

Alan Grafen

THERE ARE NOW MANY studies in which data on lifetime reproductive success (LRS) are available, and in this chapter I wish to consider the kind of problems that arise in drawing inferences from such data. By LRS I mean direct observation of the total number of offspring produced in a lifetime by each member of a set of known individuals. The most complete proposed set of analyses is that of the Chicago school, namely Wade and Arnold (1980), Lande and Arnold (1983), and Arnold and Wade (1984a,b). These analyses will be scrutinized with a view to the analysis of adaptation. Various topics will be relevant. I will emphasize two distinctions: the first is between the analysis of adaptations and the detection of selection in progress, and the second is between natural and artificial variation. I will discuss the likely causes of natural variation and stress the necessity of entertaining and relying on hypotheses beyond the limits of the data, considering particular examples of the truism that correlations do not establish cause. Two general conclusions will, I hope, remain with the reader: that LRS data are not a panacea for epistemological complaints, and that a theory-free analysis is likely to be an interest-free analysis.

Another distinction is important in understanding my purpose in this chapter. On the one hand is the general idea of studying animals intensively for a long period, so that whole lifetimes are observed, and attaching importance to the number of offspring individual animals produce. I believe that the application of this idea is a genuine advance in technique and that we understand animals better for it. On the other hand are particular proposals about the use of LRS data and about what inferences may be drawn from variances, correlations, and regressions involving LRS. Wade and Arnold (1980), Clutton-Brock (1983), Lande and Arnold (1983), Arnold (1983a,b), Arnold and Wade (1984a,b), and others contrast the advantages of the LRS approach with disadvantages of other methods. This advocacy is natural for a new and pathbreaking approach to field studies and new methods of analysis. However, as the publication of this book shows, the LRS method is becoming increasingly popular. It is

therefore timely to look hard at its *disadvantages* and think about what *cannot* be inferred from it. My chapter takes the value of the general approach for granted and aims to provide a critical appraisal of some specific ways of using LRS data.

Subsequent sections are organized as follows. Section 28.1 explains the distinction between adaptation and selection in progress and illustrates that different kinds of variation are best used to study them. Section 28.2 explores the likely causes of natural variation and shows how LRS regressions can become difficult to interpret as the hypothesized cause of natural variation becomes more complex. In section 28.3 the value of LRS variance decompositions will be considered. Building on the discussions in the previous sections, section 28.4 evaluates the Chicago school's proposed techniques for analyzing adaptations. In section 28.5 measures of fitness are discussed, in particular the relationship between LRS and Darwinian fitness. Section 28.6 discusses the application of inclusive fitness to data. Brief conclusions are given in section 28.7.

28.1 Adaptation versus Selection in Progress

The distinction between adaptation and selection in progress is simple yet important. An adaptation in the sense of Williams (1966a) is a feature of an organism that can reasonably be said to serve a purpose and is the result of natural selection in the past. Selection in progress is gene frequencies changing now as a result of differences in design between genetically different individuals. An organism may have an adaptation even if selection is not operating on it now. I do not know whether genetic variation is currently affecting the eye in humans, and I do not need to know in order to recognize the eye as an adaptation, to study its function, and to analyze its adaptive value. On the other side, selection in progress may be modifying an existing adaptation, creating a new adaptation, or simply changing the value of a quantitative trait back and forth as generations proceed. A paleontological analogy can be made between adaptations and the bulk of existing fossils, on the one hand, and between selection in progress and present-day corpses, some of which are currently being turned into fossils, on the other.

I am sure this distinction has been widely appreciated before, but I do not think it has been made explicitly in print, presumably because there was no need. Darwin was interested in selection in progress—partly because it was evidence for the mutability of species—and in adaptation. His theories of natural (1859) and sexual (1871) selection are still our only explanations for the existence of organic complexity and adaptation. Wright's four-volume treatise on evolution (Wright 1967–78), is almost entirely concerned with selection in progress and is a fund of information about balanced polymorphisms, selection coefficients, effective population sizes, rates of gene substitution, linkage disequilibrium, dominance, and epistasis. Fisher's book (Fisher 1958), in contrast, is mainly concerned with understanding adaptation and treats selection in progress as an

important but logically subsidiary topic. Its topics include mimicry, sex ratios, extravagant male characters, infanticide, and the heroic virtues. Kimura's neutral theory (Kimura 1983) is about what fraction of genetic variability can be attributed to selection in progress, as opposed to random drift and mutation pressure, and is all but irrelevant to the study of adaptation.

The study of adaptations begins with trying to answer the question why. Why do male red deer have antlers? Why are kingfishers brightly colored? Why are black grouse polygynous? These are the kinds of questions that have always been asked about animals, and the key to them was provided by Darwin (1859). The comparative approach, the theory of evolutionarily stable strategies, and functional morphology are methods of studying certain kinds of adaptations. I suspect that most authors and most readers of this book are interested in explaining adaptations.

