mate choice: sexual selection in a plethodontid salamander (*Desmognathus ochrophaeus*). Evolution 39: 370–386.

Howard, R.D. 1987. Reproductive success in two species of anurans. In: Reproductive Success, ed. T.H. Cluton-Brock. Chicago: Univ. of Chicago Press.

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

Kodric-Brown, A. 1985. Female preference and sexual selection for male coloration in the guppy (*Poecilia reticulata*). Behav. Ecol. Sociobiol. 17: 199–205.

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

Lande, R., and Arnold, S.J. 1983. The measurement of selection on correlated characters. Evolution 37: 1210–1226.

Lofsvold, D. 1986. Quantitative genetics of morphological differentiation in *Permomyscus*. I. Tests of the homogeneity of genetic covariance structure among species and subspecies. Evolution 40: 559–573.

Manly, B.F.S. 1985. The Statistics of Natural Selection on Animal Populations. New York: Chapman and Hall.

Maynard Smith, J. and Brown, R.W.L. 1986. Competition and body size. Theor. Pop. Biol. 30: 166–179.

Parker, G.A. 1983. Arms races in evolution — an ESS to the opponent-independent costs game. J. Theor. Bio. 101: 619–648.

Price, T.D.; Grant, P.R.; Gibbs, H.L.; and Boag, P.T. 1984. Recurrent patterns of natural selection in a population of Darwin's finches. Nature 309: 787–789.

Ridley, M. 1983. The explanation of organic diversity: the comparative method and adaptations for mating. Oxford: Oxford Univ. Press.

Seger, J. 1985. Unifying genetic models for the evolution of female choice. Evolution 39: 1185–1193.

Sutherland, W.J. 1985. Measures of sexual selection. Oxford Surveys in Evolutionary Biology 2: 90–101.

Trivers, R.L. 1972. Parental investment and sexual selection. In: Sexual Selection and the Descent of Man, ed. B. Cambell. pp. 139–179. Chicago: Aldine.

Via, S., and Lande, R. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. Evolution 39: 505–522.

Vogler, G. 1985. Multivariate path analysis of familial resemblance. Genetic Epidemiology 2: 35–53.

Wilkinson, G.S. 1987. Equilibrium analysis of sexual selection in *Drosophila melanogaster*. Evolution 41: 11–21.