A Semantic Approach to the Structure of Population Genetics

Elisabeth A. Lloyd

Philosophy of Science, Vol. 51, No. 2. (Jun., 1984), pp. 242-264.

Stable URL:
http://links.jstor.org/sici?sici=0031-8248%28198406%2951%3A2%3C242%3AASATTS%3E2.0.CO%3B2-H

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

A SEMANTIC APPROACH TO THE STRUCTURE OF POPULATION GENETICS*

ELISABETH A. LLOYD

Department of Philosophy
Princeton University

A precise formulation of the structure of modern evolutionary theory has proved elusive. In this paper, I introduce and develop a formal approach to the structure of population genetics, evolutionary theory's most developed sub-theory. Under the semantic approach, used as a framework in this paper, presenting a theory consists in presenting a related family of models. I offer general guidelines and examples for the classification of population genetics models; the defining features of the models are taken to be their state spaces, parameters, and laws. The suggestions regarding the various aspects of the characterization of population genetics models provide an outline for further detailed research.

1. Introduction. Known to be unlike Newtonian mechanics but also unlike Creationist biology, modern evolutionary theory has a structure which has proved difficult to characterize. Recently, John Beatty (1980, 1981, 1982) and Paul Thompson (1983) have approached the problem of describing the structure of evolutionary theory using the semantic approach developed by P. Suppes, B. C. van Fraassen and F. Suppe.

Advantages of the semantic view over the logical positivist approaches which until recently, dominated discussions of theory structure, have been presented by Suppes (1957, 1961, 1962, 1967), van Fraassen (1970, 1972, 1974, 1980), Suppe (1972, 1973, 1974, 1977), and Stegmuller (1976), and will not be discussed here.

The semantic approach, in which I include both the "set theoretic" and "state space" approaches, has been used to describe the structure of Newtonian mechanics, equilibrium thermodynamics, quantum mechanics, and parts of biological theory (Sneed 1971; Stegmuller 1976; Suppes 1957; Wessels 1976; Moulines 1975; van Fraassen 1970, 1972, 1974; Suppe 1974a; see Suppe 1974, 1979 for a summary of the semantic view and its literature).

I shall assume the following positions in this paper, without defense: the semantic view—in particular, the state space version—is more suited

*Received June 1983; revised January 1984.

†I wish to thank Bas van Fraassen, Richard Lewontin, David Hull, Lindley Darden, Mary Williams, Steven Orzack, and two anonymous referees for this journal for their questions, criticism, and suggestions.

Philosophy of Science, 51 (1984) pp. 242–264.

to a description of evolutionary theory than any axiomatic view (Beatty 1980, 1981, 1982; Thompson 1983); the semantic view can provide a richer and more useful description of the structure of a theory than axiomatic approaches (see especially Suppe 1972, van Fraassen 1980); and finally, the semantic view is capable of formally describing theories which are not describable using any axiomatic approach (van Fraassen, forthcoming).

My goal in this paper is to provide an introduction and further development of the semantic approach to the structure of population genetics theory, the most formal and developed sub-theory of contemporary evolutionary theory.[1] In the process of providing a formal framework for the detailed description of the theory, I provide a means by which precise analysis of theoretical problems can be carried out. Ultimately, the utility of describing population genetics (and evolutionary theory as a whole) through the semantic view rests on the ability of the semantic view to provide an analytical framework sensitive to the relevant theoretical problems. A working model, i.e. an actual semantic description of the theory, must therefore be available before we can evaluate its power as an analytical tool.

After a brief look at the semantic view in the rest of Section 1, I present a range of population genetics models in Section 2, in order to illustrate the variety and general character of the theory, and the suitability of the semantic approach. In Section 3, I discuss particular problems encountered in describing population genetics models using the semantic view of theory structure as a framework. Promising lines for further research that makes use of this approach are noted throughout the paper.

1.1 The Semantic View of Theory Structure. According to the semantic view of theory structure, a scientific theory specifies certain kinds of systems. The systems specified by a scientific theory are *ideal*; they are used in scientific explanation through claims that certain systems in the natural world are of the kind defined by the theory. The semantic view offers a formal approach to analyzing these systems, which are usually understood as mathematical structures. There are different ways to describe these structures formally: the set-theoretic predicate approach, de-

[1]Populations genetics theory, because of its advanced formal development, lends itself to analysis by the semantic view. My use of population genetics as a starting place for the analysis of the structure of evolutionary theory as a whole does *not* imply, however, that population genetics is assumed to constitute the "core" or "foundation" of the theory. On the contrary, I assume that population genetics, as a set of structures, is embedded in the larger structure called evolutionary theory. It seems likely that if the semantic approach can be used to describe the most structured segment of the theory, it may provide a good approach to the theory as a whole.

veloped by Suppes, Sneed, and Stegmuller, involves description by a set-theoretic predicate; the state space approach, employed by van Fraassen and Suppe, describes the structures as configurations of certain mathematical spaces. In this paper we will be using van Fraassen's version of the state space approach (see Suppe 1979 for discussion of differences among the various semantic approaches).