The study of selection in progress is also fairly old. Animal breeders who keep track of their stocks are interested in selection in progress. The school of ecological genetics is devoted to the study of selection in progress in nature and has made many fascinating discoveries (Ford 1975). To distinguish between different hypothesized modes of evolution—for example, the Fisher-Haldane mode of a succession of more-or-less independent gene substitutions and the Wrightian shifting balance—it is important to study selection in progress. I believe that the authors and readers of this book are less interested in currently changing gene frequencies than in adaptations.

To illustrate the differences in the kind of study necessary to investigate these two distinct problems, I shall use as an example the spot number on the hind wing of *Maniola jurtina*, a character much studied by ecological geneticists whose work is reviewed by Ford (1975) and Brakefield (1984). The present book is about measuring the reproductive success of individuals, a technique the ecological geneticists did not use. The experiments I propose will therefore be hypothetical, and I do not wish to suggest that they are superior to those in fact used.

Suppose first that we wish to discover the adaptive significance of spot number. The obvious experiment is to paint spots on or off the hind wings and to compare (for example) predation, mating success, and thermoregulation in the groups with different numbers of spots. If we found that spottier butterflies were eaten less often but had the same mating success and temperature control, we could conclude that the function (or better, a function) of the spots was to avoid predation. LRS is not very useful here because it is too all-encompassing a measure. We wish to know *why* the butterfly has the spots, not *how much* more successful more spotted individuals are than less spotted. To understand the adaptive significance of spot number, we want to pin down more exactly the mechanism of advantage. We can make a start by finding if spot number correlates with components of LRS, hoping to find where it is useful to look more closely for the reasons behind the advantage of spots.

If, on the other hand, we are interested in selection in progress on spot number, then we are looking for evidence of gene-frequency changes at the loci that affect spot number. It would be pointless in this case to create variation by painting spots. Ford and colleagues measured the frequency of adult morphs in successive generations and sought to exclude the other possible causes of the observed changes. With data on LRS of individuals, we could find the covariance between LRS and natural spot number. According to the “secondary theorem of natural selection” of Robertson (1966, 1968), the selective change in a character is equal to its genetic covariance with LRS divided by mean LRS. The genetic covariance is equal to the phenotypic covariance (the one we observe) multiplied by the heritability of the character. So by showing that the heritability was not zero (Brakefield 1984), which is to say there is additive genetic variance for spot number, the covariance between spot number and LRS could be used to demonstrate that there was selection in progress at the loci affecting spot number. It would not demonstrate that spot number was part of the causal chain from genes to differential success of individuals, for it could be another, pleiotropic, effect of the genes that determine spot number. In fact, Ford reports that the selection they detected by measuring spot numbers was probably the result of differential parasitism of the butterfly larvae by a hymenopteran. The reason for the correlation between spot number as an adult and susceptibility to parasite attack as a larva is not known.

We have examined distinct ways of studying the two distinct problems. One particularly important difference between them is the kind of variation exploited. It was simplest, though not necessary, to use artificial variation in the study of adaptation, and it was necessary to use natural variation to study selection in progress.

The distinction between adaptation and selection in progress does not mean there are no connections between them. One obvious connection is that current adaptations are the result of selection that was in progress at some time in the past. Another connection arises in some modern theories of sexual selection and the maintenance of sexual reproduction. Hamilton and Zuk (1982) proposed that sexual selection is a defense against certain kinds of parasitism, in which females choose comparatively unparasitized males so that their offspring will in turn be comparatively resistant to parasites. This is a good case for illustrating the distinction between adaptation and selection in progress, because an adaptation in one character (female choice) is based on the continuing existence of selection in progress not in itself, but in another character (resistance to parasites).

28.2 Using Natural Variation to Study Adaptation

It is a truism that correlation does not prove causation, and this is the reason artificial variation was used in the proposed experiment to discover the adaptive significance of spot number. If we control the variation ourselves, we know its cause and can eliminate the possibility that any

other variable correlated with spot number is the true cause of differences in LRS. Had we used natural variation in spot number, it would have been much more difficult to eliminate this possibility. The gain from using natural variation is that we need only measure (components of) LRS and ten characters in order to investigate the adaptive significance of ten characters: to conduct ten experiments to do the same thing would be much more work. It is prudent to be suspicious of bargain offers. Of course all methods of investigation have their drawbacks, and in analyzing LRS data I would certainly perform the LRS analyses first and unleash my suspicions on the results afterward.

That natural variation may mislead us does not mean that it will. In this section I consider some plausible causes of natural variation and ask how they would affect an attempt to discover the adaptive significance of spot number by means of correlations using the natural variation in spot number and LRS (or components of LRS).

Why might spot number vary in nature? The first candidate component of variation in spot number is mutational variance held in check by stabilizing selection. This is variation caused by the occurrence of slightly deleterious mutations that are eliminated so slowly that they remain and contribute to variance for some time. If this is the cause, then the natural variation is just as good as the same quantity of artificial variation. Apart from mutations, no other differences exist between many-spotted and few-spotted butterflies.

