A Structural Approach to Defining Units of Selection

Elisabeth A. Lloyd

Philosophy of Science, Vol. 56, No. 3. (Sep., 1989), pp. 395-418.

Stable URL:
http://links.jstor.org/sici?sici=0031-8248%28198909%2956%3A3%3C395%3AASATDU%3E2.0.CO%3B2-F

Philosophy of Science is currently published by The University of Chicago Press.

A STRUCTURAL APPROACH TO DEFINING UNITS OF SELECTION*

ELISABETH A. LLOYD†

Department of Philosophy
University of California, Berkeley

The conflation of two fundamentally distinct issues has generated serious confusion in the philosophical and biological literature concerning the units of selection. The question of how a unit of selection of *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) is undergoing selection processes. In this paper, I begin by refining a definition of the unit of selection, first presented in the philosophical literature by William Wimsatt, which is grounded in the *structure* of natural selection models. I then explore the implications of this structural definition for empirical evaluation of claims about units of selection. I consider criticisms of this view presented by Elliott Sober—criticisms taken by some (for example, Mayo and Gilinsky 1987) to provide definitive damage to the structuralist account. I shall show that Sober has misinterpreted the structuralist views; he knocks down a straw man in order to motivate his own causal account. Furthermore, I shall argue, Sober's causal account is dependent on the structuralist account that he rejects. I conclude by indicating how the refined structural definition can clarify which *sorts* of empirical evidence could be brought to bear on a controversial case involving units of selection.

1. Defining the Units of Selection. Hull has described evolution by natural selection as a combination of several interrelated processes: replication, interaction, and evolution. According to Hull, replicators pass on their structure largely intact from generation to generation. An *interactor* is an entity that directly interacts as a cohesive whole with its environment in such a way that replication is differential (1980, p. 318). The process of evolution by natural selection, then, is "a process in which the differential extinction and proliferation of interactors cause the differential perpetuation of the replicators that produced them" (Hull 1980, p. 318; compare Brandon 1982, pp. 317–318).

According to Hull, there are two "units of selection" questions: at what level does replication occur, and at what level does interaction occur (1980,

*Received September 1987.

†I would like to thank Dick Lewontin 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 Spenser, David Hull, Sam Mitchell, Jim Griesemer, and James Woodward.

Philosophy of Science, 56 (1989) pp. 395–418.

p. 318). I shall adopt Hull's distinction in my treatment of the units of selection problem. Nearly all of the literature on the units of selection controversy involves a debate about interactors, that is, about which level of entities can and do interact with their environment in such a way as to influence the process of evolution. The "genic selectionists", most notably Dawkins (1976, 1982) and Williams (1966), claim that replicators are the only "real" units of selection. I have addressed the numerous problems with this claim elsewhere (1988, chap. 7). In this paper, I shall discuss only issues concerning interactors as units of selection.

1.1. Interactors. Questions about interactors focus on the description of the process itself—the interaction of entity and environment, and how this affects evolution—rather than on the *outcome* of this process (see Vrba and Gould 1986, p. 217). Since it is, after all, organisms that are born and die, there is a danger of ascribing all selection to the single level of organisms, that is, to organismic selection, when in fact, entities at other levels are interacting with their environments in ways that affect both the survival and reproduction of organisms and various properties of the gene pool.

The aim of this paper is to clarify and defend an approach to delineating interactors. I offer my approach, which is a refinement of Wimsatt's (1980) definition, in section 2. In section 3, I consider an attack made by Sober on the structuralist approach. I argue that he has misinterpreted, and therefore failed to consider seriously, the approach taken by Wimsatt, Arnold and Fristrup, and others. Finally, in section 4, I illustrate the utility of my additivity definition, through an analysis of the controversial case of the myxoma virus. I argue that the additivity definition helps clarify the role of key assumptions in the myxoma debate, as well as indicating directions for future research.

2. The Additivity Definition. Richard Lewontin's (1970) formulation of the "logical skeleton" of the principle of evolution by natural selection is a popular approach to delineating interactors (for example, Maynard Smith 1976; Wimsatt 1980, 1981; Hull 1980; Ruse 1980; Sober 1981, 1984; Buss 1983; see Brandon and Burian 1984 for an excellent overview of the units of selection controversies). The three conditions presented by Lewontin—phenotypic variation, differential fitness, and heritability of traits relating to 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). Although Lewontin's formulation is meant to serve as a set of necessary and sufficient conditions for evolution by natural selection, 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 (that is, an interactor) in evolution. As Wimsatt has argued, the set of three conditions articulated by Lewontin defines types of entities that either *are* units or are *composed* of units of selection (1981, pp. 143–144). Wimsatt suggests the following definition both as a corollary to Lewontin's third principle—heritability of traits relating to fitness—and as a sufficient condition for a unit of selection in evolution:

Wimsatt's Definition

> 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)

One problem with Wimsatt's definition is that the term, "context-dependent", is too vague.[1] I believe that Wimsatt's intended interpretation of the term is, "not transformable into additive variance", as he implies in the body of the paper. In order to clarify the theoretical bases of Wimsatt's definition, and to facilitate discussion, I have reformulated it more precisely, as follows:[2]

Assume that for each entity there is a unique entity-type. The entity-type ranges over (for instance) Z = {gene, chromosomal regions, genotype, genome, individual organism, kin group, population}. Each element of Z represents a unique biological level (listed from "low" to "high" levels). There may be many different *kinds* of entity at a given level, for example, there may be many possible combinations of alleles (kinds) which are all genotypes (entities of type "genotype").

Additivity Definition

> A unit of selection is any entity-type for which there is an additive component of variance for some specific component of fitness, F^*, among all entities within a system at that level which does *not* appear as an additive component of variance in (some decomposition of) F^*

[1]For example, "context" is often interpreted to include environmental context. In Wimsatt's definition, however, "context" refers only to other individual entities of the same entity-type (1981, p. 144).

