

Evaluation of Evidence in Group Selection Debates¹

Elisabeth A. Lloyd

University of California, San Diego

The conflation of two fundamentally distinct issues has generated serious confusion in the philosophical and biological literature concerning the units of selection. The questions of how a unit of selection is defined, theoretically, is rarely distinguished from the question of how to determine the empirical accuracy of claims--either specific or general--concerning which unit(s) are undergoing selection processes. In this paper, I begin by refining a definition of the unit of selection, first presented by William Wimsatt, that is grounded in the structure of natural selection models. I then explore the implications of this definition for empirical evaluation of group selection models. I examine an objection to Wimsatt's definition raised by Elliott Sober, and I conclude by indicating how the proposed refined definition can clarify which sorts of empirical evidence could be brought to bear on group selection controversies.

1. Defining the Units of Selection

Much of the recent philosophical and biological literature on the units of selection problem makes explicit use of the "logical skeleton" of the principle of natural selection presented by the biologist Richard Lewontin in his paper, "The Units of Selection" (1970) (e.g., Wimsatt 1980, 1981; Sober 1981, 1984; Hull 1980; Ruse 1980; Walton ms.; see Brandon and Burian 1984 for an overview of the units of selection controversies).

The three principles presented by Lewontin--phenotypic variation, differential fitness, and heritability of fitness--are meant to "embody the principle of evolution by natural selection." (1970, p. 1). The generality of these principles is noted by Lewontin, who writes, "any entities that have variation, reproduction, and heritability may evolve." (1970, p. 1). That is, although Lewontin's formulation of Darwin's principles serves as a set of necessary and sufficient conditions for evolution to occur, it seems to be a necessary but not a sufficient set of conditions for a type of entity to act as a unit of selection. As Wimsatt has argued, the set of three conditions defines types of entities that either are units or are composed of units (1981, p. 142). Wimsatt suggests the following definition both as a corollary

to Lewontin's third principle--heritability of fitness--and as a sufficient condition for a unit of selection:

Definition 1

A unit of selection is any entity for which there is heritable context-independent variance in fitness among entities at that level which does not appear as heritable context-independent variance in fitness (and thus, for which the variance in fitness is context-dependent) at any lower level of organization. (Wimsatt 1981, p. 144).

I reformulate Wimsatt's definition as follows:

Definition 2

Assume that for each entity there is a unique entity-type. The entity-type ranges over (for instance) Z= (gene, genotype, genome, individual organism, deme, population, species). Each element of Z represents a unique biological level. There may be many different kinds of entity at a given level, e.g., there may be many possible combinations of alleles (kinds) which are all genotypes (entities of type "genotype"). Any entity-type for which there is an additive component of variance for some specific component of fitness, F^* , among all entities at that level, which does not appear as an additive component of variance in F^* among all entities at any lower level, is a unit of selection.

Note that this definition allows for several units of selection to be described simultaneously in the same system.

Wimsatt's basic idea, which I can only touch on here, is as follows: Additivity of the variance in fitness plays a central role in population genetics theory. According to Fisher's fundamental theorem of natural selection, additive variance in fitness determines the rate of evolution (Fisher 1930). Heritability, in turn, depends on the proportion of the additive component of variance in fitness to the total variance (in his paper, variance measures are always understood in relation to a specific range of environments). In other words, the effect of selection depends on the additive component of variance, not on variance as a whole. (Additivity means linear functionality at that level with those fitness parameters. See Wimsatt 1981, p. 144; Lewontin 1978 on the "principle of quasi-independence"). Wimsatt concludes that additivity of fitness contributions is important because it is equivalent to the third of Darwin's principles, the heritability of fitness (1981, p. 144; This is substantiated by biologists--see Roughgarden 1979; Wade and McCauley 1980, pp. 810-811; Arnold and Fristrup 1982, p. 116).

2. Models, Empirical claims and units of selection

The principles of variation, differential fitness, and heritability (with its detailed requirements for fitness parameters emphasized by Wimsatt) can be interpreted as a general outline or set of defining characteristics for natural selection models.