In general, a structure presented by a theory (understood as intended to represent empirical phenomena) is a *model* of the theory if it satisfies the theorems of the theory. In a semantic definition, the set of sentences that are theorems of the theory is defined *not* in relation to a set of axioms, but by directly defining the class of structures; for any given language L, the theorems of the theory in L are the sentences of L which are satisfied in all these structures. Reference to syntax or to a syntactically defined set of theorems is thus unnecessary. The models picked out are mathematical models of the evolution of states of a given system, in both isolation and interaction, through time. This selection is achieved by conceiving of the ideal system as capable of a certain set of *states*—these states are represented by elements of a certain mathematical space, the *state space* (van Fraassen 1970, p. 238; 1972, pp. 303, 305). (NB In this paper, "models" and "systems" *always* refer to *ideal* systems; when the actual biological systems are being discussed, they will be called "empirical" systems.) The variables used in each mathematical model represent various measurable or potentially quantifiable physical magnitudes. Classically, any particular configuration of values for these variables is a *state* of the system, the state space or "phase space" being the collection of all possible configurations of the variables.

The theory itself represents the behavior of the system in terms of its states; the rules or laws of the theory (i.e. laws of coexistence, succession, or interaction) can delineate various configurations and trajectories on the state space. Under the semantic view, these structures, "being phase spaces of configurations imposed on them in accordance with the laws of the theory", are themselves seen as constitutive of the theory (Suppe 1977, pp. 226–227).

Description of the structure of the theory itself therefore involves only the description of the set of models presented by the theory. It is crucial, then, to discuss the various necessary components of describing a model.

Construction of a model within the theory involves assignment of a location in the state space of the theory to a system of the kind defined by the theory. Potentially, there are many kinds of systems that a given theory can be used to describe; limitations come from the dynamical sufficiency (i.e. whether it can be used to describe the system accurately and completely) and the effectiveness of the laws used to describe the

system and its changes. Thus, there are two main aspects to defining a model. First, the state space must be defined—this involves choosing the variables and parameters with which the system will be described; second, coexistence laws, which describe the structure of the system, and laws of succession, which describe changes in its structure, must be defined.

Defining the state space involves defining the set of all the states the system could possibly exhibit. Certain mathematical entities—in the case of the models we shall be looking at, these are *vectors*—are chosen to represent these states. The collection of all the possible values for each variable assigned a place in the vector is the state space of the system. The system and its states can have a geometrical interpretation: the variables used in the state description (i.e. state variables) can be conceived as the axes of a Cartesian space. The state of the system at any time may be represented as a point in that space, located by projection onto the various axes.

The family of measurable physical magnitudes, in terms of which a given system is defined, also includes a set of parameters. The biologist R. C. Lewontin defines parameters as values that are not themselves a function of the state of the system[2] (1974, pp. 7–8). Thus, a parameter can be understood as a fixed value of a variable in the state space—topologically, setting a parameter seems to amount to limiting the number of possible structures in the state space by reducing the dimensionality of the model (see Section 3.4).

Laws, used to describe the behavior of the system in question, must also be defined in a description of a model or set of models. Laws have various forms: in general, coexistence laws describe the possible states of the system in terms of the state space, while changes in the state of the system are described by laws of succession. Suppe has given a complete, formal classification of succession and coexistence laws according to the semantic view (Suppe 1976); detailed discussion of the various evolutionary laws in terms of his system cannot be done here. Rather, we will discuss, in Section 3.3, certain problems of classification encountered in analyzing evolutionary laws.

At this point, I would like to draw a distinction. Consider the problem of determining the most appropriate state space with which to represent genetic changes in populations; this is, to an extent, an *empirical* question. Determination of the types or categories of state spaces *used* in pop-

[2]Lewontin notes that although parameters can involve time and can change *over* time, they are not correlated to the variable value as it changes over time (personal communication).

ulation genetics, however, and the relation of these state space types to determination of the structures comprising the theory, are *philosophical,* rather than empirical, questions.