[2]I take my definition to be equivalent to Wimsatt's, under his intended interpretation. I also believe that it articulates the principles behind a variety of specific approaches to defining higher levels of selection, including the views of Wade (1978, 1985), Crow and Aoki (1982), Arnold and Fristrup (1982), Damuth and Heisler (1987), among others; (see my (1988) chap. 5).

among all entities at any lower level.[3]

Note that this definition allows for several distinct kinds of units of selection to be described simultaneously in the same system. That is, an entity-type can function both as a unit of selection and as a part of a unit of selection at a higher level.

2.1. Theoretical Context of the Additivity Definition. In justifying a version of the additivity approach, Crow and Aoki cite the "secondary theorem of natural selection": the "rate of change of the mean value of a character correlated with fitness in any subpopulation is the additive genetic covariance of that character and fitness" (1982, p. 2628). That is, the rate of evolution of a trait depends on the additive genetic (co)variance between the character and fitness. In the additivity definition, we are simply *generalizing* the basic principle that relates the efficacy of natural selection to additivity (see, for example, the theoretical justification given by Crow and Aoki 1982, p. 2628). I conclude that the additivity of variance in fitness is important because it is a way of expressing and delineating the heritable traits that affect fitness. For example, in a case involving a single locus, if furry animals are more fertile or more viable, furriness will increase at a rate predictable by two factors: the additive genetic variance of furriness and the genetic correlation of furriness with fitness (see Wimsatt 1981, p. 144; Roughgarden 1979; Wade and McCauley 1980, pp. 810–811; D. S. Wilson 1983, p. 184; Arnold and Fristrup 1982, p. 116).

The additivity definition should also make intuitive sense. One operative notion in all examples of group selection is the presence of some sort of interaction effect or context-dependence (Wimsatt (1981) and Sober (1984) agree on this point). The standard expression of such interaction in mathematical or statistical language is in terms of non-additivity and non-linearity. The basic point is that biologists want empirically adequate models to describe these interactions. My claim is that models describing evolution by natural selection must be of a basic form—described very abstractly by Lewontin's three conditions—in order to provide empirically adequate descriptions of changing natural systems. *The additivity definition is a detailed description of one aspect of this basic model form*, applicable when there are certain interactions and population structures.

[3]"Additivity", in this definition, should be understood as shorthand for "transformable into additivity", for the duration of the paper. Simple additivity, which means (approximately) linear functionality at that level with those fitness parameters, is too narrow, as Lewontin has pointed out to me. Non-additivity of variance can arise from non-monotonicity or from simple non-rectilinearity; only the former relation is of interest here. See Li (1964), chap. 33, and Steel and Torrie (1960), pp. 129–131, 156–159, for some commonly used methods for transforming non-rectilinear into additive functions.

Specifically, the additivity definition describes a *role* in the model that corresponds conceptually with what is informally called an *interactor*. For example, *if* a model is judged to be empirically adequate, and its fitness parameters at the group and organismic levels have the relation described in the additivity definition, *then*, according to my view, one claims that the group is a unit of selection in this system.

Some readers may be wondering how this view relates to Sober's causal approach to the units of selection. I argue in section 3 that the additivity definition provides the theoretical underpinning for the causal definition adopted by Sober.

Ultimately, one *can* interpret the additivity definition as individuating causes. The notion of interactors is quite naturally interpreted as causal, and my definition is meant to delineate the role of interactor in evolutionary models. There are also philosophical arguments available for linking the particular form of definition I've chosen with the delineation of causes. Paul Humphreys, for example, has argued that something resembling additivity is central to the whole notion of something being a distinguishable causal factor; an event or entity is not a cause unless it makes some sort of similar contribution to the occurrence of an effect over a range of circumstances (1985).

I do not wish to make a commitment to causes here. Nevertheless, I want to emphasize that the additivity definition *can be interpreted as delineating causes*. Under a causal interpretation, the additivity definition would be understood as picking out entities with traits that interact causally with their environments in a manner that can produce evolutionary change. One of the advantages of the structural framework I use—the semantic approach—is that it leaves the theory open to both realist and anti-realist interpretations. I would like to maintain this advantage. Hence, claims made about models are compared in terms of empirical adequacy. This could be understood, by some realists, as a minimum standard for a superior causal story (see section 3.3).

2.2. Models, Empirical Claims and Units of Selection. Following the state space version of the semantic view that I have defended elsewhere, I take models to be purely abstract entities—structures described in mathematical language, usually—which have by themselves no empirical content (see van Fraassen 1980; for applications of the semantic view to evolutionary theory, see Thompson 1983, 1985, 1986; Beatty 1980, 1981, 1982; and Lloyd 1984, 1986, 1987, 1988). Empirical claims are made *about* relations between models and natural systems; 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. I take it that Lewontin described the general outline of models—a model

type—that is used to explain evolution by natural selection. The additivity definition, above, can be seen as delineating a certain role or set of relations within such a model type. I claim that the additivity definition is a formal, structural representation of the role of *interactor*, arising directly from the description of the structure of natural selection models.

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 nature and the rest of the natural system.

Note that in making an empirical claim about a unit of selection, a mathematical relation in the model—variance in fitness—is to be compared with a relation measured from the natural system, also called "variance in fitness". Variance in fitness in the model, however, is to be understood as an *expectation*, while variance in fitness as usually measured from the natural system is a sample *statistic*, taken simply from *actual* relative frequencies. This distinction between the theoretical definition and the actual statistic from the natural system is extremely important. The same statistical methods can be used to analyze both sets of values (see Kempthorne 1957, p. 234).

Empirical claims may be made at various levels; a controversial *general* claim, for instance, is that group selection is a relatively important, significant component in the process of evolution. (This claim is sometimes framed in terms of the relative importance of various causal factors.) Empirical evaluation of such general assertions, however, rests on the ability to evaluate single cases.

I believe the central issue in units of selection controversies lies in the comparison of two or more models. It is possible to produce two (or more) models with *different* units of selection that have the same expected outcome or result. For example, both an organismic-level and a group-level model might predict that the same type would be favored in evolution, within a specific range of environments. The myxoma case, discussed in section 4, is such a case. The two models can be conceived as *competing* descriptions of a natural system.

