



The “averaging fallacy” and the levels of selection

SAMIR OKASHA

*Department of Philosophy, University of Bristol, Bristol BS8 1TB, UK
(e-mail: Samir.Okasha@bristol.ac.uk; phone: +44 117 928 7610)*

Received 24 January 2003; accepted in revised form 19 June 2003

Key Words: Averaging fallacy, Genic selection, Group selection, Levels of selection

Abstract. This paper compares two well-known arguments in the units of selection literature, one due to Sober and Lewontin (1982), the other due to Sober and Wilson (1998). Both arguments concern the legitimacy of “averaging” fitness values across contexts and making inferences about the level of selection on that basis. The first three sections of the paper shows that the two arguments are incompatible if taken at face value, their apparent similarity notwithstanding. If we accept Sober and Lewontin’s criterion for when averaging genic fitnesses across diploid genotypes is illegitimate, we cannot accept Sober and Wilson’s criterion for when averaging individual fitnesses across groups is illegitimate, and vice versa. The final section suggests a possible way of reconciling the two arguments, by invoking an ambiguity in the concept of “genic selection”.

Introduction

In their recent examination of the units of selection problem, Elliott Sober and David Sloan Wilson (1998) argue that the importance of group selection in evolution has been obscured because of what they call the “averaging fallacy”. Those who commit the averaging fallacy are prone to claim that certain selection processes are driven by purely individual-level selection, when in fact they involve at least a component of group selection. Indeed, Sober and Wilson maintain that “the controversy over group selection and altruism in biology can be largely resolved simply by avoiding the averaging fallacy” (1998: 34).

The averaging fallacy occurs as follows. A population is sub-divided into a number of groups each of which contains individuals of two types, in varying proportions; overall (population-wide) fitness values for the two individual-types are calculated by averaging the fitness of each type across all the groups in which it occurs; it is then claimed that the individual-type with the higher overall fitness increases in frequency by individual selection. This is fallacious, according to Sober and Wilson, because it ignores the effects of group structure. It may be the case that the individual-type that is fittest

overall is actually *less* fit within each group, but that groups in which that individual-type is common are fitter than groups in which it is rare. If so, then the individual-type is subject to both group selection and individual selection, Sober and Wilson argue. Within each group, individual selection works against the type in question; but this is counteracted by differential group productivity, i.e., group selection, which favours the type. Averaging fitness values across all groups obscures the multi-level nature of the selection process.

Sober and Wilson's argument is interestingly reminiscent of an earlier argument of Sober and Lewontin (1982), much discussed in the units of selection literature of the 1980s. Sober and Lewontin's argument was directed against the genic selectionism of Dawkins (1976) and GC Williams (1966); it purported to show that in certain circumstances, the gene cannot be regarded as the unit of selection, on pain of obscuring important information about the causal structure of the selection process. One such circumstance is *heterozygote superiority*, according to Sober and Lewontin. In a standard one-locus two-allele population genetics model, if the heterozygote has higher fitness than both homozygotes, then selection occurs at the level of the diploid genotype, not the individual gene, they argued. In such a model, genic fitness¹ values can be calculated by averaging over all the genotypes in the population, and these values will predict the evolution of the system, i.e., will predict which allele will increase in frequency. But this does not mean that the gene is the unit of selection, according to Sober and Lewontin – the genic fitness coefficients are statistical artifacts, not reflections of the underlying dynamics. Again, averaging obscures the causal facts about the level at which selection is operating, making it appear as if low-level selection is doing all the work, when in fact it is not. An extended version of this argument is also found in Sober (1984).

It is natural to think that there is a parallel between the Sober and Wilson (1998) argument and the Sober and Lewontin (1982) argument. This is certainly the view of many commentators. For example, Sterelny and Griffiths (1999), discussing Sober and Wilson's defence of group selection, say that the averaging fallacy is "of the same kind that gene selectionists have been accused of perpetrating" (p. 168). There is some evidence that this was the view of Sober and Lewontin too; a footnote to their 1982 paper says that "the averaging of effects can also be used to foster the illusion that a group selection process is really just a case of individual selection" (p. 129n). There is also evidence that Sober and Wilson regard the two arguments as of a piece; a footnote in their recent book directs the reader to both Sober and Lewontin (1982) and Sober (1984) for "further philosophical discussion concerning how averaging can obscure causal facts" (1998: 341n).

This paper demonstrates that the two “averaging” arguments are actually incompatible if taken at face value, their apparent structural similarity notwithstanding, and explains why. I stress that my purpose is *not* to convict Sober (or anyone else) of inconsistency, but rather to clarify the logical relation between two important and influential arguments which seem very similar, and which, taken singly, both seem plausible. In the final section, I suggest a possible way of reconciling the two arguments. If we distinguish between genic selection as a *causal process*, and genic selection as a useful *perspective* from which to view evolution, then we can interpret the two “averaging” arguments as addressing different questions, hence eliminating the incompatibility. I start by looking at Sober and Wilson’s (1998) argument in more detail.

Multi-level selection, the evolution of altruism and the averaging fallacy

To fix ideas, we consider a simple model for the evolution of altruism, identical in all relevant respects to the one Sober and Wilson (1998) present in *Unto Others*, which itself derives ultimately from Wilson’s (1975) trait-group model. There are two types of organism in the population, selfish (S) and altruist (A). Reproduction is assumed to be asexual, and like begets like: the offspring of selfish individuals are selfish, and likewise for altruists. The population is divided into a number of sub-groups which differ in their proportion of altruists. Organisms spend part of their life-cycle in these groups, during which fitness-affecting interactions take place, followed by reproduction. After reproducing, the organisms die immediately. The groups then dissolve, and the offspring organisms blend into the global population and mix up. New groups are then formed, and the cycle repeats. This is shown in the diagram below.