With this distinction in mind, I examine a few examples of population genetics models in Section 2. My purpose is threefold. First, having presented the general terms in which I propose to describe population genetics models, I illustrate these terms through a few actual models. This is an easy and natural task, since much of the theory is *presented* in these same terms. The second goal, then, is to demonstrate that the state space version of the semantic approach provides a natural reconstruction of the theory—less arbitrary than, for example, an axiomatic approach—since it makes sense of the theory as presented. Third, I hope to show, particularly through the example in Section 2.3, that the semantic approach highlights some features of population genetics theory which are theoretically important. Detailed discussion of the description of population genetics models is presented in Section 3.

2. Models in Population Genetics. Population genetics, as characterized for example by Lewontin, is the "study of the origin and dynamics of genetic variation within populations" (1974, p. 12). The notion of "gene frequency" is fundamental; description of both changes and equilibria of gene frequencies in populations is a primary goal of population genetics theory.

The Hardy-Weinberg law, an equilibrium law of gene frequencies, serves as the foundation of population genetics theory. Take a single locus (gene) with only two alleles, A and a (alternate types of that gene), in a population of diploid organisms (organisms with paired chromosomes). Take the frequency of allele A to be p, the frequency of a to be q (with $p + q = 1$). The Hardy-Weinberg law gives the genotype frequencies of the zygotes (the potential next generation) by the equation:

$$p^2AA + 2pqAa + q^2aa = 1$$

The system represented by the above equation is a "one-locus" system, i.e. calculations are performed assuming the complete isolation and independence of the alleles at each locus. Furthermore, a completely random mating pattern is assumed, i.e. the genotype makes no difference to the probability of mating to any given genotype. It is also assumed that each genotype contributes equally to the pool of gametes from which the zygotes are randomly "chosen", but this is not generally the case. The comparative contribution of each genotype to the next generation is its fitness value. More complicated models, involving the individual (w) or population fitness value ($\bar{w}$), in conjunction with the basic Hardy-Weinberg law, are necessary in order to describe all but the most simplistic

system. In the rest of this section, I shall present a few examples of these more complex models.

2.1 Deterministic Models. Consider the case in which carriers of a certain genotype contribute a larger proportion of gametes (reproductive cells carrying only half of the complement of the organism's chromosomes) to the gene pool than the other genotypes (of the locus under consideration). Some modification in the Hardy-Weinberg equation is necessary, since it represents equal contributions from genotypes to the gene pool. The difference in contribution is a measure of the "fitness" or "selective value" (w) of a given genotype. The fitness of the genotype contributing the most is taken by convention as 1; the other genotypes have fitnesses of (1-s), where the value of s is the selection coefficient of that genotype.

In a case of simple dominance, where the fitness of genotypes AA and Aa equals 1, and the fitness of aa is (1-s), we can predict the frequencies (in the ideal system) of the genotypes after selection through a modification of the Hardy-Weinberg equation:

$$p^2AA + 2pqAa + (1 - s)q^2aa = 1 - sq^2$$

We can then calculate the frequency p' of the A allele in the next generation:

$$p' = (p^2 + pq)/(1 - sq^2) = p/(1 - sq^2)$$

So the increment, Δp of the frequency of allele A in one generation is:

$$\Delta p = spq^2/(1 - sq^2)$$

(from Dobzhansky 1970, p. 102). If s is very small (0.01 or less), it is possible to calculate analytically an equilibrium value p_E (such that $p_E' = p_E$) for the frequency of A. Calculations of the number of generations taken for a given change in gene frequency are then also possible (Maynard Smith 1968, pp. 74–75; Lewontin 1967, p. 81). This sort of model is a *deterministic* model since, given the initial conditions of the population—in this case the initial gene frequency—and any set of parameters—in this case the selection coefficient—the precise condition of the population at some future time can be predicted (Lewontin 1967, p. 81; see Section 3.3).

More than one parameter can be incorporated into the basic model based on the Hardy-Weinberg equations. For example, mutation rates can be included, so that the frequency in the next generation depends both on selection and on mutation. Take the mutation rate from a to A as μ, where μ is defined as the probability that a has mutated to A within the time of one generation. The frequency in the next generation is calculated as fol-

lows (where s is the negative fitness coefficient of A):

$$p' = \frac{2p(1-s) + 2\mu q(1-ps)}{2 - 2s(p^2 + 2pq)} = \frac{p - ps + \mu q - \mu pqs}{1 - sp(p + 2q)}$$

In other words, the frequency of allele A in the next generation is calculated in terms of both parameters, μ and s. Once again, this is a deterministic model, since a definite gene frequency results. The model can be simplified greatly by assuming that p is very small, which is plausible under the assumption that the A allele is deleterious, and hence would be maintained only at low frequency. If p is small, then we can approximate using

$$p' = p - ps + \mu q$$

and at the equilibrium state

$$p_E = p_E - sp_E + \mu(1 - p_E)$$

or, if μ is small relative to s, we can approximate by

$$p_E = \frac{\mu}{s}$$

(Maynard Smith 1968, p. 79).