Evaluation of a claim about a unit of selection in a particular case involves determining the correspondence between each model and the natural system in question. Variance in fitness plays a decisive role in judging units of selection, according to the additivity definition. Hence, empirical evaluation of a units of selection claim must involve examination of how well the expected fitness values in the competing models match the actual statistics taken from the natural population (see Lloyd 1987, for a more complex view of confirmation).

2.3. The Additivity Definition and Methodological Considerations. One way to measure the additivity of variance of a fitness parameter ("fitness parameter" refers to the coefficient of a component of fitness) in the natural system is by a statistical method called the analysis of variance (abbreviated ANOVA). It is beyond the scope of this paper to address the differences among alternative statistical methods, for example, analysis of covariance and regression analysis (see my 1988). The analysis of variance is a method of partitioning the distribution of variation into between-group and within-group components; the method requires several kinds of information. First, fitnesses for individuals of each kind (of gene or phenotype) *within* each group are required. The average fitness for the group can then be used to determine the variance for each kind from the group mean. Finally, the average fitness *overall*, for the global population, can be used to determine the variance for each kind from the *global* mean fitness. The variance can then be broken down into the component relating to the group mean and the component relating to the global mean (for example, Crow and Aoki 1982).

The statistical tools used in describing correlation and partitioning variation require certain methodological precautions in their application. Lewontin, in his penetrating criticism of the uses of the analysis of variance in human genetics, argues that the outcome of an analysis of variance depends upon several factors, including: the actual functional relations embodied in the norm of reaction; the actual distribution of frequencies of each kind of entity; and the actual structure of environments (1974, pp. 406–408). Biologists working on the units of selection problem have explicitly recommended consideration of a *range* of population compositions and environments, in order to reduce the likelihood of getting misleading results from the sorts of unusual frequency or environment effects discussed by Lewontin (Wade and McCauley 1980; McCauley and Wade 1980; Arnold and Fristrup 1982; see also Wimsatt 1980, p. 254; 1981, p. 150). Furthermore, in using the statistical tools, some larger, environmental context is being held constant before the analysis of variance is done; for example, most population genetics models restrict the larger context to "Mendelian" reassortment of (chromosomal) genes.

The distribution of types (either genotypes or traits) can also affect the results of an analysis of variance (see Lewontin 1974, pp. 403–406). Arnold and Fristrup (1982) suggest some methodological "rules of application" in order to avoid this problem. Arnold and Fristrup state that, since their models deal with *traits* or *characters*—rather than genes—it is extremely important to develop a clear picture of the relationship between a particular character value and individual fitnesses *within* all groups (see Lloyd 1988 for discussion of the covariance approach). It is necessary to have a good understanding of just how much a particular trait

contributes to fitness, which in turn requires knowledge of a range of group compositions in which individuals can find themselves (Arnold and Fristrup 1982, pp. 122–123). Wimsatt recommends the same approach (1981, p. 150).

All of the above recommendations regarding empirical evaluation of a unit of selection claim presuppose a great deal, including: what is to count as a group; what is to count as a "mixed population", that is, what traits are significant on the individual level; and how "environment" is to be defined. This ecological and natural historical information about the natural system is therefore necessary to any evaluation of a claim regarding units of selection.

Let us summarize the discussion thus far. Certain interrelations among fitnesses are expected to hold, given that some entity is a unit of selection in a particular system. Biologists, in attempting to determine whether the natural system exhibits one set of fitness interrelations or another, may utilize one of several statistical tools. These tools are well known to yield biased or nonrepresentative results unless certain conditions are met. Hence, the biologists promoting the use of statistical tools for unit "detection" have also indicated the necessary conditions for application of the tools (see Griesemer and Wade's excellent analysis (1988)).

If the conditions of application are met, and the statistical analysis yields a result conforming to the expectation of specific unit of selection model, then the biologist might make a claim that he or she has evidence supporting a particular units of selection claim. The satisfaction of the application conditions is an *essential* part of viewing the outcome of the statistical analysis as evidence for a claim involving units of selection. A number of critics, including Sober, have ignored this key aspect of the structural approach to determining units of selection, as we shall see below.

One of the most common criticisms made regarding the additivity criterion is that it is unable to distinguish frequency-dependence from group selection. I have rebutted this claim elsewhere in some detail (Lloyd 1988, chap. 5), and will not consider it in this paper. The criticism made by Sober, discussed in section 3, below, is closely related to the frequency-dependence critique; the basic claim is that the additivity definition is unable to distinguish two different processes.

3. Sober's Causal Account. Sober, in motivating his causal characterization of group selection in *The Nature of Selection* (1984), rejects a structuralist approach to the problem. He claims that "insofar as 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). Sober characterizes the structuralist approach

to units of selection as "the ANOVA criterion", and claims that "although the analysis of variance may yield intuitive results for some cases, its limitations are immediately evident when we look at others" (1984, pp. 271–272). He offers the homogeneous populations problem, discussed below, as a basis for rejecting an approach based on theory structure. This example holds a crucial place in the arguments presented in the second half of his book; Sober uses it repeatedly to deflect the claim that a structuralist definition would be adequate and effective, and thereby to motivate his own causal account (for example, 1984, pp. 304, 323, 349).

I shall argue in section 3.2 that Sober's "ANOVA criterion" is a straw man; he has misrepresented—and thereby failed to consider seriously—structuralist views. I argue that Sober's causal definition and the additivity approach are equally unsuccessful in resolving the homogeneous populations problem, as it stands. Finally, I show that Sober's own causal solution is founded on the additivity definition he rejects. An important virtue of the additivity definition is demonstrated throughout this discussion; it clarifies the decisive role of certain types of evidence through grounding in the basic structure of evolutionary models.

3.1. 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. How can we tell whether the six-footers' success is a result of group selection or of organismic selection? If organismic selection were operating favoring tallness, we would expect the six-footers to do the best individually, the six-foot groups would become full faster, and they would send out more migrants. On the other hand, suppose group selection were operating, and an individual organism's fitness is determined not by its own height, but by the average height of the group it is in. If selection favored the taller groups, then the six-footers would do better, and there would be more six-foot groups, just as in the organismic selection case.