In this discussion, models are taken as purely abstract entities--structures described in mathematical English, usually--which have by themselves no empirical content. Empirical claims are made about models; a system in nature (henceforth, "natural system") is "described" or "explained" by a model when the model is homomorphic in certain respects to the natural system. The definition of natural selection models offered in section 1 is fairly abstract--few of the parameters are specified. Definition 2, above, can be seen as describing a certain role or set of relations within a natural selection model. An empirical claim, then, that some entity-type X is a unit of selection (with respect to fitness component Y) amounts to the claim that entity-type X within the model bears certain formal relations (involving variance in fitness parameters) with the rest of the model, and that this set of relations is homomorphic to the relations between the corresponding entity-type in the natural system and the rest of the natural system (for a detailed presentation of the semantic view of theories, used here, see Suppes 1957, 1967; van Fraassen 1970, 1972, 1980; and Suppe 1972, 1977; for application of the semantic approach to evolutionary theory, see Beatty 1980, 1981, 1982, Thompson 1983, 1985, and Lloyd 1983, 1984). The 'additivity criterion', as I shall call the above approach to empirical evaluation of units of selection claims, arises directly from the description of the structure of natural selection models.

Since empirical claims are made about models with respect to specific natural systems, empirical evaluation of such claims involves determining the extent of match between the model and the natural system in question. Variance in fitness is the primary property of natural selection models used to judge units of selection, according to Definition 2. Additivity of variance of a fitness parameter ('fitness parameter' refers to the coefficient of a component of fitness) in the natural system is determined by a statistical method called the analysis of variance (abbreviated ANOVA) (it can also be determined by alternative statistical methods, e.g., analysis of covariance and regression analysis). The analysis of variance is a method of partitioning the total variance into between-group and within-group components. In other words, it indicates how much of the total variation in a set of data is due to differences between groups as opposed to differences within groups. The analysis of variance can be used as a statistical tool to characterize the actual variance in fitness for some entity-type in the natural system. Empirical evaluation of units of selection claims therefore centers on determining whether the theoretical (or expected) variance in fitness represented in the abstract model matches the actual variance in fitness measured in the natural system.

3. Sober's causal definition

Elliott Sober, in his recent book, The Nature of Selection, takes a causal, rather than structural, approach to understanding units of selection (1984, p. 583). He claims that since group and individual selection differ "in virtue of their causal structure, it is unrealistic to think that a population genetical model will define what group selection is." (1984, p. 324). This claim is part of Sober's argument that Wimsatt's proposed definition of a unit of selection is fundamentally unsatisfactory.

I find Sober's causal definition, summarized below, and Definition 2 above to be equivalent (Wimsatt also finds Sober's definition to be equivalent to his own (1981, pp. 150-151)). Sober's suggestion, which can be generalized to other levels, is as follows:

Group selection occurs in a set of populations exactly when there exists some property P such that:

- 1) Groups vary with respect to whether they have P, and
- 2) There is a common causal influence on those groups that makes it the case that
- 3) Being in a group that has P is a positive causal factor in the survival and reproduction of organisms (1984, p. 314).

Sober is offering a definition of the group as a unit of selection, rather than a test for determining whether group selecting is occurring. Sober notes that "it may turn out, of course, that the analysis of variance generates the same answer to the causal question as the characterization suggested here." (1984, p. 304). He claims, though, that "although the analysis of variance may yield intuitive results for some cases, its limitations are immediately evident when we look at others." (1984, p. 271). He offers the homogeneous populations problem, discussed below, as a problematic case for the additivity criterion.

I shall argue that Sober's causal definition and the additivity criterion are equally unsuccessful in resolving Sober's homogeneous population example. Understanding this enables us to see that Sober's causal approach, rather than replacing the additivity criterion, complements it by emphasizing important aspects of hypothesis generation and testing. The additivity criterion in turn, clarifies the decisive role of certain types of evidence through grounding in the basic structure of natural selection models.

4. The homogeneous populations problem

Sober's example involves a set of six populations, each internally homogeneous for height: the first population consists in one-foot-tall individuals, the second in two-foot-tall individuals, and so on, up to the sixth population composed of six-foot-tall individuals. When a population reaches a certain census size, it sends out migrants, which form their own colonies. Each colony is also internally homogeneous for height, and it is assumed that like produces like (Sober 1984, pp. 258-259).