Now consider fitnesses. The fitness of any organism depends on whether it is selfish or altruistic, and also on which group it is in. The higher the proportion of altruists in an organism’s group, the higher that organism’s fitness – for it is on the receiving end of more altruistic actions. *Within* any group, selfish organisms are obviously fitter than altruists – they enjoy the benefits of others’ altruism, without incurring any of the costs. Following Sober and Wilson (with minor simplifications and changes in notation), we can model the situation using the following fitness functions:

$$W_A(x) = 1 - c + b(x - 1)$$

$$W_S(x) = 1 + bx$$

$W_A(x)$ means the fitness of an altruist in a group containing x altruists; $W_S(x)$ means the fitness of a selfish organism in such a group. Each organism has a

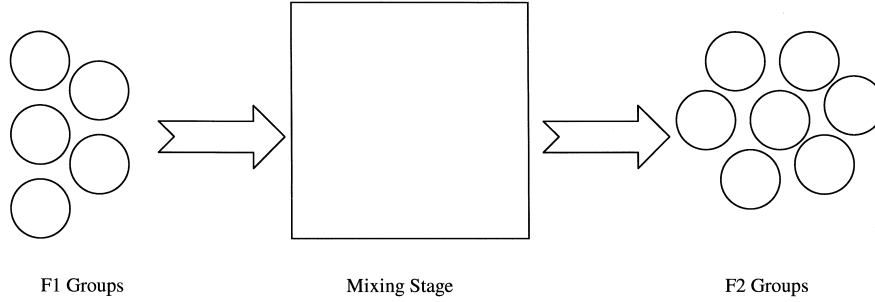


Figure 1. Group selection model for the evolution of altruism.

“baseline” fitness of 1; c denotes the fitness cost of behaving altruistically, b the fitness benefit that an organism receives from the presence of each other altruist in its group. We assume $b > 0$ and $c > 0$; it follows that $W_S(x) > W_A(x)$ for all x , i.e., within each group, selfish organisms are fitter than altruists.

For simplicity, we assume that the population of size N is divided into groups of equal size n ; hence there are N/n groups in total. By “group fitness” we mean total group productivity, i.e., the total number of individuals the group contributes to the next generation. So

$$W_G(x) = xW_A(x) + (n - x)W_S(x)$$

where $W_G(x)$ is the fitness of a group containing x altruists.

$$\text{This expression simplifies to } W_G(x) = x(b(n - 1) - c) + n \quad (1)$$

Note that since b and c are constants, group fitness is a linear function of the proportion of altruists in the group. So long as $b > c/(n - 1)$, this function is monotonic increasing – raising the proportion of altruists raises the fitness of the group.² Sober and Wilson do not explicitly discuss the condition $b > c/(n - 1)$, but they assume throughout that group fitness is enhanced by having a greater proportion of altruists in the group, so they are obviously assuming that the condition is satisfied; in any case, $b > c/(n - 1)$ is a necessary (though not sufficient) condition for altruism to evolve.

To calculate the average fitness of altruists, we average over all the group-types in the population that contain altruists, weighting by the frequency of each group-type, and the proportion of altruists *within* each group-type; similarly for the average fitness of selfish organisms. Therefore:

$$W_A = \frac{\sum_1^n xW_A(x)f(x)}{\sum_1^n xf(x)}$$

$$W_S = \frac{\sum_0^{n-1} (n - x)W_S(x)f(x)}{\sum_0^{n-1} (n - x)f(x)}$$

where $f(i)$ means the frequency of groups containing exactly i altruists and $(n - i)$ selfish types, in the overall metapopulation of groups.

The condition for the increase of altruism in the population as a whole is $W_A > W_S$ – on average, altruists must be fitter than their selfish counterparts. So if we can compute the values of W_A and W_S , we can predict the evolution of the system. Nonetheless, it is wrong to attribute the spread of altruism, in any particular case, to the fact that $W_A > W_S$, according to Sober and Wilson. This is to commit the averaging fallacy. It ignores the fact that altruists are actually *less* fit within each group, hence are selected against by within-group individual selection, but increase in frequency because of the positive correlation between group fitness and proportion of altruists in the group, i.e., are selected for by between-group selection. Averaging obscures the causal processes involved in the spread of altruism, they insist. Only a “multi-level” perspective, which recognizes selection at both the individual and the group level, is true to the causal facts of the situation.

Intuitively, Sober and Wilson’s argument is quite compelling. A proper understanding of why altruism increases in frequency, if it does, must surely allude to the group structure of the population, and the greater productivity of groups containing more altruists. Averaging over all groups and saying that altruism is “fitter overall”, though computationally useful, does indeed seem to mislead about the levels of selection, and obscure the true causal structure of the selection process.

Sober and Wilson do not provide a *general* characterization of the averaging fallacy; they explain it on the basis of examples. This prompts the question: exactly which feature of the above model makes it fallacious to average fitnesses across all groups? There are three possible answers:

- (a) the fact that the fitness of an organism depends on the composition of its group;
- (b) the fact that the groups vary in fitness;
- (c) the fact that the organism-type with the highest overall fitness has lowest fitness within each group.