The problem worrying Sober is that we seem unable to compare the adequacy of the two competing models of this system if we consider only the outcomes of the models, that is, the expected frequencies of each type overall. 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 claims that the two hypotheses are "predic-

tively equivalent", and that an investigation into the "causes" of fitness differences is necessary (1984, p. 259).

Using the additivity definition, there would be no way to draw a conclusion regarding group selection with the information given. We could not perform a statistical analysis, because we lack information about the variation within groups—we could not get the variance of each type from the group mean. Still, 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 in this case.

3.2. Sober's "ANOVA Criterion" vs. the Additivity Definition. 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). He notes that, in 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 definition 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. (Sober 1984, p. 272)

Sober believes that this "absent value" objection against his "ANOVA criterion" is also decisive against the structuralist views of Wimsatt, Arnold and Fristrup (1984, p. 275, fn. 41). This is based on a misconstrual of their proposals.

Sober's complaint against the additivity approach is that it supposedly depends on the "*actual* array of fitness values"; this dependence makes the additivity account vulnerable to the absent value problem (1984, p. 275). However, Wimsatt, Arnold and Fristrup (also Wade, whom Sober does not recognize as implementing the additivity approach) all *explicitly* reject the sort of simplistic application of the statistical tools that Sober makes in his homogeneous populations example. The additional information about heterogeneous populations (the "terra incognita" of the ANOVA criterion) is, in fact, explicitly required by Arnold and Fristrup in their specification of the conditions under which their analysis of units should be applied:

> If the variation within groups is negligible compared to the variation among groups, or, in the extreme case, when individuals are perfectly segregated into groups by character value . . . . *It would be a serious error* to attempt to derive our estimator [of the relationship between character value and individual fitness within groups] from an analysis that initially ignored the grouping of individuals. (Arnold and Fristrup 1982, p. 123; my emphasis)

Here, Arnold and Fristrup are following a standard statistics textbook treatment of the problem. Under the standard applications of ANOVA, some sort of independent random sampling needs to be assumed; with a single independent variable which can take just two values (treatment or control), this amounts to assuming that the treatment and control are random with respect to other relevant factors. Satisfaction of this assumption is usually incorporated into experimental design. With more than one independent variable (such as in the group selection case) one also needs information about whether the independent variables are correlated.

In the homogeneous populations example, the problem is that one possible explanatory variable (mean group height) is perfectly correlated with another possible explanatorily relevant variable (individual height). Statisticians call this "perfect multicollinearity", and textbooks emphasize that techniques of statistical analysis such as ANOVA are unreliable in such cases, and also in cases where the correlations are high but less than perfect (see, for example, Mosteller and Tukey 1977, especially pp. 280–285, 319–320; Steel and Torrie 1960, pp. 124–131, 194–199).

Sober has also misinterpreted Wimsatt's position on the conditions of applying the structural definition. Sober describes Wimsatt's position as follows:

> Wimsatt stresses that his characterization should apply to the conditions that *actually* obtain in the populations considered; one should not imagine some counterfactual situation that the populations might have been in (but were not) and then see what the criterion of context dependence would imply. (Sober 1984, p. 275)

Taken in context, Wimsatt's point is parallel to claiming that, if one would like to describe the spectra of a given light source, one should use *that* light source and not some other in drawing conclusions. Wimsatt's intention is quite clear from the statement that immediately followed the passage referred to by Sober:

> The issue is perhaps confused by the fact that in order to determine what aspects of environment the fitness of a given chromosome depends upon and what aspects it is invariant over, one may need to look at a variety of different mixed populations in a variety of en-

> vironments. But the relevant environment to use in determining whether its fitness is context-dependent or independent in *this* environment is still *this environment*. (Wimsatt 1981, p. 150)

This claim is precisely what we would expect, given the methodological rules of application of the statistical tool (which in Wimsatt's case is regression analysis).

Sober's attempt to show that Wimsatt's definition is not meant to sustain the necessary sort of counterfactual claim fails. Sober has missed the key distinction necessary to Wimsatt's view, while, admittedly, Wimsatt himself has not made this distinction sufficiently clear. This crucial distinction is between the theoretical description of a system with a specific unit of selection, and the testability of that description. If certain sorts of correlations exist in the natural system, it is not possible to compare the empirical adequacy of two competing models of that system; one cannot determine which set of statistical relations (expectations) in the competing models better matches the natural system.

Sober offers several other counterexamples to his "ANOVA criterion", all of which fail as counterexamples to the additivity criterion because they violate standard methodological rules of application (1984, pp. 272–275). I will not take the space to detail each case here; one example will illustrate the point. Sober suggests the following case. Suppose there are two populations at opposite ends of the universe, and one outproduces the other. Sober thinks that, on the additivity criterion, biologists would perform an analysis of variance and conclude that group selection is operating (1984, p. 274). This is false; according to the methodological constraints advocated by the biologists, the analysis of variance should be performed only within a specific range of shared environments, and data should include a variety of types of systems tested across a variety of environments within that range. Hence, this example does not meet even minimal requirements for the use of the statistical tool.[4]

The fact that the authors (Wimsatt; Arnold and Fristrup) dismissed by Sober explicitly required additional information and strict conditions for the use of the statistical tests indicates that they were not, as Sober implies, naively inferring from statistic to cause, nor is it correct to view the structuralist definition itself as rooted in the "actual".

The situation is more clearly described using the semantic approach to theories: the comparative fit between two distinct theoretical models and a natural system is being evaluated. Causal differences in the models are expressed structurally as different expectations about fitness interrela-

[4]Contrast Sober's example with Sober and Lewontin (1982), in which they discuss the relatively strict conditions under which organisms can be seen as experiencing a common selection regime (1982, pp. 170–171).

tions. The statistical tools are used to generate evidence about which fitness relations hold in the natural system in question. It may be objected that "real causal" differences will not always match the expected fitness interrelations in the model. As we shall see below, however, Sober's causal approach itself is also vulnerable to this objection. This is, I would claim, a fact of life of empirical science. In fact, in evaluating particular cases, Sober actually uses the structuralist approach without acknowledging it.