Suppose that the six-foot-tall groups outproduced the groups with shorter individuals. Sober's question is this: how can we tell whether the six-footers' success is a result of group selection or of individual selection? Sober claims that there is no "predictive" difference for the two hypotheses, and that an investigation into the "causes" of fitness differences is necessary (1984, p. 259).

Sober considers the situation in which individual selection is actually at work in the six-footer case. He claims that since there is variation in fitness between groups, and that this variation in fitness

is heritable, that a biologist is likely to assume that group selection is operating. In other words, Sober is claiming that the application of a group-level model to this set of populations would fulfill the requirements of a selection model, at least as they are presented by Principles 1-3, above. Yet such an application would be incorrect, says Sober, since only individual selection is actually operating. "It follows that (heritable) variation in the fitness of groups is not sufficient for the existence of group selection." Sober concludes (1984, p. 258).

Consider, in contrast to the individual selection case, the situation in which an organism's "chance of surviving and being reproductively successful is determined not by its own height but by the average height of the group it is in." (Sober 1984, p. 259). I would agree with Sober that we have a genuine case of group selection here, in which groups are selected for their average height. Note that under the strict provision that all groups are homogeneous, there seems to be no obvious way to tell whether group selection or organismic selection is operating. Sober, in facing this problem, concludes: "To distinguish group from organismic selection, we must consider why there are fitness differences of a certain kind...the crucial distinction arises at the level of causality." (1984, p. 259).

5. Group or Individual Selection?

Sober, in evaluating the example of the homogeneous populations, claims that the group selection hypothesis and individual selection hypothesis are "predictively equivalent" (1984, p. 259). I would like to emphasize that this is true only under narrowly constrained circumstances. If the groups were heterogeneous, and if group selection were in operation, then one would expect, for example, quite a different survival and reproductive rate for short individuals in tall (average) groups than for short individuals in short (average) groups. That is, the group selection and individual selection models are not predictively equivalent in general.

Once again, the analysis of variance can be seen as a mechanism or test for determining whether the variance in fitness parameters at a certain level is additive. Sober claims that the homogeneous population problem "reveals a rather straightforward defect of the anova characterization" of a unit of selection that he calls the "absent value problem" (1984, p. 271).

Before considering Sober's objection, let us examine the results of applying the suggested additivity criterion to the homogeneous populations case. We can attempt to use an analysis of variance; it becomes obvious, however, that such an analysis cannot be completed because there is no variety in the contexts in which a phenotype can find itself. (Wimsatt emphasizes that context-dependence must be evaluated under a variety of conditions in the relevant variables, 1981.) Still, there are no data to indicate that group membership does influence the fitness of phenotypes, hence, according to definition 2, the biologist should conclude that the individual organism is the unit of selection.

Suppose, though, that there is group selection operating on these populations. Under the additivity criterion, there would be no way to determine this with the information given. There would be no dependence of the variance in fitness on the (group) context; therefore, group selection would not even be considered. I emphasize that the limitation of information is the key to the failure of the additivity criterion to detect group selection in this case. Clearly, if groups were at all heterogeneous--containing both short and tall individuals--and it were the case that groups of taller-than-average-height were more fit, this fact would result in non-additivity of variance of individual fitness parameters (and in addition, the fitness parameters of groups with a given average height would be additive, hence satisfying the second clause of the definition). In other words, if the additional information were provided from heterogeneous groups, the additivity criterion would produce the correct result (see Arnold and Frisrup 1982, p. 123).

Returning to Sober's criticism, we see that he has seen the crucial role of a certain type of information applying the additivity criterion. Sober notes that, with the homogeneous populations example, an analysis of variance cannot be carried out; the analysis of variance calculations have "missing values" (1984, p. 272). He is also aware of exactly which circumstances would provide the information needed to make the additivity criterion work: "It is the ANOVA's obsession with the actual that gets in the way here... . To discover which of these selection hypotheses is true, we want to ask a hypothetical question. What would happen if populations were not internally homogeneous? But here we enter terra incognita as far as the analysis of variance is concerned." (1984, p. 272).