If the population is in equilibrium between mutational variance and stabilizing selection, LRS should be higher at the intermediate values of spot number and lower at each extreme. The slope of a linear regression of LRS against spot number would therefore be zero or very small. That LRS is lower at both ends would produce a negative quadratic term in a polynomial regression. Lande and Arnold (1983) discuss using regressions in this way. As before, however, this regression with LRS tells us how advantageous different spot numbers are but says nothing about why. To discover the adaptive significance, it would be of more interest to know the effect of spot number on particular causes such as predation or mating success.

Simple mutational variation checked by stabilizing selection is unlikely to be the sole source of variability in spot number. (Spot number has now become simply a less abstract way of saying "the character of interest." Ford [1975] and Brakefield [1984] should be consulted for facts about the real spot number.) Let us consider the effect of purely environmental variation. Suppose a warm pupation site is an unplannable piece of good luck and that it has an effect on spot number that no genetic combination could reliably produce. Suppose further that those individuals that pupate in a slightly warmer place have more spots while those that pupate in a cooler place have fewer, and that no other variables are affected in this way. Now the linear component of a regression of LRS on spot number need no longer be zero, since it is possible even in equilibrium that the possession of more spots is consistently advantageous. The effect

of spot number on predation could be investigated by regressing predation rate on spot number, and it could be established that the adaptive significance of spot number was as an antipredator device. With this cause of natural variation, correlation with components of LRS is a reliable guide to adaptive significance.

I was careful to say that simple mutational variance was as good as *the same quantity* of artificial variance. We cannot choose how much mutational or environmental variance there is, and the ability to detect effects will be roughly proportional to the amount of variance. Experiment then still has the advantage that the quantity of variance can be chosen. The experimenter of course has problems of his own, including the naturalness of his manipulations and the disturbance they cause.

With these two benign causes of natural variation, the existence of correlation is evidence for causation. However, we cannot be sure that these are the only two causes acting in a particular case. We consider next other possible causes of natural variation that lead to less welcome conclusions. We have discussed additive genetic variation and environmental variation that affect spot number only. Problems arise when the cause of natural variation means that other characters are correlated with spot number.

The first malignant cause is pleiotropic mutational variance. If mutations affect a number of characters, then those that are disadvantageous in all their effects will be more quickly eliminated than those that are advantageous in some effects and slightly disadvantageous on balance. This implies that the mutations that persist will tend to produce a negative correlation between those characters that are positively associated with fitness (Falconer 1981, 300). Consequently an individual who has a disadvantageous spot number will tend to have, say, an advantageous wingspan, and similarly in reverse. By using natural variation in spot number we will measure the combined effects of spot number and the correlated part of wingspan. The consequence of this is to diminish the strength of associations between spot number and LRS and, to a lesser extent, with its components. Wright (1968–78, 1:61) states that “the available evidence indicates that pleiotropy is virtually universal.” We define characters according to our interests, and we can hardly expect the same divisions to be observed by the biochemistry of development.

The second factor is the “silver spoon” effect, a particularly likely example of a common environmental cause that influences many characters. I define the silver spoon effect as positive correlations between characters in the adult that are positively associated with fitness, brought about by the common underlying cause of favorable or unfavorable environmental events during development. It differs from the example of warm pupation sites only in that many characters are affected, and this is exactly what causes problems. Let us switch examples and think of juvenile red deer. There are many accidents of childhood, including the extent of parental care, the severity of winters, and the chance occurrence of epidemics. They are likely to produce some individuals that are

vigorous and fit in both senses of the word and others that are weak and unfit. The syndrome of characters called "quality" may often be determined by the silver spoon effect. If this does occur, we will observe a positive correlation between LRS and a whole host of characters, most of which will be correlated with each other. The components of LRS would also tend to be intercorrelated. To discover the adaptive significance of antler size, we would have to partial out the effects of all the correlated characters. The effect of early food availability is documented for domesticated animals by Sadleir (1969). Lande (1982) discusses the likelihood of negative genetic correlations but positive phenotypic correlations between characters positively associated with fitness.

The third factor has been called "making the best of a bad job" (Dawkins 1980). This applies particularly to behavioral characters, and the idea is that the observed differences may be a response by individuals to their particular circumstances. These preexisting circumstances are then confounded in the analysis with the character. This effect is fairly subtle, and I can think of no plausible animal example. I believe this is because we would have to understand a species rather better than we generally do to recognize it. As a human analogy, consider a group of children who have to catch a bus to school in the morning. Some will walk and some will run. Will the walkers or the runners be more likely to catch the bus? I would bet heavily on the walkers. For any given child on a given morning, running would increase the chance of arriving on time. However, those who choose to run will be those who got up late and are at risk of missing the bus, and those who choose to walk are those who got up early and will catch the bus with no difficulty. A correlation of speed from home to bus with catching the bus, naively interpreted, would therefore reveal that running made a child less likely to catch the bus. The difference in behavior is the result of a decision that produces a correlation between the behavior and an underlying variable. Particularly for reproductive strategies, we must suspect that differences in behavior are "adaptive reactions" of this sort of preexisting relevant physical and social circumstances.