These answers are importantly different. (a) implies that the only situation in which averaging would *not* be fallacious is where organismic fitnesses do not depend on group composition, i.e., where an organism’s fitness is a function of its own phenotype alone. (b) implies that the only situation in which averaging would not be fallacious is where all groups have identical fitness. (c) implies that averaging would not be fallacious in any situation where the organism-type with highest overall fitness also has highest fitness within each group. In Sober and Wilson’s examples, conditions (a) (b) and (c) are all satisfied. But the conditions are obviously non-equivalent. The relations between them are as follows. *Modulo* the assumption that all groups are

the same size, then (c) entails both (a) and (b), i.e., if the ordering of within-group relative fitnesses is “reversed” when we average across all groups, then the groups must vary in fitness, and the fitness of an organism must depend on the composition of its group.³ (b) entails neither (a) nor (c), with or without the assumption of constant group size. (a) entails neither (b) nor (c), with or without the assumption of constant group size.

Which of the three features do Sober and Wilson think makes averaging fallacious? I doubt that it is (c). For their view is not that group selection always leads altruism to evolve but rather that it *may* do – it depends on whether it is powerful enough to counteract the individual selection for selfishness. So they allow that group selection can operate in a situation, but altruism not increase in frequency in that situation. In such a situation, condition (c) will not be satisfied – selfishness will have higher overall fitness, as well as being fitter within each group. Since the averaging fallacy is supposed to obscure group selection, I take it that Sober and Wilson think it is fallacious to average whether or not (c) is satisfied.

It is doubtful that condition (a) *alone* makes averaging fallacious. In the model above, if all the groups contained identical proportions of altruists, then it would presumably *not* be fallacious to average fitness over all groups, even though an organism’s fitness would still depend on the composition of its group. With no between-group variance in proportion of altruists, no group selection occurs, so averaging would not obscure the true causal facts. This suggests that condition (b) is the key. However, (b) alone also seems insufficient. Suppose two types of organism, A and B, are distributed into groups of equal size in varying proportions; As are fitter than Bs, so there is variance in group fitness. But the fitness of an organism depends entirely on its own phenotype, so each A has identical fitness whatever group it is in, and likewise for each B. (A’s could be taller, for example). In this case, averaging does not seem fallacious. The overall fitness of each type will simply be identical to its fitness within each group.

This suggests that the *conjunction* of conditions (a) and (b) is what makes averaging fallacious, in the Sober and Wilson model for the evolution of altruism. Groups must vary in fitness, and there must be group-level effects on organismic fitness, if averaging is to obscure the causal facts. In what follows, I assume that this is Sober and Wilson’s view. However, my argument does not depend on this assumption. If their real view is that (a) alone, or (b) alone, is the feature that makes averaging fallacious, then my argument goes through nonetheless.

Genic selectionism and heterozygote superiority

The Sober and Lewontin (1982) argument was an attack on the genic selectionist idea that natural selection is a force that operates at the level of individual genes, favouring those genes whose phenotypic effects cause them to spread faster than their alleles. Sober and Lewontin argued that the well-known phenomenon of heterozygote superiority (heterosis) belies the claims of genic selectionism.

To fix ideas, we consider a simple one-locus population genetics model with two alleles A and a, whose frequencies are p and q respectively; $p + q = 1$. We assume that mating is random, so initial (pre-selection) genotypic frequencies are in Hardy-Weinberg equilibrium. We then introduce viability selection. Genotypic fitnesses are assumed constant, and are denoted by W_{AA} , W_{Aa} and W_{aa} .

	AA	Aa	aa
Initial frequency	p^2	$2pq$	q^2
Fitness	W_{AA}	W_{Aa}	W_{aa}

To calculate post-selection frequencies, we multiply initial frequency by fitness for each genotype, then normalize by dividing by mean population fitness \bar{w} in the usual way; $\bar{w} = (p^2W_{AA} + 2pqW_{Aa} + q^2W_{aa})$.

	AA	Aa	aa
Post-selection frequency	$(p^2W_{AA})/\bar{w}$	$(2pqW_{Aa})/\bar{w}$	$(q^2W_{aa})/\bar{w}$

New allelic frequencies are $p' = (p^2W_{AA} + pqW_{Aa})/\bar{w}$; $q' = (q^2W_{aa} + pqW_{Aa})/\bar{w}$. If p' and q' are different from p and q, then evolution by natural selection has occurred: one allele has increased in frequency at the expense of the other.

In this simple model, fitnesses are ascribed to diploid genotypes, not individual genes. But it is straightforward to calculate genic fitness coefficients W_A and W_a , attaching to the individual alleles A and a themselves, which suffice to predict the system's evolution. There are a number of ways to do this. The method employed by Sober and Lewontin is to stipulate that $pW_A = p'\bar{w}$, i.e., A's initial frequency multiplied by its fitness equals A's post-selection frequency normalized by average fitness, and then solve for W_A . An alternative (but equivalent) approach will be used here, for reasons that will become clear. We calculate W_A by averaging the fitness of the A allele across all the genotypes in which it occurs, weighting by the frequency of each genotype and the proportion of A's *within* the genotype; this is exactly

analogous to the method for calculating the average fitness of altruists and selfish types that we used in the group selection model above.