3.3. Sober's Causal Solution. Sober argues that it is necessary, in order to solve the homogeneous populations problem, to take causal mechanisms into account. In this 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).

This very sensible suggestion of Sober's is, in fact, a textbook application of the additivity approach (Arnold and Fristrup 1982, p. 122; Wimsatt 1981, p. 150).

The second technique supposedly does not require intervention into the system. The biologists, says Sober, can find out what selection forces are at work by looking for "sources". 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 height. Knowledge of this fact seems to indicate that it is "statistical properties of the group" that make it more or less vulnerable. Hence, Sober argues, "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 solution rests on varying the group context of two otherwise identical organisms and observing or predicting the resulting differences in fitness.

In a later discussion of this situation, Sober 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 required by the additivity definition. Just because information is needed does not mean it is available, however. 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.

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 had found the real forces? Sober tells us; if environmental considerations give us reason to think that group selection is operating, then the variance in individual fitness parameters is expected to have certain properties—the very formal properties represented in the additivity definition. Hence, Sober's approach implicitly relies on the additivity definition. The question still remains whether the system indeed has those properties; more information—perhaps obtained through a perturbation experiment—is necessary to provide evidence either way, contrary to Sober's implication.

3.4. Summary. Sober rejects the additivity definition for failing a certain test. However, in his solution to the problem using the causal view, Sober *imports* precisely 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 support their claim by linking it to certain empirical properties of the system—precisely those picked out by the additivity definition.

I take it that one of Sober's main points in his objection to his "ANOVA criterion" is that actual statistics do not determine the ideal (expected) variances of the model, and therefore actual statistics taken from a population cannot answer the empirical question conclusively. This is true. In other words, he is concerned about an accidental fitting of the model and natural system. Sober has not recognized, however, that this very legitimate worry has already been addressed by those promoting a structural approach to units of selection through specifying methodological norms regarding testing and variety of evidence.

Sober is also right in claiming an important place for a causal picture of the natural system in determining units of selection, but he claims too broad a place, thereby blurring the nature of the relations between evidence and theory.

Take the homogeneous populations example; there is an overlap of group

and individual selection model predictions, under the assumption of particular parameter values. In this case, even if group selection were operating, we would still have additivity, but it would be accidental. The way to make progress on the problem is to change the testing situation by reshuffling the group composition. Causal and ecological information is necessary for such testing, because the reshuffling of group composition assumes classes of individuals that are divided into significant traits. The significance of a trait depends on the whole ecological picture involving the organism and the relation between its traits and the environment (see Arnold and Fristrup 1982; Wimsatt 1980; and Lewontin 1974 for detailed discussion of this point).

Sober, in the process of emphasizing the above important methodological aspect of the units of selection problem, has neglected the fact that those defending the structuralist approach have explicitly discussed it as well. Then what, one may ask, is the difference between Sober's causal view and the structuralist view that I have explicated? One difference is that, in his applications of the causal view, Sober relies on the structuralist definition. I submit that this is because the structural definition is more basic. (See Brandon 1986 for a discussion of some shortcomings of Sober's definition.)

Finally, the additivity criterion has the advantage, over Sober's causal view, of revealing the connection between the theory of evolution by natural selection and the evidence required for a units of selection claim. There is some sense in which Sober's focus on causes and his neglect of the intended uses of statistical analyses leads us dangerously far from the theory-evidence relations. The claim that one has located the real selective causes acting on an evolving system is dependent upon the empirical adequacy of certain models. The causes are inferred from a satisfactorily confirmed claim about a model; developing a plausible causal hypothesis is not good enough.

For example, biologists have a causal story for organismic selection in the controversial case of the myxoma virus, but there is also a group selection causal story. Sober, in his discussion of this case, does not keep the empirical problem in focus. Through examining the myxoma case below, I show that, to biologists, pinpointing the correct cause involves inferring it from certain empirical properties of the natural system—precisely those properties at stake in the additivity definition.

4. Example: Myxoma

4.1. The Debate. The importance of theoretical and methodological assumptions in the evaluation of units of selection is particularly clear in the case of the myxoma virus. The virus, which infects rabbits, was in-

troduced into Australia in order to control the rabbit population. At first, the virus killed at least 99% of the exposed rabbits, but it subsequently became less effective (Fenner 1965, p. 492). When wild rabbits were tested against laboratory strains of the virus, it was found that they had become resistant. The development of such resistance would be expected from simple organismic selection. Viruses taken from the wild, though, when tested against laboratory rabbits, were found to have become less virulent than the laboratory strains (Fenner 1965).

Is the decrease in virulence in myxoma the result of group selection or of individual selection? Futuyma argues that if the fitness of an individual parasite is lowered by the death of its host, avirulence is advantageous. The myxoma virus is spread from host to host by mosquitoes which only bite live rabbits. Rabbits tested with pure, highly virulent strains were usually dead within 9–13 days, while those infected with an avirulent type lived an average of 26 days following infection (Levin and Pimentel 1981, pp. 312–313). Hence, virulent strains would have a lower likelihood of being spread (and therefore have lower fitness). Fenner concluded that the critical factor in the evolution of avirulence was the longer survival times of rabbits infected with less virulent strains (1965, pp. 493–494). Futuyma concludes that the avirulence evolved to "benefit individual parasites" (Futuyma 1979, p. 455, cited in Wilson 1983).

Alexander and Borgia also argue that the myxoma virus evolved through organismic selection. They base their conclusion on the (undefended) assumption that when a mosquito bites a rabbit, either a single virus particle or a set of particles of the same strain is injected (1978, pp. 452–453).