These adaptive reactions are in one sense just correlations between characters, but I believe they have a special power to mislead, for three reasons. First, the individual animal is likely to have much more information about its state of health and nutrition than we have. Second, in deciding between fighting or reproductive strategies, a small difference in state can cause a large difference in the strategy chosen, particularly if the options available are discrete. Third, it is natural for us to ascribe differences in success to correlated differences in behavior, perhaps because we think of our own behavior as causing but uncaused. Important but poorly observed information about an animal's state may therefore be converted into easily observed behavior by the animal's adaptive reaction to that state. We would then assign to the behavior what are really consequences of the underlying state.

This completes my catalog of malignant causes of natural variation

that would invalidate simple inferences about adaptive significance, though the catalog is by no means exhaustive. In no case was a correlation with LRS itself of much importance. If the purpose of the exercise is to understand the adaptive significance of a character, correlations with predation rate or harem size or fighting ability are more to the point. This is because an adaptive explanation is about why something is advantageous, not just how advantageous it is. The main conclusion of this section is that what a correlation with LRS or its components tells you depends on the causes of the natural variation in the character. It will not always be easy to discover those causes.

28.3 The Partitioning of Natural Variation in LRS

The stages between which variation may be divided can be illustrated using the example of *Maniola jurтина*. An egg has a certain chance of surviving to become a larva, which has a certain chance of surviving to become a pupa, which has a certain chance of surviving to become an adult. Once adult, its reproductive success is the product of longevity and fecundity per unit time. Individuals may differ to varying degrees in their success in these five stages.

There are a number of proposals for partitioning variance in LRS between different stages of the life cycle and reproductive cycle (Wade and Arnold 1980; Arnold and Wade 1984a,b; Brown, this volume, chap. 27), similar to the key factor analysis of Varley and Gradwell (Varley and Gradwell 1960; Southwood 1978) and an extension of work by Crow (1958). Crow suggested a partition of his "index of selection," and Jacquard (1974, 322–30) gives a number of applications of this partition to human populations. The idea is that by decomposing the variance between survival as egg, larva, and pupa, adult longevity, and fecundity, we can identify the stages at which individuals differ greatly. The total variance in LRS, its difference between the sexes, or the variance among males in mating success are sometimes thought to be useful in deriving a measure of sexual selection (Wade 1979; Payne 1979; Wade and Arnold 1980). My aim in this section is to discuss what significance we can attach to these variances under different possible sets of causes of natural variation, when our purpose is to explain adaptation. Since the various causes are explained in the previous section, I shall not explain them again.

The first candidate cause of variability is mutational variation held in check by stabilizing selection. Surprisingly, the amount of genetic variability in fitness in this case seems not to have been investigated directly. Turelli (1985), reconciling conflicting results of Latter (1960) and Bulmer (1972, 1980) on the one hand and Lande (1976) on the other, shows that the variance of an ordinary character will be greater with higher mutability of the loci involved and with weaker stabilizing selection. Fitness is no ordinary character, and it appears likely from Turelli's finding that genetic variance in fitness will be greater with higher mutability and strong stabilizing selection. Thus the genetic variance in fitness at a stage does

increase with the strength of stabilizing selection, but comparisons between stages are vitiated unless we know the mutabilities of the sets of loci affecting fitness at the different stages.

Once environmental variation is allowed, however, the inference from variance at a stage about the strength of selection acting at that stage becomes even weaker. The larger part of variance in success in the various stages of life may well be environmental, as suggested by the effect of food availability during youth on subsequent success in domesticated animals (Sadleir 1969) and the low heritabilities generally reported for characters strongly related to fitness (Falconer 1981).

The claim made by Wade and Arnold (1980) is that the variance assigned to a stage sets an upper limit to the selection that can go on at that stage. (They use language differently: By "selection" they mean the phenotypic change within a generation, and by "response to selection" they mean the phenotypic change between generations.) This upper limit is to directional selection in progress, as distinct from stabilizing selection in progress and as distinct from adaptive value. Upper limits are likely to be good guides in two kinds of circumstances. One is when the upper limit is likely to be nearly attained, as would be the case if the internal volume of a glove were used as a guide to the volume of the hand that wears it. (It is a poor guide to the quantity of air trapped inside the glove.) The other case is where the upper limit contains two elements in roughly constant proportions, one of which is the value of interest. For example, the volume of a container filled with air is a good guide to the volume of free oxygen it contains because fluctuations in pressure and in relative concentrations of the different atmospheric gases are comparatively small. We have seen that the limit to selection in progress is probably not very nearly attained in general, and I know of no reason to believe that genetic and environmental variation should be of roughly constant proportions in different characters or life stages. But for the purposes of understanding adaptation, the most important point is that the variance at a stage, if it measures anything at all, measures something to do with selection in progress, not with adaptive value.