Presuming meiosis is fair, this gives:

$$W_A = (W_{AA}p^2 + W_{Aa} \left(\frac{1}{2}\right) 2pq) / (p^2 + \left(\frac{1}{2}\right) 2pq) = pW_{AA} + qW_{Aa}$$

$$\text{Similarly, } W_a = (W_{aa}q^2 + W_{Aa} \left(\frac{1}{2}\right) 2pq) / (q^2 + \left(\frac{1}{2}\right) 2pq) = qW_{aa} + pW_{Aa}$$

These genic selection coefficients now predict the evolution of the system. If $W_A > W_a$ then the A allele will spread; if $W_a > W_A$ then the a allele will spread; if $W_A = W_a$ then the system is in allelic equilibrium. Genic selectionists, who argue that the gene is always the unit of selection, rest their case (in part) on the possibility of calculating genic selection coefficients of this sort. Though fitnesses were initially ascribed to diploid genotypes, by averaging over all genotypes we can determine which allele is fitter overall; selection can therefore be thought of as operating at the genic level, favouring the fitter allele. As GC Williams (1966) wrote: “one allele can always be regarded as having a certain selection coefficient relative to another at the same locus at any given point in time . . . Adaptation can thus be attributed to the effect of selection acting independently at each locus” (p. 112).

Sober and Lewontin argue that in some cases, averaging over genotypes and attributing selection to differences in genic fitness is unexceptionable. However, in cases of heterozygote superiority, i.e., where $W_{Aa} > W_{AA}$ and $W_{Aa} > W_{aa}$, the averaging strategy obscures the causal structure of the selection process. In such cases, genic fitnesses W_A and W_a can of course be calculated, and will suffice to predict the gene frequency change from one generation to another. But in these cases, W_A and W_a are mere statistical artifacts, Sober and Lewontin argue. Selection is really acting on diploid genotypes, not individual genes.

What is so special about heterozygote superiority? The crucial feature, according to Sober and Lewontin, is that the effect of individual alleles on genotypic fitness is context-dependent. In some genotypic contexts, having a copy of the A allele *raises* an organism’s fitness, while in other genotypic contexts it *lowers* an organism’s fitness. The A allele therefore does not have a “uniform causal role”, and so cannot itself be subject to selection; the same is true of the a allele. “If we wish to talk about selection for a single gene”, Sober and Lewontin write, “then there must be such a thing as the causal upshot of possessing that gene” (1982: 122). This is the condition that is violated in the heterosis example, and is why the averaging strategy misleads about the level of selection.

Sober and Lewontin's argument obviously implies, and Sober's (1984) presentation of the argument explicitly states, that it *is* permissible to think of the gene as the unit of selection in diploid population genetics models so long as each individual allele *does* have a "uniform causal role". In these cases, averaging is not fallacious. Sober (1984) also provides a more explicit account of what "uniform causal role" means (though the germ of the idea is clearly contained in the earlier paper).⁴ For a gene to have a uniform causal role, the presence of the gene must raise (lower) genotypic fitness in at least one genotypic context, and not lower (raise) it in any context. This idea is best explained graphically. The diagrams in Figure 2 plot genotypic fitness against proportion of A alleles within the genotype. Clearly, there are only three possible proportions – 1, $\frac{1}{2}$ and 0 – depending on whether the genotype is AA, Aa, or aa.

In each of these diagrams, the A allele plays a "uniform causal role", according to the Sober and Lewontin criterion. In diagram (i), aa and Aa genotypes have equal fitness, while AA has highest; in diagram (ii) fitnesses are additive, genotype fitness increasing linearly as the proportion of A alleles increases; in diagram (iii), fitnesses are also additive, but genotype fitness declines as the proportion of As increases. So selection at the genic level will operate in all three cases, favouring the A allele in cases (i) and (ii), disfavouring it in case (iii). The averaging strategy is not fallacious in these cases, for Sober and Lewontin. By contrast, consider diagram (iv), which represents heterozygote superiority, with AA the fitter of the two homozygotes, i.e., $W_{Aa} > W_{AA} > W_{aa}$. In this case, the A allele does not have a uniform causal role, according to Sober and Lewontin, so averaging is fallacious.

Sober and Lewontin's criterion for when averaging is fallacious can therefore be expressed as follows. Averaging is permissible where genotype fitness is a monotone function, increasing or decreasing, of the proportion of either allele;⁵ otherwise averaging is fallacious. I refer to this as the *monotonicity criterion*. The reason for formulating the criterion this way will become apparent. In the heterosis case, genotype fitness is a nonmonotone function of the frequency of both alleles; that is what makes averaging fallacious.

Sober and Lewontin's argument has received extensive discussion in the literature. Some (e.g., Lloyd (1988), Brandon (1992)) have endorsed the argument, while others (e.g., Sterelny and Kitcher (1988), Waters (1991)) have rejected it, claiming that the heterosis example involves no more than the standard relativity of fitness to environment (genetic environment in this case), and is thus fully compatible with genic selectionism. (Interestingly, Maynard Smith (1987) says that he was initially convinced by Sober and Lewontin's argument but later changed his mind). The details of this debate will not be examined here. My aim is to show that *if* we accept Sober and

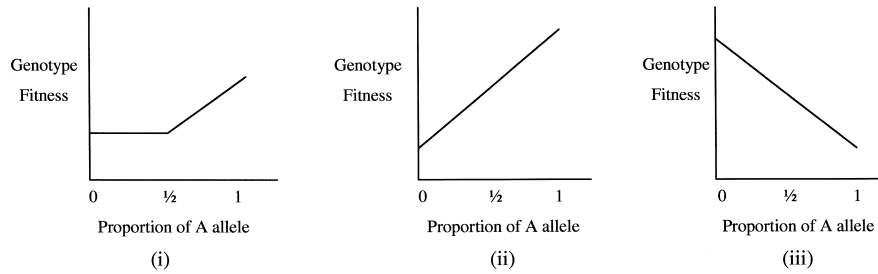


Figure 2. Monotonicity.