Lewontin, in contrast to the above authors, views each set of virus particles injected into a rabbit as a deme, that is, a population of genetically different virus strains, some more virulent, some less. When a rabbit dies, the deme goes extinct. If a number of avirulent particles have the misfortune of being injected into a rabbit along with a quite virulent particle, their relative fitness will be greatly affected by the company they keep (Lewontin 1970). Levin and Pimentel, in their mathematical treatment of Lewontin's proposal, argue that "within a parasite group or colony, selection is for higher growth rates, despite the fact that this endangers the survival of the host and ultimately of the whole colony" (in this case, the virulent types are superior competitors within demes) (1981, p. 308); virulence may be a consequence of the capacity to multiply rapidly. It is this reproductive success that leads Lewontin to claim that the reduction in virulence "cannot be explained by individual selection" (1970, p. 15).

When the global population consisting of the ensemble of all colonies is considered, though, it is necessary to consider the parasite genotype's influence on the survival probability of the host, in computing the overall

fitness of the genotype. On the global level, then, Levin and Pimentel expect there to be selection against the most virulent type (1981, p. 308). The result is an increase in avirulence in the *global* population, despite the fact that the avirulent strains lack a selective advantage *within* demes (Lewontin 1970, pp. 14–15).

Alexander and Borgia correctly remark that Lewontin's interpretation "requires that, as a rule, less- or more-virulent viruses be mixed in the same rabbits" (1978, p. 453). (This point is not addressed by Levin and Pimentel, who simply see each host as a heterogeneous group (1981, p. 314).) The problem of viewing ensembles or groups that may be homogeneous as biologically meaningful and diverse "groups" is not addressed. In another treatment of the problem, M. E. Gilpin simply assumes that a variety of genotypes is injected (1975, pp. 96–97). Alexander and Borgia conclude: "If the population of rabbits is composed largely of individuals infected with pure more-virulent and pure less-virulent strains (i.e., clones), the relevant selection on the virus might appropriately be described as occurring on the individual level" (1978, p. 453).

4.2. Discussion. It is clear that the theorists do not agree on the composition of the groups. The additivity criterion clarifies the pivotal nature of their assumptions about group composition. In order to determine whether there is additivity of variance of fitness parameters, information about group composition and contribution to the global gene pool must be both collected and analyzed. D. S. Wilson has explained the essential role of such information. Wilson argues that it is quite possible to construct an organismic selection model which will produce the same outcome (gene frequency values) as a group selection model. The overall genotype fitness (averaged over all local groups) is used to calculate the global gene frequencies. He writes, "Indeed, using this methodology, it is easy to conclude that the [group level] character evolves by individual selection because it has the highest relative fitness throughout the global population and because evolution within local groups was not monitored" (1983, p. 171).

Note that in the myxoma case, both individual selection models and group selection models predicted an overall increase in the avirulent strains; the final outcome in both cases was the same. Those who conclude that avirulence evolves by organismic ("individual") selection are comparing only the global fitnesses of virulent and avirulent types. Wimsatt has argued that biologists often assume that variance in fitness is additive. Given this assumption, there is no question of performing an analysis to test for non-additivity (1980, p. 230).

In addition, G. C. Williams' parsimony claim discourages such testing. Williams claims that group selection models should not be considered,

unless the organismic selection model is inadequate empirically. He writes:

> In explaining adaptations, one should assume the adequacy of the simplest form of natural selection, that of alternative alleles in Mendelian populations, unless the evidence clearly shows that the theory does not suffice. (Williams 1966, p. 55)

The additivity criterion reveals the potential for dogmatic abuse of such sensible-looking advice.[5] "The evidence", as Williams puts it, is not found, it is created. *Not performing* analyses of group composition and contribution to the global gene pool virtually guarantees (except in cases where the organismic model's predictions of gene frequency are significantly in error) that no evidence will be found which will reveal the inadequacy of the simpler (genotypic or genic) models. The myxoma case is an excellent example. There is an organismic selection model that is able to account for the global gene frequencies; hence, if we follow Williams' advice, we would not even consider a group selection model. Yet there is some reason to think that group selection might be occurring—should this be dismissed out of hand because of the mere presence of a successful organismic level model? Understanding the structure of selection models and the information needed to perform an adequate comparison reveals quite clearly the dogmatism of Williams' maxim, as it is usually applied.

The theorists do not agree on the composition of the groups being injected, which in turn has predictable consequences—according to the additivity criterion—on whether they support a group or organismic selection model. Those biologists who assert that the groups are homogeneous (and they give their ecological and causal reasons for accepting this assumption) find organismic selection. (Sober is quite right in insisting that the issue is the homogeneity of the groups, and not their relatedness, that is, whether they are clones (1984, p. 334).) Similarly, those who assume that each group is heterogeneous (they also have their causal reasons—Lewontin, for example, cites the likelihood of multiple infection, and the spread of heterogeneity (personal communication)) can utilize the additivity criterion to conclude that group selection may be operating.

But surely there is a fact of the matter. Either the injected particles are mixed, or they are not. This is an empirical question. 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, statistical analyses could be done; if they are homogeneous, experiments could perhaps be done to manipulate the populations in order to get the necessary information.

[5]Note, however, that the additivity criterion also has a bias toward attributing selection to the lowest levels (see discussion in Wimsatt 1981, p. 146).

In this case, both sides have provided ecological and causal support for 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.

4.3. Sober's Causal Solution. Sober presents a causal analysis of the myxoma debate, in which he continues his attack on what he takes to be the structuralist view. In addressing this case, Sober 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 specific 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 yields statistics that conform to the particular set of model relations described in the additivity definition. I conclude that Sober's view is dependent upon the additivity criterion.

Sober's discussion of this case also reveals his misunderstanding of the structuralist approach:

> Biologists sometimes hold that the group selection account depends on the assumption that a rabbit is innoculated with viruses of different degrees of virulence. Clearly, within-rabbit variation is necessary for there to be any competition within a group. However, the claim here is that within-group variation is necessary for the existence of between-group selection. This is the very antithesis of the analysis of variance criterion of group selection, which would conclude in this case that there *must be group selection and cannot possibly be individual selection.* I have already noted that I dissent from the ANOVA characterization. It is curious that this mode of analysis and one that demands that members of the group not be clones of each other are in such *blatant contradiction.* (Sober 1984, pp. 333–334; my emphasis)

There are several problems with Sober's account. First, the biologists Sober refers to here are Alexander and Borgia (mentioned above), who, he reports, "hold that if rabbits were bitten only once and innoculated with a pure strain, then the group selection hypothesis would not be tenable" (Sober 1984, p. 332; see Alexander and Borgia 1978, pp. 452–453). The biologists' views are clearly consistent with the additivity definition; if there is no visible non-additivity of variance in fitness at the individual level, the group level will not be considered (given this limited

information). Sober's causal approach, however, fails to account for the biologists' views.