Let us now turn to the significance of variance in LRS within sexes, and in particular to whether it can be used to detect or measure sexual selection, as has been suggested by Payne (1979), Wade (1979), and Wade and Arnold (1980). In this section I consider questions about the adaptive significance of character, not about selection in progress. The kind of question we are trying to answer is if measuring variance in LRS can help us decide whether antlers in red deer are the product of sexual selection. I shall argue that the answer is no, that the natural variance in LRS is entirely irrelevant to this problem. The substance of the argument follows from the distinction between adaptive significance and selection in progress. Variance now in LRS may be relevant to sexual selection in progress, but it need have nothing whatever to do with the nature of selection in the past.

Suppose that male red deer live for ten years without mating, in their eleventh year develop large antlers and defend a harem, and then die. Suppose what is even more unlikely, that each male gains exactly the same number of matings in its life. We cannot conclude that antlers are not sexually selected, despite the zero variance in LRS between males. There will also be zero variance in components of LRS such as number of mates. It may be that all males have the same size antlers and that any variant would be punished by being unable to hold a harem (say because females would leave a small-antlered male, while a large-antlered male would be too encumbered to fight). In this case the adaptive significance of antlers is to attract females, and antlers are certainly the result of sexual selection. The point of this extreme example is to show that current variance in LRS in males and sexual selection are in principle quite separate.

It is interesting to compare these arguments with those of Arnold (1983a, 67), who discusses the analysis of sexual selection using variances in LRS and its components. He states that the aim of the method is “merely to characterise sexual selection by its statistical effects on phenotypic characters within a generation,” that this “of course, tells us nothing about how selection actually worked in the past,” and that “the goal is to understand the process of sexual selection by direct measurement of its contemporary impact.” It seems fairly clear that the focus of interest here is measurement of selection in progress, not the analysis of adaptation.

I believe that most evolutionists and behaviorists would say they were primarily interested in adaptation, as opposed to selection in progress, once the distinction is brought to their attention. Their primary concern is why male red deer have such big antlers, not whether there are genes now changing in frequency that affect antler size. Of course the gene-frequency changes are interesting, just as the physiological mechanisms of antler creation and the morphological modifications for bearing the antlers are interesting. But they are not central.

In principle, therefore, variance in LRS and analysis of adaptations due to sexual selection are separate. It is likely, though, that variance among males in LRS will be higher in sexually selected species. This is because monogamous species tend to be less strongly sexually selected and will probably have a lower variance in LRS. A positive correlation between variance in LRS and degree of sexual selection would therefore not be surprising, and it would not refute the logical point that one is neither an explanation nor a measure of the other.

28.4 The Proposed Methods of the Chicago School

In this section I will consider the methods proposed by Wade and Arnold (1980), Lande and Arnold (1983), and Arnold and Wade (1984a,b). Three problems with their methods arise from earlier discussion, but it is not my intention to discourage the application of their methods. (Experiments of

course have problems of their own.) With suitable data I would certainly use them, because selection in progress is interesting and because they provide hints about adaptation. My aim is only to give some cautions about the interpretation of their results.

In making claims for their methods, Arnold, Wade, and Lande do not always distinguish clearly between the analysis of adaptation and the detection of selection in progress. It is clear, however, that the design of their methods is to detect selection in progress. Witness the connection, of which they make much, between their methods and the dynamic equations of evolutionary change, and the idea that they are measuring the *potential for selection*, of which only a certain fraction becomes the *response to selection*. Their technique could be compared with other, more direct, techniques for detecting selection in progress, such as those described by Dobzhansky (1970) and Ford (1975). More relevant here is that some of their proposed methods may be useful for the study of adaptation while others may not, and those that are useful for this distinct purpose will require different caveats and face different problems of interpretation.

The variance partitions, magnitudes of variances in LRS, and regressions of LRS on characters are valuable additions to our knowledge of the natural history of a species. However, they tell us little about selection in progress unless we can estimate the environmental causes of variation at each stage, and they reveal little if anything about adaptation. Like the existence of sexual dimorphism, they may *suggest* that males in a species are sexually selected but cannot be used to *show* that they are. The technique that *can* tell us about adaptation is the regression of components of LRS on characters.