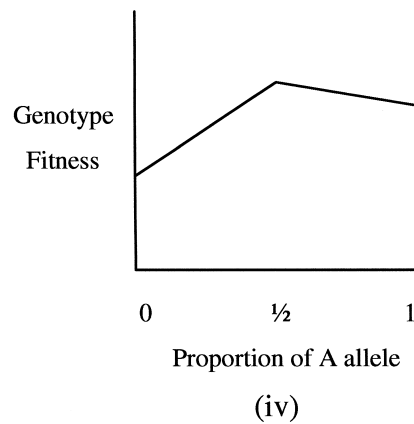


Figure 3. Non-monotonicity.

Lewontin's criterion for when it is fallacious to average over diploid genotypes, parity of argument forces us to reject Sober and Wilson's criterion for when it is fallacious to average over groups of organisms, and vice versa.

How to compare the two averaging arguments

The technique for comparing Sober and Wilson's (1998) argument and Sober and Lewontin's (1982) argument is essentially straightforward, and derives from a point that is alluded to by Sober and Wilson (1998), and developed in detail by Wilson (1990) and Kerr and Godfrey-Smith (2002). The point is that the basic multi-level selection framework used to model the evolution of altruism, which pictures selection as operating simultaneously at the group level and the individual level, can in fact be applied to a standard one-locus population genetics model of the sort examined above. One only has to think of diploid organisms as groups containing two alleles each; the fitness of a group (i.e., an organism) depends on the relative proportions of A and a

alleles it contains; these fitnesses then determine how many alleles the group contributes to the gamete pool, from which new groups of size $n = 2$ are then formed. The gamete pool therefore corresponds to the mixing stage in Figure 1 above. With fair meiosis, both alleles within any group (organism) have equal fitness, so all the variance in fitness must be between groups. But if meiosis is not fair, i.e., if either allele is a segregation-distorter, then there will be both within-group selection and between-group selection.

It is important to stress that this is not merely an analogy. As Wilson (1990) and Kerr and Godfrey-Smith (2002) show, there is a formal equivalence between one-locus two-allele population genetics models interpreted this way, and multi-level selection theory. This makes it easy to compare the two “averaging” arguments. There is potential for terminological confusion here, as we wish to use “group” to refer both to individual organisms (in the diploid population genetics case), and to refer to groups of organisms (in the evolution of altruism case). To avoid this problem, I introduce the neutral terms “collective” and “particle”, and let “group” retain its original meaning. In the evolution of altruism scenario, the collectives are groups and the particles are individual organisms; in the diploid population genetics scenario, the collectives are organisms and the particles are alleles (or genes).

Consider first the Sober and Lewontin monotonicity criterion. Translated into our new terminology, this criterion states that averaging is permissible where collective fitness is a monotone function of the proportion of each particle-type, and fallacious otherwise. This immediately implies that averaging is *not* fallacious in the evolution of altruism model. Recall equation 1, which gives group fitness as a function of the proportion of altruists:

$$W_G(x) = x(b(n - 1) - c) + n \quad (1)$$

$W_G(x)$ denotes the fitness of a group containing x altruists and $(n - x)$ selfish types. Since b and c are constants, and since we are assuming with Sober and Wilson that $b(n - 1) > c$, i.e., that raising the proportion of altruists increases the group’s fitness, it follows that $W_G(x)$ is a linear monotone increasing function of x , with slope $b(n - 1) - c$. This is shown in the diagram below.

Therefore, if the monotonicity criterion is correct, averaging is not fallacious in the evolution of altruism model.⁶ If it is permissible to average over collectives so long as collective fitness depends monotonically on proportion of particle-types, as Sober and Lewontin argue, then it is permissible to average over groups in the evolution of altruism model, contra Sober and Wilson. To put the point slightly differently, if Sober and Lewontin’s *reason* for saying that the diploid genotype, rather than the individual gene, is the unit of selection in the heterosis case is correct, parity of argument implies that the individual organism, rather than the group, is the unit of selection in

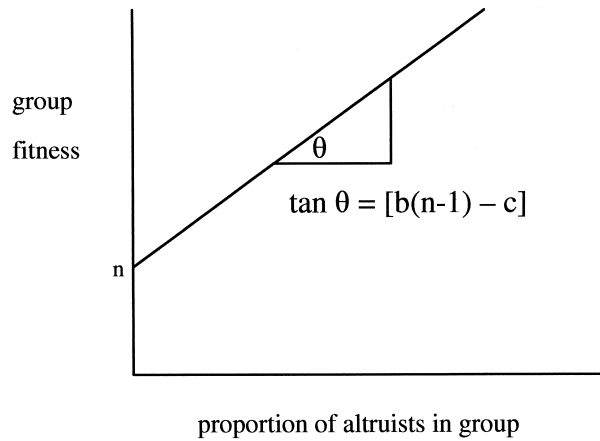


Figure 4. Group fitness as a function of proportion of altruists.

the evolution of altruism case. But as we have seen, Sober and Wilson insist that the evolution of altruism model *does* involve group selection.

Now let us run the argument in the other direction. Suppose that the Sober and Wilson criterion for when averaging is fallacious is correct. Averaging is fallacious when (a) the fitness of a particle depends on the composition of its collective, and (b) the collectives vary in fitness. This implies that averaging is fallacious in any diploid population genetics model where the three genotypes differ in fitness, *whether or not there is heterozygote superiority*. Unless $W_{AA} = W_{Aa} = W_{aa}$, then there will be variance in collective fitness, and the fitness of at least one of the particle-types will depend on which collective it is housed in. So if Sober and Wilson's criterion is correct, then it is *always* fallacious to average over diploid genotypes and attribute the system's evolution to genic selection, except in the limiting case where all the genotypes are equal in fitness. Heterozygotic superiority, or non-monotonicity more generally, has got nothing to do with it. This clearly conflicts with the position of Sober and Lewontin (1982); their argument was that averaging over diploid genotypes is perfectly acceptable in many cases.