Second, Sober is incorrect in reporting that the biologists think that within-group variation is necessary for the existence of group selection. Their claim is better understood in the context of the homogeneous populations claim; within-group variation is necessary only for *determining* whether group selection is operating. It could well be operating among homogeneous groups, but neither the additivity criterion nor Sober's causal analysis could determine this.

Third, Sober claims that on the ANOVA criterion, one would "conclude in this case that there must be group selection and cannot possibly be individual selection" (1984, pp. 333–334). This may be true on Sober's own ANOVA criterion, but it is simply false on the actual structuralist views he claims to be attacking. On the additivity criterion, given the lack of other information, one would conclude in favor of individual selection, *if* one made any inference at all (which is dubious, given the violation of the methodological prerequisites). Hence Sober's "blatant contradiction" between the structuralist views and biologists' practice (see above quote) does not even exist (see also Sober 1984, p. 349).

Finally, the additivity criterion allows us to isolate a crucial theoretical difference missed by Sober. In discussing the various interpretations of the myxoma situation, Sober reports that Futuyma "does not explicitly dissent from any of the empirical assumptions that Lewontin (1970) makes" (1984, p. 332). While it is true that Futuyma's differences with Lewontin in empirical assumptions are not explicit, they are very important. It is quite clear that Futuyma assumes that the individuals in the injected groups are of the same genotype (same virulence) while Lewontin assumes they are of different genotypes. As we saw in section 4.2, this empirical assumption is pivotal in testing the competing units of selection hypotheses, in this case. According to my understanding of Sober's causal view, he should agree with this analysis, and should require the same empirical information, just as he eventually did in the homogeneous populations example. The fact that he himself did not perform an adequate analysis therefore is not decisive, though I do think this failure undermines any claim for heuristic advantages of his causal approach.

5. Methodological Considerations Revisited. Sober, in attempting to eliminate the rival structuralist approach while motivating his own causal view, states:

> It is not to be doubted, however, that further stipulations may be imposed that ensure that an ANOVA criterion will coincide with a plausible causal analysis. My feeling is that these further require-

> ments will be sufficiently substantive to permit one to conclude that the analysis of variance is not what lies at the heart of the idea of a unit of selection. (Sober 1984, p. 275)

And later:

> [ANOVA] 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. (Sober 1984, p. 317)

First, only Sober thinks that the analysis of variance "lies at the heart of" the idea of units of selection on the structuralist view; Wimsatt, in contrast, claims: "My definition is anchored in the theoretical structure of the modern mathematical theory of evolution as well as in the classical formulation of 'Darwin's principles'" (1981, p. 151; see section 2.1, above). The conditions involving variance in the additivity definition are not intended to stand on their own merits; the point is that evolution by selection models, as they are widely understood, require that the variance in fitness parameters be additive. Furthermore, the analysis of variance was not being presented by Wimsatt or Arnold and Fristrup as an "analysis of causation"; its purpose is as a statistical tool in developing evidence regarding the interrelations of group and individual level fitness parameters and traits (see Arnold and Fristrup 1982, pp. 121, 128; also Griesemer and Wade 1988).

The key point is this. There is a difference between having an adequate theoretical definition—for example, a description of how the theoretical model must look if a model-entity is to be considered a unit of selection—and having all conceivable cases in nature be susceptible to determination under that definition. Sober has presented a number of cases which are not (immediately) testable. But his definition is no better off with regard to these cases.

The causal information emphasized by Sober is identical to the information demanded by the methodological rules of application for the statistical tools used in conjunction with a structuralist definition. Sober has a very strong case for claiming that this information about the system is necessary for the empirical evaluation of claims about units of selection. I have argued, however, that Sober has thrown out the baby with the bathwater. Sober has confounded the adequacy of a theoretical description of a unit of selection with the testability of that description. This is not surprising, since he has also confounded the structural definition itself

with the statistical tools used to evaluate applications of that definition. Through these confusions, Sober ends up rejecting the structural and conceptual underpinnings of his own definition.

In this paper, I have clarified the structuralist approach to the units of selection problem, using the semantic view of theories as a framework. I have argued that Sober's attack on the "ANOVA criterion"—which he believes to undermine the structuralist approaches taken by Wimsatt, Arnold and Fristrup—fails, since his "ANOVA criterion" differs in crucial ways from their views. Furthermore, I have shown that his causal view is actually based on the additivity definition, which is a more precise version of Wimsatt's original "context-dependence" definition. Finally, I have demonstrated the power and utility of the additivity definition through the analysis of the controversial case of the myxoma virus.

REFERENCES

Alexander, R. D., and Borgia, G. (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* (2): 113–129.

Arnold, A. J., and Wade, M. J. (1984), "On the Measurement of Natural and Sexual Selection: Theory", *Evolution 38*: 709–719.

Beatty, J. (1980), "Optimal-Design Models and the Strategy of Model Building in Evolutionary Biology", *Philosophy of Science 47*: 532–561.

———. (1981), "What's Wrong with the Received View of Evolutionary Theory?" *PSA 1980*, vol. 2. East Lansing, Michigan: Philosophy of Science Association, pp. 397–426.

———. (1982), "The Insights and Oversights of Molecular Genetics: The Place of the Evolutionary Perspective", *PSA 1982*, vol. 1. East Lansing, Michigan: Philosophy of Science Association, pp. 341–355.

Brandon, R. N. (1982), "The Levels of Selection", *PSA 1982*, vol. 1. East Lansing, Michigan: Philosophy of Science Association, pp. 315–323.