In general, in deterministic models the initial conditions of the population are represented by an ordered *set* of values of variables, i.e. a vector. The above examples used a set of only one variable, p. A parameter set is also specified, μ and s in the previous example; the value for the variable after a certain time interval is given by equations incorporating the parameters. Such equations embody the dynamical *laws* of change for the system; they entail a theory about the equilibrium states of the system.

2.2 Stochastic Models. With some evolutionary processes, a number of different results are possible. The mathematical models must, in these cases, represent the relative chances of the occurrence of each of the possible results. In one example of such a probabilistic or "stochastic" model, the probability that an allele with selective coefficient s will reach fixation (i.e. have frequency of 1) within a population of effective size N over many generations is evaluated. The result of this type of model will be a probability distribution rather than the single value specified by a deterministic model. That is, the model will specify the probabilities of the various possible final states, but will *not* say which one will occur, even if we know *only* one will occur. The model can be understood as having "ergodic properties", i.e. at equilibrium there is some final prob-

ability $p(1)$ of the system being in state 1, another probability $p(2)$ of the system being in state 2, etc. (Lewontin 1967, p. 81).

Thus, p_s in

$$p_s = \frac{1 - e^{-4Nsp}}{1 - e^{-4Ns}} \qquad \text{(Kimura \& Ohta 1971, pp. 9–10)}$$

(where p is the frequency of the allele at the beginning of the process) can be understood roughly as the proportion of total populations of effective size N, which, confronted with an allele with selection coefficient s, would eventually reach a frequency of 1 for that allele (i.e. eliminate all other alleles at that locus).

The need for stochastic models arises when it is necessary to know more than the average of a range of values, i.e. when we need to measure variability. The basic way to handle essential (i.e. necessary) variability is to use an appropriate probability distribution that represents the chance that an individual selected at random will be found to have any given value or range of values. A number of different types of distributions are useful.[3]

In most stochastic models in population genetics, the biologist attempts to predict the way in which the "ensemble of populations" changes in time, and what the equilibrium distributions look like. This is basically statistical mechanics, and the problems can be solved by borrowing methods from that branch of physics (Lewontin 1968, p. 82). For example, in order to solve the distribution function of the gene frequency at equilibrium, change in the ensemble is often approximated as a partial differential equation in time, though this is not always possible or practical (see Bailey 1968, p. 42).

As the mathematical models used to represent genetical phenomena incorporate more parameters and information—in order to make them match the empirical results more closely—it becomes more difficult to

[3]For continuous measurements, a normal (Gaussian) probability curve

$$y = \frac{1}{\sigma\sqrt{2\pi}}\, exp - \frac{(x - \mu)^2}{2\sigma^2} \quad (-\infty < x < \infty)$$

provides an accurate representation. The equation is put in terms of two parameters, μ: the mean, and σ: the standard deviation. In application, the likelihood of observing an individual with the character in the range from x_1 to $x_2 = \int_{x_1}^{x_2} y dx$.

For discrete variables, a binomial distribution can be used. Taking n individuals for which the (independent) chance of having a certain character is p, the chance of observing r individuals with that character is

$$\frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \quad (0 \leq r \leq n)$$

(equations from Bailey 1967, p. 25).

arrive at precise mathematical solutions. Yet there is still a need to formulate the complex models in well-defined mathematical terms (Bailey 1968, p. 43). In cases in which approximations cannot be done, simulations are often used. These simulations are "realizations of a stochastic model which are strictly analogous to possible realizations of a real-life process" (Bailey 1968, p. 43). In other words, a computer is used to produce a large number of artificial (as opposed to actual, laboratory) realizations of the stochastic process in question. A large number of runs are executed, using alternative combinations of the values of the parameters, which are fixed for each particular model. With the collection of model results in hand, the biologist can then compute the means, variances, etc. for the models. Simulation models can also aid future research by providing 1) information about what measurements might be useful, and 2) a means of estimating parameter values (Starfield et al. 1980, pp. 338, 353; Bailey 1968, p. 42).

2.3 Example. Lewontin and Dunn's work on polymorphism in the house mouse, discussed below, provides an interesting example of both the simulation of a stochastic process, and an explicit comparison between deterministic and stochastic models.