On the Sober and Wilson view, the only cases of genic selection will involve segregation-distortion, where the alleles within a given genotype differ in fitness. If there is segregation-distortion with constant genotypic fitnesses, then it is a case of exclusively genic selection, and averaging is permissible; if the genotypes vary in fitness too, as in most real cases of segregation-distortion, Sober and Wilson will say that genic selection and organismic selection are occurring simultaneously, just as group selection and organismic selection occur simultaneously in the evolution of altruism model. So averaging would be fallacious: it would wrongly suggest that *all* the selec-

tion was at the genic level. In my view this is a plausible and instructive way to think about segregation-distortion; but the foregoing shows that it is not compatible with the Sober and Lewontin criterion for when averaging across diploid genotypes is fallacious.

It follows that the two averaging arguments are mutually incompatible. If Sober and Wilson are right that averaging is fallacious in the evolution of altruism model, it follows that averaging is almost always fallacious in diploid population genetics models, and thus that pure genic selection is very rare. If Sober and Lewontin are right that averaging is fallacious in the heterosis example, where monotonicity is violated, but not fallacious otherwise, it follows that averaging is not fallacious in the evolution of altruism model, and thus that altruism evolves by individual selection, not group selection. Though structurally similar at first sight, the two arguments turn out to conflict.

The underlying reason for the conflict is that the arguments offer different *types* of criterion for when averaging is fallacious. One looks at the effects of particles on the fitness of the collective; the other looks at the effects of the collective on the fitness of particles. The Sober and Lewontin criterion asks whether the presence of an additional particle-type always makes the same *sort* of difference to the fitness of the collective in which it is housed; if so, then it is legitimate to average over collectives, otherwise not. The Sober and Wilson criterion asks whether the fitness of a particle-type depends on the collective in which it is housed, and whether the collectives vary in fitness; if so, then it is illegitimate to average over collectives; otherwise, it is legitimate. The two criteria are thus fundamentally different in nature.

A possible reconciliation

Strictly speaking, what the foregoing shows is that the two averaging arguments are incompatible, *if* we take them to be addressing comparable questions. This is because my argument above assumes that the expression “level of selection”, as it occurs in both averaging arguments, has a univocal interpretation. This is a natural assumption to make. However, it *may* be possible to reconcile the two averaging arguments by dropping the univocity assumption. That is, when Sober and Lewontin ask whether selection is at the genic or the genotypic level, and when Sober and Wilson ask whether selection is at the group or individual level, we interpret the expression “level of selection” differently in the two cases.

This proposal may sound implausible, in view of the formal equivalence between multilevel selection theory and diploid population genetics. However, an ambiguity in the concept of “genic selectionism” suggests a

possible way of carrying out the proposed reconciliation. Genic selectionism, or the idea that the gene is the “unit of selection”, has sometimes been presented as an empirical thesis about the world, at other times as a heuristically valuable perspective from which to view evolution. This ambiguity permeates the early work of Dawkins, in particular. On the one hand, Dawkins used the famous Necker-cube analogy to argue that genic selectionism and orthodox organismic selection are alternative ways of looking at a single process; this suggests that genic selectionism is a heuristic perspective. On the other hand, Dawkins cited phenomena such as meiotic drive, “outlaw” genes, selfish DNA etc., which cannot be interpreted in terms of organismic advantage, as evidence for genic selectionism; this suggests that genic selectionism is an empirical thesis about the world, not just a heuristic perspective.

One way of resolving this ambiguity is to insist on a distinction between selection processes that occur at the genic level, and a genic or gene’s eye perspective on selection processes that occur at other levels. In cases of meiotic drive, “outlaw” genes and selfish DNA, the selection process is at the genic level – there is selection between genes within the same organism. In cases of orthodox organismic selection, the selection process is at the organismic level, for there are no fitness differences between genes within organisms, but rather between organisms. However, it is still possible to adopt a gene’s eye perspective on such selection processes, for the end result of the process is the increase in frequency of one gene at the expense of its alleles. Therefore, it is a factual matter whether a given selection process occurs at the genic level – some do and some do not; but *any* (microevolutionary) selection process, at whatever level, can be viewed from the genic perspective.

If this way of resolving the ambiguity in Dawkins’ concept of genic selectionism is accepted, then an interesting reconciliation between the two averaging arguments suggests itself, as follows. We accept the Sober and Wilson (1998) criterion for determining the level at which a selection process occurs: if the particles within a collective differ in fitness, then there is selection at the particle level; if the collectives themselves differ in fitness, and if the fitness of a particle depends on the composition of its collective, then there is selection at the collective level. So in the diploid population genetics case, selection at the genic level requires that segregation in the heterozygote be distorted; otherwise all the selection is at the genotypic level. This entails that we must reject the Sober and Lewontin (1982) monotonicity criterion as a criterion for determining the level at which selection is occurring.

However, we can interpret the Sober and Lewontin (1982) criterion in another way, bearing in the mind the distinction between the *process* of genic selection and a *genic perspective* on selection processes at other levels. We

can interpret the monotonicity criterion, not as a criterion for determining whether the process of selection is at the genic or genotypic level, but rather as a criterion for determining whether or not a genic *perspective* on the selection process is heuristically useful or not. Where genotypic fitness is a monotone function of the proportion of either allele, then it is heuristically useful to adopt the gene's eye view – even though the selection *process* may be occurring exclusively at the genotypic level. Thinking of one gene as being “fitter on average” than its allele, and so increasing in frequency, gives us a heuristically useful perspective on the selection process. However, where monotonicity is not satisfied, adopting the gene's eye view is not heuristically useful. A weaker version of this idea would hold that monotonicity is a *necessary* condition for the gene's eye viewpoint to be heuristically useful, but not sufficient.