Let us consider the multivariate regression techniques proposed by Lande and Arnold (1983) and developed by Arnold and Wade (1984a,b). Using natural variation, they estimate a number of quantities for each character that are relevant to selection in progress. Three of them are the opportunity for selection (the variance in the character), the selection differential (the covariance of the character with an LRS component), and the selection gradient (the slope of the best-fitting straight line relating the LRS component to the character, holding other variables constant). The first two are not likely to be relevant to adaptation, but the third is. The meaning of a selection gradient may be seen as follows. Suppose antler length is measured in inches and LRS is measured in percentage of the mean LRS. Then a selection gradient for antler length of three means that if two stags are the same in every respect except that the first one's antlers are longer by one inch, then the first will have an LRS that is (on average) higher by three. "The same in every respect" means the same in every other character included in the analysis, not the same in every other character that affects fitness, which is of course the unattainable ideal. It will be seen that the technique avoids many of the problems of the correlation of characters.

How valuable is this selection gradient in the analysis of adaptation? The problems are two we examined in section 28.2, namely, correlation of

characters and the amount of natural variation and the additional problem of *incompleteness*.

One strength of the Lande and Arnold (1983) analysis is that correlation between characters that are included in the multivariate regression will not confound the estimates of selection gradients, but as they point out, the estimates may be confounded by missing characters. How likely is it that all or most relevant characters will be included? In *Maniola jurtina*, the relevant correlate of adult spot number was thought to be parasitism as a larva. A character may be influenced by pleiotropic loci with (to a student of behavior) obscure physiological effects. The silver spoon effect described in section 28.2 is likely to be the result of unobserved environmental effects. To the extent that we can find surrogates for these effects, such as weight at one year old, their influence on the selection gradients of other characters can be removed; but to the extent that those surrogates are imperfect, that influence will remain. The “adaptive reaction” of section 28.2 is also a problem for the reasons explained there, so long as we remain ignorant of the behavioral rules the animals use. Clear and observable differences in behavior may be consequences of hard to observe underlying differences in health or vigor that are the true causes of variation in success.

The next problem in the interpretation of selection gradients is the quantity of natural variation. If this is small, then it will be difficult to detect an effect that genuinely exists. The variation that is relevant in the Lande and Arnold (1983) analysis is not simply the variance in a character, but only the part of it that cannot be predicted by variation in other characters. A selection gradient insignificantly different from zero may occur because the character is selectively irrelevant or because the character does not have enough independent variation. Andersson (1982) found that male widowbirds with larger tails were more successful than those with smaller tails, using an experiment. The experimentally manipulated tails were about five standard deviations from the mean. To detect the same size of difference using natural variation, even supposing it to be benign, would have required much larger sample sizes.

The third problem is that the data set may be incomplete. The most likely cause is that early mortality may prevent the measurement of important variables. In *Maniola jurtina*, the spot number of individuals parasitized as larvae by hymenopterans is believed to be different from those that escape parasitism. By using natural variation, we cannot discover this, because the spot number can be measured only in adults. The penalty for a male red deer planning to have a large body size may be that by allocating resources to growth instead of to defense against disease, it is more likely to die as a juvenile. So the “largest” males die young and cannot have their adult size measured. There will often be an “invisible fraction” of individuals who do not appear in the regression but should. The close connection between the multivariate regression and the dynamic equations of evolutionary change depends on the completeness of the sample of individuals in the regression. Sex-limited characters may be

genetically correlated with other characters in the sex in which they are not expressed, and if this is so, then a whole sex may belong to this "invisible fraction."

In conclusion, the Chicago school's methods are interesting analyses to perform on suitable data. They provide clues about the adaptive significance of characters and are interesting in their own right as a contribution to the natural history of the species. For the reasons given above, the regression results must be treated with caution when applied to the analysis of adaptation. The variance partition is not a measure of the extent to which the behavior or morphology of a species may be attributed to sexual selection.

28.5 Measurements of Fitness

In this section I discuss ways of measuring fitness. A fieldworker who has measured the LRS of a number of individuals may like to feel that the bulk of the work is over, that it is necessary only to find the correct statistical technique, and a convenient computer program to implement it, before the biological significance of the study will be revealed. The feeling is tempting because fitness is the central concept of evolutionary biology, and LRS seems to be fitness. I hope to convince the reader that LRS is in an important sense not Darwinian fitness. In the next section the important case of inclusive fitness is discussed.

Following Williams (1966a, 158), we may interpret Darwinian fitness as a property of a design, not of an individual, and particularly not as the number of offspring an individual happens to produce. Rather it is the number of offspring that a given design of animal will on average produce. Suppose a new mutation occurs, and the reproductive success of its few bearers fluctuates greatly between generations because of random factors peculiar to individuals, such as being struck by lightning. If Darwinian fitness *were* LRS, we would be forced to say that the Darwinian fitness of the new design fluctuated greatly between generations. Williams would want to say this only if the environment changed so much that a bad design in one year was a good design in the next. Too much should not be made of this distinction, but it is important that we are trying to estimate something we cannot directly observe. With a large enough sample, we could get as close as we like.

LRS is clearly a very important measurement in estimating Darwinian fitness, but this does not mean they are the same thing. Darwinian fitness is the property of a design, while LRS is a property of an individual. LRS can be known exactly, while we estimate Darwinian fitness with error. The difficulties encountered in previous sections with ascribing differences in LRS to differences in a character are difficulties in estimating Darwinian fitness from LRS. Any difficulty in discovering LRS is purely observational.