6. Sober's causal solution

In order for the additivity criterion to work, information is needed regarding what would happen if the groups were heterogeneous. Since this information is not available, the additivity criterion is judged by Sober to be inadequate as a general definition of a unit of selection, since there is no guaranteeing that populations will be "conveniently arranged" (1984, p. 272).

According to Sober, questions regarding units of selection must take causal mechanisms into account. In the homogeneous populations example, he says, "two techniques are available for finding out which causal mechanism was actually at work." (1984, p. 260).

First, one can manipulate the system. Sober suggests that populations could be rearranged into groups composed of individuals with different heights (heterogeneous groups); the biologist would then compare what happens to a six-footer in a population with one average height with what happens to six-footers in a population with a different average height. A series of comparisons could be run which would give evidence about whether an individual's fitness is fixed by its own height or by the average height of the group (1984, p. 260).

Clearly, if information concerning heterogeneous populations were available, the additivity criterion itself would yield the same answer.

The second technique supposedly does not require intervention into the system. The biologist, says Sober, can find out what selection forces are at work by looking for "sources" of forces. The biologist must see "what forces a system experiences by examining its environment." (1984, p. 260). For example, Sober continues, suppose predation were the main source of selection--predators do not single out prey, rather they take bites out of entire groups--and they prefer groups of very small average size organisms. Knowledge of this fact seems to indicate that it is "statistical properties of the group" that make it more or less vulnerable. Hence, "a large organism in one group might have a very different vulnerability to predators than a large organism in another group, owing to the fact that the containing groups differ." (1984, p. 260). But the groups are supposed to be homogeneous, so how could the containing groups of "large" organisms differ? Note that the logic of his argument rests on varying the group context of two otherwise identical organisms and noticing the resulting differences in fitness.

In a later discussion of this situation, different average heights of groups are seen as "causal background contexts against which to assess the causal role of the individual property of being tall." (1984, p. 315). Sober notes that if the populations are homogeneous, that such an analysis cannot be done. He claims that "we need to consider not simply the fitnesses that organisms actually have but the fitnesses they would have if they were in different groups, or if they had different heights" (1984, p. 317). Again, this is precisely the information needed for the additivity criterion to work. Just because information is needed does not mean it is available, however. Can a claim about group selection be substantiated without this information? If it cannot, then Sober has told us nothing new about determining the empirical standing of the two alternative hypotheses.

Either we started with homogeneous populations or we did not. If we did not, then analysis of variance can reveal immediately whether a group selection hypothesis should be considered, i.e., whether an individual selection model is insufficient. If we did start with homogeneous populations, appropriate manipulations of the populations may be carried out to provide enough information for the additivity criterion to operate. It seems that Sober wishes to claim that a biologist performing a causal analysis can somehow "see" the real natural system and how it works--hence no manipulations are needed.

My question is this. Suppose biologists were to look for the "real forces" operating on a system in nature by examining its environment. How would they know that they found the real forces? Sober gives us the answer: if environmental considerations give us reason to think that group selection is operating, then the variance in individual fitness parameter is expected to have certain properties--the very properties represented in definition 2. The question still remains whether the system does or does not have those properties. If evidence can be obtained that it does, then the additivity criterion is also effective.

In summary then, Sober has rejected the additivity criterion for failing a certain test. Sober, in his solution to the problem using the causal view, imports exactly the information needed to make the additivity criterion effective. If group selection is operating in a

set of completely homogeneous populations, neither the additivity criterion nor Sober's causal view could give good grounds for claiming that it is. Even if some biologists thought they had located a cause for group selection, this is not enough; they must show that they have pinpointed the correct cause by linking it to certain empirical properties of the system--precisely those picked out by the additivity criterion.

Consider the case of the Myxoma virus. Is the decreased virulence in this rabbit-killing virus the result of group selection or of individual selection? Some scientists assume that the groups of viruses living within the rabbits are homogeneous (e.g., Alexander and Borgia 1978; Futuyama 1979; Fenner 1965), while others assume the heterogeneity of the groups (e.g., Lewontin 1970; Levin and Pimentel 1981; Gilpin 1975). The theorists do not agree on the composition of the groups, which in turn has predictable consequences--according to the additivity criterion--on whether they support a group or individual selection model for Myxoma. Those biologists who assert that the groups are homogeneous (and they give their ecological and causal reasons for accepting this assumption) find individual selection. Similarly, those who assume that each group is heterogeneous (they also have their causal reasons) can utilize the additivity criterion to conclude that group selection is operating.