Interpreted this way, the Sober and Lewontin argument does not conflict with the Sober and Wilson argument, for they are addressing different questions. The latter is about the level at which a selection process occurs, the former is about the utility of a particular way of thinking about a selection process. This reconciles the two arguments. But is the Sober and Lewontin argument, re-interpreted this way, actually a plausible one? Is it really true that where we are dealing with selection in a diploid species, the utility of the gene's eye view is contingent on genotypic fitness being a monotone function of the proportion of different alleles? Why should this be so?

This is a difficult question, which cannot be properly tackled here. But a few remarks are in order. One aspect of Dawkins' “gene's eye” view of evolution is the idea that individual genes are the fundamental “units of self-interest” in evolution, while organisms are their passive vehicles, mere epiphenomena of the evolutionary process. So we should think of organismic adaptations, and organisms themselves, as side-effects of the competition for increased representation among gene-lineages. Though the gene's eye view of evolution is not without its critics, it has proved heuristically useful to at least some evolutionists. Social behaviours that evolve by kin selection, in particular, can be usefully understood as strategies “devised” by genes to secure their future propagation; most people find the alternative “inclusive fitness” approach to the evolution of social behaviour much less intuitive. “Extended phenotype” examples are also usefully thought about from the gene's perspective.

It could perhaps be argued that where monotonicity is violated, the utility of the gene's eye viewpoint is reduced. Thinking in terms of the “good of the gene”, as Dawkins does, makes good sense where a given gene has a relatively context-independent effect on organismic fitness. The phenotypic effect of the gene can then be thought of as a strategy “devised” by the gene

to give it a selective advantage. But as early critics of Dawkins pointed out, if a gene's effect on the phenotype is dependent on which other genes are in the genome, thinking of selection as operating independently at each locus is a serious over-simplification. And where monotonicity is violated, we are dealing with an extreme case of context-dependence – the *type* of effect that a gene has on the fitness of its host organism depends on which gene is at the homologous locus. So in a sense, the gene's chances of propagation are not in its own hands – they depend on factors extraneous to it. The gene is not in control of its own destiny. By contrast, where genotypic fitness depends monotonically on proportion of alleles at the locus, then a particular gene does exert control over its own destiny, for it has a consistent type of effect on organismic fitness – it either raises it or lowers it. So the plausibility of thinking of genes as the real beneficiaries of the evolutionary process, manipulating the world around them for their own benefit, may depend on whether the monotonicity criterion is satisfied.

This is of course a far from conclusive argument. Indeed, a conclusive argument for or against the relevance of monotonicity is probably not possible, for whether one finds the “gene's eye” view of evolution heuristically valuable is to some extent a subjective question. Moreover, it is unclear why monotonicity, rather than for example additivity, should be the relevant factor. A number of authors have argued that if genes interact nonadditively in the determination of organismic fitness, this is problematic for genic selectionism (Wimsatt (1980), Lloyd (1988), Wright (1980), Gould (1999)). (As with the Sober and Lewontin monotonicity criterion, we should interpret the additivity criterion as a criterion for when the genic perspective is heuristically useful, *not* as a criterion for when selection is at the genic level, given that we are accepting the Sober and Wilson criterion for the latter.⁷) Additivity is of course a special case of monotonicity – so requiring additivity rather than monotonicity would lead us to adopt the gene's eye view in a much fewer range of cases. In a one-locus diploid model, requiring additivity would lead us to adopt the gene's eye view only if there were no dominance at all, which is obviously a much rarer circumstance than the absence of heterosis. In my view, it is unclear how to decide whether additivity or monotonicity is the relevant requirement, if either.

Conclusion

To conclude, I have shown that the two averaging arguments, though structurally similar, are in fact incompatible with one another if interpreted at face value. If we accept the Sober and Lewontin criterion for when averaging over diploid genotypes is fallacious, we cannot accept the Sober and Wilson

criterion for when averaging over groups is fallacious, given that diploid population genetics is formally isomorphic to multi-level selection theory. The reasons for the incompatibility were explored. Finally, I suggested a possible way of reconciling the two arguments, by interpreting the two criteria as addressing different questions; the merit of this proposed reconciliation is a matter that requires further examination.⁸

Notes

¹ What both Sober and Lewontin and I call “genetic fitness” is sometimes called “*marginal genetic fitness*” in the population genetics literature – it is the fitness of a gene averaged across all the genotypic contexts in which it occurs. Some authors (e.g., Kerr and Godfrey-Smith (2002)) have used “genetic fitness” differently, to mean the fitness of a gene *in* a given genotypic context, rather than the average across genotypes. Though this usage has much to recommend it, because it parallels the use of “individual fitness” to mean the fitness of an individual organism in a given group rather than averaged across groups, I do not adopt it here.

² Intuitively, this means that if one selfish organism in a group “converts” to being an altruist, the resulting net fitness boost to all other members of the group is greater than the loss of fitness incurred by the converting organism. It is important to see that this is not guaranteed by the fact that $W_S(x) > W_A(x)$ for all x ; see Kerr and Godfrey-Smith (2002) for further discussion of this point, and its significance.