The crucial extra stage in estimating Darwinian fitness from LRS comes in deciding what character is to be studied. Plotting mean LRS

against size for range of size-class intervals will estimate the Darwinian fitness of different-sized animals. The Lande and Arnold (1983) regression is a more sophisticated version of this, and their selection gradient will estimate the marginal effect on Darwinian fitness of an increase in size. Special problems arise when we try to apply this method to social behavior, as a special case of the "invisible fraction" discussed in the previous section.

Suppose we wished to determine whether an altruistic act was favored by selection or not, by means of the LRS method. The obvious way to proceed is to find the average LRS of altruists and compare this with the average LRS of nonaltruists. The problem here is that altruism will often be conditionally expressed, and thus many of the apparent nonaltruists may be altruists for whom the appropriate circumstances did not arise. If the "genuine" nonaltruists and the "altruistic" nonaltruists benefited from the altruism of others in the same way, this would be unimportant. However, if altruism is directed at relatives, then "altruistic" nonaltruists will be the recipients of help more often than "genuine" nonaltruists. The result will be that the success of nonaltruists will be overestimated, and so the selective advantage of altruism will be underestimated. I have discussed this elsewhere (Grafen 1984, 1985). In the next section I discuss how the advantage of altruism could be estimated. The main point for now is that LRS fails to give the right answer when we cannot identify the sets of individuals whose average LRS would give the correct estimates of Darwinian fitness.

Next in this section on measurements of fitness, a few comments on the problem of overlapping generations. By fitness I have meant (expected) lifetime number of offspring. Fisher (1958) and, in more detail, Charlesworth (1980) have pointed out that offspring further in the future should be discounted by a factor that depends on the growth rate of the population. Selection favors early reproduction in a growing population and late reproduction in a shrinking population, if the total number of the individual's offspring must remain constant. Equally, in a static population, selection is indifferent between early and late reproduction if the total number of offspring remains constant.

Populations that are neither extinct nor continually increasing must have a long-run average growth rate of nearly zero. If individuals can detect and respond in their behavior to the growth or shrinking of the population, and if the number of offspring they can produce is independent of when they are produced, then they may indeed be selected to value differently offspring produced at different times. If, on the other hand, they lack this adaptive flexibility, or the population never grows or shrinks very fast anyway, then they will tend to behave appropriately to the long-run average growth rate of nearly zero. This is the justification for the simplifying assumption that expected number of offspring is the proper measure of fitness.

Finally, at what stage should fitness be measured? I have used expected number of *offspring*. Why not grandoffspring, or descendants at

some hypothetical time in the infinite future? Suppose our aim is to understand the evolution of a character, which for simplicity we shall assume to take one of two possible forms. Then our choice of fitness measure must satisfy two properties: that we can work out the fitness measure from our knowledge of or assumptions about the character, and that if one form has a higher fitness than other, it will increase in frequency in the population. Subject to these conditions, the simpler the fitness measure the better. The use of number of offspring is suitable if the character in *the adult* does not affect its offspring's survival or fitness after the age of counting.

A very convenient assumption in measures of fitness is independence of control. When we count offspring as a measure of the parent's fitness, we implicitly assume that up to the stage of counting the survival of an offspring does not depend on its own phenotype. This independence of control is unlikely to be fulfilled by a simple measure of fitness that extends beyond offspring to future generations. This assumption is not as restrictive as it seems, for the independence of phenotypic effects applies only to the character being studied, not to the phenotype as a whole. Cheverud (1984) presents a model with cross-generational effects.

The best measure of fitness is then determined by the character we have in mind rather than by any deep principles. This is another illustration of the fact that Darwinian fitness and LRS are different, because Darwinian fitness should be computed differently for different characters according to the stage of reproduction at which they take effect. For example, a character that determines survival as a juvenile will be best studied using survival to breed as a measure of fitness. Using number of offspring surviving to breed would be assessing the phenotype of the offspring rather than the phenotype of the adult. Contrariwise, we might study a character that affected the survival to first breeding of the individual's offspring. Then number of surviving offspring is the appropriate measure of fitness because it assesses the individual's phenotype and because the offspring's own value of the character does not confound the analysis. Using number of offspring born as a measure of fitness does justice to neither of the examples, although it could be used. One individual's fitness would then depend on another individual's phenotype, which is the circumstance for which Hamilton (1963, 1964, 1970) developed his theory of inclusive fitness. Simplicity suggests using an individual measure where possible.

One much-discussed character for which a fitness measure would have to go beyond offspring is sex ratio. In simple models, grandoffspring are best counted. The effect of an individual's phenotype on *number* of descendants is not apparent until the grandoffspring generation, but independence of control is maintained because the offspring's phenotypes do not take effect until the generation after that. The point, then, is that the best place to measure fitness depends on the character being studied. If no place can be found that satisfies independence of control, then more complicated modeling is necessary.