But surely there is a fact of the matter. Either the groups of virus particles are heterogeneous or they are not. The additivity criterion clarifies what could be done in order to make progress on this debate; the composition of the groups of viruses could be determined; if they are heterogeneous, an analysis of variance could be done; if they are homogeneous, experiments could perhaps be done to manipulate the populations in order to get the necessary information.

In this case, both sides have provided ecological and causal facts supporting their views. The reason that they do not have the answer yet is made clear by the additivity criterion; they lack the information to do an adequate empirical comparison.

Sober, in addressing this case, concludes that his own causal definition "delivers the correct conclusion that the reduction in virulence is a case of group selection (provided that Lewontin's facts are right, of course)." (1984, p. 333). But, as the additivity criterion makes quite clear, Lewontin's empirical assumption regarding the composition of the group is precisely what is at stake. Assuming that a causal picture of the system is correct is not enough to settle the debate; the claim must be justified by demonstrating that the system in nature produces statistics that conform to the particular set of model relations described in definition 2.

In summary, Sober's complaint about the analysis of variance is that it is

a way of partitioning measurements on actual organisms only; it is not sensitive to the sort of counterfactual considerations to which an analysis of causation must attend. However, if it were reformulated so that the anova table represented the fitness values that would obtain in certain counterfactual circumstances, it would

no longer be subject to the absent value problem. Nevertheless, it would still be neither necessary nor sufficient as an analysis of causation. (1984, p. 317).

The analysis of variance is not being presented by Wimsatt or by myself as an "analysis of causation"; its purpose is as a statistical tool in determining the interrelations of group and individual level fitness parameters. The fact that "the analysis of variance is an imperfect guide to causation" comes as no surprise to those familiar with Lewontin's article, "The Analysis of Variance and the Analysis of Causes" (Sober 1984, p. 275; Lewontin 1974). The conditions involving variance in Definition 2 are not intended to stand on their own merits; the point is that selection models, as they are widely understood, require that the variance in fitness parameters be additive. When considering whether a selection model is applicable to an entity at a certain level, the analysis of variance can be used--carefully, following specific methodological maxims--to determine whether the natural system is indeed described by the model. In other words, the analysis of variance is used as a test to help determine the empirical adequacy of a specific empirical claim. Now, one may want to conclude that if the model is found to be empirically adequate, then real causes in nature have been found. But this is certainly a separate issue from the determination of comparative empirical adequacy of two competing models.

Ecological (causal) information is necessary for constructing appropriate tests of units of selection claims. The requirement, agreed upon by all, for information about "heterogeneous" groups assumes the existence of classes of individuals defined according to "significant" traits. These classes must originate from the evolutionary, ecological, causal picture of the organism and its environment. I suggest that Sober, in the process of emphasizing this important aspect of the units of selection problem, has needlessly given up the advantages of the additivity criterion, not the least of which is the connection it reveals between natural selection theory and the evidence required for a units of selection claim.

Notes

¹I would particularly like to thank Dick Lewontin for discussing the units of selection problem with me, and for his helpful advice, criticism, and suggestions. For their valuable comments and criticism, I would also like to thank Bas van Fraassen, Evelyn Fox Keller, Bill Wimsatt, Steve Orzack, Dick Burian, Robert Brandon, Deborah Gordon, Hamish Spencer, David Hull, and Jim Griesemer.