———. (1986), "Review of *The Nature of Selection*", *The Philosophical Review 95*: 614–617.

Brandon, R. N., and Burian, R. M. (eds.) (1984), *Genes, Organisms, Populations: Controversies over the Units of Selection*. Cambridge, Mass.: MIT Press.

Buss, L. W. (1983), "Evolution, Development, and the Units of Selection", *Proceedings of the National Academy of Sciences, USA 80*: 1387–1391.

Crow, J. F., and Aoki, K. (1982), "Group Selection for a Polygenic Behavioral Trait: A Differential Proliferation Model", *Proceedings of the National Academy of Sciences USA 79*: 2628–2631.

Damuth, J., and Heisler, I. L. (1988), "Alternative Formulations of Multilevel Selections", *Biology and Philosophy* (Oct.): 407–430.

Dawkins, R. (1976), *The Selfish Gene*. New York: Oxford University Press.

———. (1982), *The Extended Phenotype*. San Francisco: W. H. Freeman.

Fenner, F. (1965), "Myxoma Virus and *Oryctolagus cuniculus*", in H. G. Baker and G. L. Stebbins (eds.), *The Genetics of Colonizing Species*. New York: Academic Press, pp. 485–501.

Futuyma, D. J. (1979), *Evolutionary Biology*. Sunderland, Mass.: Sinauer.

Gilpin, M. E. (1975), *Group Selection in Predator-Prey Communities*. Princeton, NJ: Princeton University Press.

Griesemer, J. R., and Wade, M. J. (1988), "Laboratory Models, Causal Explanation, and

Group Selection", *Biology and Philosophy* (Jan.): 67–96.
Hull, D. L. (1980), "Individuality and Selection", *Annual Review of Ecology and Systematics 11*: 311–332.
Humphreys, P. W. (1985), "Quantitative Probabilistic Causality and Structural Scientific Realism", *PSA 1984* vol. 2. East Lansing, Michigan: Philosophy of Science Association, pp. 329–342.
Kempthorne, O. (1957), *An Introduction to Genetic Statistics*. New York: Wiley.
Levin, S. and Pimentel, 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.
Li, C. C. (1964), *Introduction to Experimental Statistics*. New York: McGraw Hill.
Lloyd, E. A. (1984), "A Semantic Approach to the Structure of Population Genetics", *Philosophy of Science 51*: 242–264.
———. (1986), "Thinking about Models in Evolutionary Theory", *Philosophica 37*: 87–100.
———. (1987), "The Confirmation of Ecological and Evolutionary Models", *Biology and Philosophy 2*: 277–293.
———. (1988), *The Structure and Confirmation of Evolutionary Theory*. Westport, Conn.: Greenwood Press.
Maynard Smith, J. (1976), "Group Selection", *Quarterly Review of Biology 51*: 277–283.
Mayo, D., and Gilinsky, N. (1987), "Models of Group Selection", *Philosophy of Science 54*: 515–538.
McCauley, D. E., and Wade, M. J. (1980), "Group Selection: The Genetic and Demographic Basis for the Phenotypic Differentiation of Small Populations of *Tribolium castaneum*", *Evolution 34*: 813–821.
Mosteller, F., and Tukey, J. W. (1977), *Data Analysis and Regression*. Reading, Mass.: Addison-Wesley.
Nunney, L. (1985), "Group Selection, Altruism, and Structured-Deme Models", *American Naturalist 126*: 212–230.
Roughgarden, J. (1979), *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. New York: MacMillan.
Ruse, M. (1980), "Charles Darwin and Group Selection", *Annals of Science 37*: 615–630.
Sober, E. (1981), "Holism, Individualism, and the Units of Selection", *PSA 1980*, vol. 2. East Lansing, Michigan: Philosophy of Science Association.
———. (1984), *The Nature of Selection*. Cambridge, Mass.: MIT Press.
Sober, E., and Lewontin, R. C. (1982), "Artifact, Cause, and Genic Selection", *Philosophy of Science 49*: 157–180.
Steel, R., and Torrie, J. (1960), *Principles and Procedures of Statistics: With Special Reference to the Biological Sciences*. New York: McGraw Hill.
Thompson, P. (1983), "The Structure of Evolutionary Theory: A Semantic Perspective", *Studies in the History and Philosophy of Science 14*: 215–229.
———. (1985) "Sociobiological Explanation and the Testability of Sociobiological Theory", in J. H. Fetzer (ed.), *Sociobiology and Epistemology*. Dordrecht: Reidel, pp. 201–215.
———. (1986) "The Interaction of Theories and the Semantical Conception of Evolutionary Theory", *Philosophica 37*: 28–37.
van Fraassen, B. C. (1980), *The Scientific Image*. Oxford: Clarendon Press.
Vrba, E. S., and Gould, S. J. (1986), "The Hierarchical Expansion of Sorting and Selection: Sorting and Selection Cannot Be Equated", *Paleobiology 12* (2): 217–228.
Wade, M. J. (1978), "A Critical Review of the Models of Group Selection", *Quarterly Review of Biology 53*: 101–114.
———. (1985), "Soft Selection, Hard Selection, Kin Selection and Group Selection", *American Naturalist 125*: 61–73.
Wade, M. J., and McCauley, D. E. (1980), "Group Selection: The Phenotypic and Gen-

otypic Differentiation of Small Populations", *Evolution 34*: 799–812.
Williams, G. C. (1966), *Adaptation and Natural Selection*. Princeton, N.J.: Princeton University Press.
Wilson, D. S. (1983), "The Group Selection Controversy: History and Current Status", *Annual Review of Ecology and Systematics 14*: 159–187.
Wimsatt, W. (1980), "Reductionist Research Strategies and Their Biases in the Units of Selection Controversy", in T. Nickles (ed.), *Scientific Discovery: Case Studies*. Dordrecht: Reidel, pp. 213–259.
———. (1981) "Units of Selection and the Structure of the Multi-Level Genome", *PSA 1980*, vol. 2. East Lansing, Michigan: Philosophy of Science Association.