I conclude this section by repeating its important points. There is an important sense in which LRS is not Darwinian fitness, but rather Darwinian fitness is an abstraction that may be estimated, not always straightforwardly, from LRS. LRS data are not, as it may seem at first thought, an infallible oracle for answering important evolutionary questions about a species. Indeed, we have yet to come across a question for which they are decisive, and in every case it would be valuable to have other kinds of data as well.

28.6 Inclusive Fitness

In this section I discuss the application of inclusive fitness (Hamilton 1964, 1970) to data. Elsewhere I have discussed errors in alleged definitions of inclusive fitness and also worked out an example from data as an illustration of how inclusive fitness should be applied (Grafen 1982, 1984). Here I shall repeat the most salient points and explain their relation to the dichotomy of natural versus artificial variation. First, it is better to apply Hamilton's rule (shortly to be defined) than inclusive fitness to data. The two are formally the same, but as historical fact Hamilton's rule has been correctly applied, whereas inclusive fitness has been incorrectly applied to data (Grafen 1984). The reason is that applying Hamilton's rule suggests to the user's mind the correct logic of differences.

To apply Hamilton's rule, we need to estimate the three terms that appear in it. The rule is that "a social action is favored by natural selection if $rb - c > 0$ ": r is the relatedness of the donor to the recipient, b is the benefit to the recipient of the action and c is the cost of the action to the donor. I have discussed relatedness and derived and defended Hamilton's rule as an evolutionary principle at some length in a recent paper (Grafen 1985). In field studies of the sort in this book, relatedness will usually be known through observed common ancestry, or sometimes it may be estimated from electrophoretic data (Pamilo and Crozier 1982). Here I will concentrate instead on the estimation of b and c .

The benefit and cost are differences in number of offspring that are caused by the performance of the social action. The meaning of b therefore involves something we do not observe; b is the number added to the recipient's number of offspring by the social action. We need to know how many the recipient would have had with the social action, and how many without. Obviously we cannot measure both in the same individual, but this does not mean we are helpless. We must use our understanding of the animal and its actions to estimate what that difference is. If, for example, the social action is help in surviving the winter, then we may know the reproductive value of an animal that survives the winter of a given age and the chance of survival with and without help by observing the survival of individuals with different fat reserves.

The complications involved, and the assumptions that need to be made, may at first seem a disadvantage in applying Hamilton's rule. This should really be turned on its head. If your aim is to understand the

adaptive significance of a social action, then these “complications” are exactly the constituents of that understanding. If you do not know the effect on survival of help, then you cannot possibly understand the adaptive significance of that help. The value of Hamilton’s rule is that it draws attention to three essential quantities, and says: If you can estimate r , b , and c , then you understand the adaptive significance of the social action. If you cannot estimate them, it is unlikely that you understand the adaptive significance of the social action.

The conclusion of more general interest is that Hamilton’s rule involves working out what would have happened in circumstances we do not observe. It is therefore an example of an artificial variation technique, though not necessarily an experimental one. These workings out are exactly what is necessary to discover the adaptive significance of the social action studied.

28.7 Conclusions

The primary purpose of the contributors to this book seems to me to be the analysis of adaptation rather than the detection or measurement of the changes in gene frequency that constitute selection in progress. The logically most straightforward way to understand an adaptation is to perform an experiment, as Andersson (1982) did with his widowbirds. An alternative is to use LRS data. Problems arise in this LRS method because it relies on natural variation in characters, when the cause of that natural variation is usually unknown. Correlations with relevant characters not included in the analysis, the quantity of variation, and the omission of dead individuals from the sample of those measured are three such problems.

The intensive kind of study that produces LRS data is likely to uncover at the same time many interesting facts about the species. The LRS data themselves add to our knowledge of the species’ natural history in an important way. They are not, however, the answers to a theoretician’s prayer. The methods of analysis of LRS data proposed by Wade and Arnold (1980), Lande and Arnold (1983), and Arnold and Wade (1984a,b) seem primarily designed to study selection in progress—that is to say, gene frequencies changing now rather than adaptation. Their multivariate regression method is a very interesting one to perform even for the study of adaptation but is often difficult to interpret because of its use of natural variation.

It is necessary to understand the causes of natural variation in order to draw conclusions from natural LRS data about adaptations and about their attribution to natural or sexual selection.

28.8 Acknowledgments

I am grateful to T. H. Clutton-Brock, M. Ridley, M. S. Dawkins, J. A. Endler, S. J. Arnold, M. A. A. Otronen, W. B. Sutherland, N. E. Pierce, J. R. Krebs, and S. A. Frank for helpful comments and discussions. The work was supported by a Julian Huxley Junior Research Fellowship at Balliol College, Oxford, and by a Royal Society 1983 University Research Fellowship.