³ Strictly, (c) entails that the fitness of *at least one* organism-type must depend on the composition of its group, not that the fitness of *every* organism-type in the population must so depend. Therefore, condition (b) in the text should be interpreted in the “at least one” not the “every” sense, for it to be true that (c) entails (b) given constant group size. Thanks to an anonymous referee for pointing this out.

⁴ In what follows, I am assuming that the position spelled out in detail in Sober (1984) is correctly attributable to Sober and Lewontin (1982). This assumption seems perfectly reasonable, given the similarities between these two works in their treatment of the heterosis example, and its implications for genic selectionism.

⁵ Strictly we should add that the function must be strictly monotonic over at least part of its domain, to rule out the case where $W_{AA} = W_{Aa} = W_{aa}$. In this case, neither allele plays a “uniform causal role”, according to Sober and Lewontin, for it is not the case that either allele raises fitness in at least one context. But a constant function counts as monotone, on the standard definition, for its first derivative never changes sign. In what follows I ignore this qualification.

⁶ Note that this result in no way depends on the fact that Sober and Wilson assume linear fitness functions. Since their definition of altruism (like that of most authors) requires that $W_G(x + 1) > W_G(x)$ for all x , it follows that in *any* model for the evolution of altruism monotonicity will be satisfied, whether or not group fitness is linearly dependent on proportion of altruists. Thanks to an anonymous referee for pointing this out.

⁷ Interpreted this way, the additivity criterion would be immune from the critiques that Godfrey-Smith (1992), Sarkar (1994) and Sober and Wilson (1994) have leveled against it. These critiques assume that the the Wimsatt/Lloyd additivity criterion *is* meant to tell us the level at which a given selection process occurs; this assumption is perfectly reasonable, as Wimsatt and Lloyd give every impression that that is indeed the question their criterion is supposed to answer.

⁸ Thanks to Peter Godfrey-Smith, Ben Kerr, Elliott Sober, Kim Sterelny and an anonymous referee for helpful comments on a previous version, and to the AHRB for financial support.

References

- Brandon R.: 1992, *Adaptation and Environment*, Princeton University Press, Princeton.
- Dawkins R.: 1976, *The Selfish Gene*, Oxford University Press, Oxford.
- Dawkins R.: 1982, *The Extended Phenotype*, Oxford University Press, Oxford.
- Godfrey-Smith P.: 1992, 'Additivity and the Units of Selection', in Hull D., Forbes M. and Okruhlik K. (eds.), *PSA 1992*, Vol. 1, Philosophy of Science Association, East Lansing, pp. 315–328.
- Gould S.J.: 1999 'The Evolutionary Definition of Selective Agency', in Singh, Krimbas, Paul and Beatty (eds.), *Thinking about Evolution*, Cambridge University Press, Cambridge.
- Kerr B. and Godfrey-Smith P.: 2002, 'Individualist and Multi-level Perspectives on Selection in Structured Populations', *Biology and Philosophy* 17(4), 477–517.
- Lloyd E.: 1988, *The Structure and Confirmation of Evolutionary Theory*, Greenwood Press, New York.
- Maynard Smith J.: 1987, 'How to Model Evolution', in Dupre J. (ed.), *The Latest on the Best: Essays on Evolution and Optimality*, MIT Press, Cambridge, MA, pp. 119–131.
- Roughgarden J.: 1979, *Theory of Population Genetics and Evolutionary Ecology: An Introduction*, Macmillan, New York.
- Sarkar S.: 1994, 'The Additivity of Variance and the Selection of Alleles', in Hull D., Forbes M. and Burian R. (eds.), *PSA 1994*, Vol. 1, Philosophy of Science Association, East Lansing, pp. 3–12.
- Sober E.: 1984, *The Nature of Selection: Evolutionary Theory in Philosophical Focus*, MIT Press, Cambridge, MA.
- Sober E. and Lewontin R.: 1982, 'Artifact, Cause and Genic Selection', *Philosophy of Science* 49, 157–180, reprinted in Brandon R.N. and Burian R. (1984) (eds.), *Genes, Organisms, Populations*, MIT Press, Cambridge, MA, pp. 109–132; page references are to the latter.
- Sober E. and Wilson D.S.: 1994, 'A Critical Review of Philosophical Work on the Units of Selection Problem', *Philosophy of Science* 61, 534–555.
- Sober E. and Wilson D.S.: 1998, *Unto Others: The Evolution and Psychology of Unselfish Behaviour*, Harvard University Press, Cambridge, MA.
- Sterelny K. and Griffiths P.E.: 1999, *Sex and Death: An Introduction to the Philosophy of Biology*, University of Chicago Press, Chicago.
- Sterelny K. and Kitcher P.: 1988, 'The Return of the Gene', *Journal of Philosophy* 85, 339–360.
- Waters C.K.: 1991, 'Tempered Realism about the Forces of Selection', *Philosophy of Science* 58, 553–573.
- Williams G.C.: 1966, *Adaptation and Natural Selection*, Princeton University Press, Princeton.
- Wilson D.S.: 1975, 'A Theory of Group Selection', *Proceedings of the National Academy of Sciences USA* 72, 143–146.
- Wilson D.S.: 1990, 'Weak Altruism, Strong Group Selection', *Oikos* 59(1), 135–140.
- Wimsatt W.: 1980, 'Reductionist Research Strategies and their Biases in the Units of Selection Controversy', in Nickles T. (ed.), *Scientific Discovery: Case Studies*, Reidel, Dordrecht, pp. 213–259.
- Wright S.: 1980, 'Genic and Organismic Selection', *Evolution* 34, 825–843.