References

- Alexander, R.D. and Borgia, S. (1978). "Group Selection, Altruism and the levels of organization of life." Annual Review of Ecology and Systematics 9: 449-474.
- Arnold, A.J. and Fristrup, K. (1982). "The theory of evolution by natural selection: a hierarchical expansion." Paleobiology 8: 113-129.
- Asquith, P.D. and Giere, R.N. (eds.). (1981). PSA 1980. Volume 2. East Lansing: Philosophy of Science Association.
- Beatty, J. (1980). "Optimal-design models and the strategy of mode building in evolutionary biology." Philosophy of Science 47: 532-561.
- (1981). "What's Wrong with the Received view of Evolutionary theory?" In Asquith and Giere (1981). Pages 397-426.
- (1982). "The Insights and Oversights of Molecular Genetics: The Place of the Evolutionary Perspective." In PSA 1982, Volume 1. Edited by P.D. Asquith and T. Nickles. East Lansing, Michigan: Philosophy of Science Association. Pages 341-355.
- Brandon, R.N. and Burian R.M. (eds.). (1984). Genes, Organisms, Populations: Controversies over the Units of Selection. Cambridge, MA: MIT Press.
- Fenner, F. (1965). "Myxoma virus and *Oryctolagus cuniculus*." In The Genetics of Colonizing Species. Edited by H.G. Baker and G.L. Stebbins. New York: Academic Press. Pages 485-501.
- Fisher, R.A. (1930). The Genetical Theory of Natural Selection. Oxford: The Clarendon Press.
- Futuyma, D.J. (1979). Evolutionary Biology. Sunderland, MA: Sinauer.
- Gilpin, M.E. (1975). Group Selection in Predator-Prey Communities. Princeton, NJ: Princeton University Press.
- Hull, D.L. (1980). "Individuality and Selection." Annual Review of Ecology and Systematics 11: 311-332.
- Levin, S. and Pimental, D. (1981). "Selection of intermediate rates of increase in parasite-host systems." American Naturalist 117: 308-315.
- Lewontin, R.C. (1970). "The Units of Selection." Annual Review of Ecology and Systematics 1: 1-18.
- (1974). "The analysis of variance and the analysis of cause." American Journal of Human Genetics 26: 400-411.
- (1978). "Adaptation." Scientific American 23: 156-169.
- Lloyd, E.A. (1983). "The Nature of Darwin's Support for the Theory of Natural Selection." Philosophy of Science 50: 112-129.
- (1984). "A Semantic Approach to the Structure of Population Genetics." Philosophy of Science 51: 242-264.

- Roughgarden, Jonathan. (1979). Theory of Population Genetics and Evolutionary Ecology: An Introduction. New York: MacMillan.
- Ruse, Michael. (1980). "Charles Darwin and Group Selection." Annals of Science 37: 615-630.
- Sober, E. (1981). "Holism, Individualism, and the Units of Selection." In Asquith and Giere (1981). Pages 93-121.
- (1984). The Nature of Selection. Cambridge, MA: MIT Press.
- Suppe, F. (1972). "What's Wrong with the Received View on the Structure of Scientific Theories?" Philosophy of Science 39: 1-19.
- (ed.). (1977). The Structure of Scientific Theories. 2nd ed. Urbana, IL: University of Illinois Press.
- Suppes, P. (1957). Introduction to Logic. Princeton, New Jersey: Van Nostrand.
- (1967). "What is a Scientific Theory?" In Philosophy of Science Today. Edited by S. Morgenbesser. New York: Meridian. Pages 55-67.
- Thompson, P. (1983). "The Structure of Evolutionary Theory: A Semantic Approach." Studies in History and Philosophy of Science 14: 215-229.
- (1985). "Tempo and Mode in Evolution: Punctuated Equilibria and the Modern Synthetic Theory." Manuscript.
- van Fraassen, B.C. (1970). "On the Extension of Beth's Semantics of Physical Theories." Philosophy of Science 37: 325-339.
- (1972). "A Formal Approach to the Philosophy of Science." In Paradigms and Paradoxes. Edited by R. Colodny. Pittsburgh: University of Pittsburgh Press.
- (1980). The Scientific Image. Oxford: Clarendon Press.
- Wade, M.J. and McCauley, D.E. (1980). "Group Selection: The Phenotypic and Genotypic Differentiation of Small Populations." Evolution 34: 799-812.
- Walton, D. (1985). "Covariance Mathematics and the Analysis of Multi-level Selection Processes: A New Formulation." Manuscript.
- Wimsatt, W. (1980). "Reductionist Research Strategies and Their Biases in the Units of Selection Controversy." In Scientific Discovery: Case Studies. Edited by T. Nickles. Dordrecht: Reidel. Pages 213-259.
- (1981). "Units of Selection and the Structure of the Multi-Level Genome." In Asquith and Giere (1981). Pages 122-183.