Lewontin and Dunn (1960) examined a situation in which the existence of a mutant *t* allele at a specific locus is widespread among the populations studied. The polymorphism (presence of more than one type of allele of the gene) is unusual. Strong selection against its maintenance in the population is assumed because it is lethal when homozygous (except in 3 cases, in which it causes male sterility). These *t* alleles, however, are also subject to a strong abnormality in the process of gamete production and maturation. Under normal conditions, 50% of the gametes of a heterozygote will contain one allele, and 50% the other. The heterozygote containing the *t*-allele, however, yields an abnormal ratio of 95:5 of *t* to normal gametes—rather than the expected 50:50—now known to result from differential mortality of the gametes as they mature (see Bennett 1975).

The problem for the biologist is to explain how the polymorphism is maintained in the population.

In general, the presence of the polymorphism is accounted for by a balance between the two forces cited above: the selection against the mutant *t* allele in the homozygote reduces the number of such alleles, while the stock of *t* alleles is constantly increased by the abnormal gamete ratios in the heterozygous males. Heterosis, i.e. superior fitness of the heterozygote, might also serve as a balancing force, but this force is omitted from these models because of lack of data (Lewontin and Dunn 1960, p. 707).

Both deterministic and stochastic models can be used to represent the key features of this qualitative account. The choice, in this case, turns on assumptions about the value of the parameter for population size.

If the breeding groups (i.e. effective population size) are assumed to be small, then chance processes, such as random drift, add a statistical element to the situation, necessitating the use of a stochastic model. That is, Wright showed that if you have a finite population size, the rates of changes in gene frequencies will depend, among other things, on random processes involving mutation rates, migration rates, selection, and "accidents of sampling". One particular result is that it is easier to reach fixation (of an allele) in small populations, in which genes are lost or fixed at random, with little reference to selection pressure (Dobzhansky 1970, pp. 230, 232–234). With small populations, then, the presence of the polymorphism is *not* understood to be purely a function of the interaction of the two forces discussed above, so a deterministic model is not appropriate (Lewontin and Dunn 1960, p. 707).

If, on the other hand, the population size is assumed to be effectively infinite, then the random effects resulting from small population size are absent, and the state of the polymorphism in the population is solely a function of the selection and abnormal segregation values (although infinite population size is not necessarily required by deterministic modelling).

The deterministic model, chosen first to account for the frequencies of this polymorphism (from Bruck 1957), uses two parameters: the proportion of mutant t gametes in the effective sperm pool, m; and p, the frequency of the non-mutant allele. The result of this model is a *single value* for the frequency of adults heterozygous for the t allele. This result was found not to correspond with the result in nature (Lewontin and Dunn 1960, p. 708).

In addition to the empirical inadequacy of the deterministic model, the biologists had theoretical reasons to believe that a stochastic model would be more appropriate for this phenomenon. That is, they noted that the effective size of a breeding unit is small; the species population as a whole consists of a number of partly separated, relatively small, breeding groups. Thus, Lewontin and Dunn decided that "a useful approach in the construction of models is to test the effects on gene frequencies of small effective size of the breeding unit" (1960, p. 708).

Lewontin and Dunn analyzed the stochastic model of the processes of the interaction of selection, segregation abnormality, and restricted population size by *simulation*. The simulation is done by making rules for the evolution of simulated populations that "conform with genetic rules of meiosis, fertilization, and selection" (1960, p. 708). Random elements are also included in the models, since chance is involved in the survival

and reproduction of any particular individual (selection), and in which gametes are chosen from the gamete pool. Randomization of the union of sperm and egg yields different frequencies on each run of the simulation (done on computer, in most cases). The idea is to collect a number of these different frequency results and get a *distribution* of the results over a number of runs (Lewontin 1962, p. 67). The parameters fixed for each run include N (effective population size), the fitnesses of the various genotypes, and m (the factor of segregation distortion). Each run is started with an exact description of the initial population (Lewontin and Dunn 1960, pp. 708–710).

Large numbers of runs are made with identical parameter sets; no two of these runs will have the same results, because of random factors. Distributions, means, and variances can be calculated from the gene frequency results obtained from all the models with a given parameter set. Lewontin and Dunn's statistical analysis of their simulated results led them to conclude that the effects of changing the population size are statistically significant, i.e. use of a smaller value for the population size parameter results in genetic drift. The actual distributions obtained by Lewontin and Dunn from the simulation of the stochastic model conform with the predictions made by Wright's mathematical model (1960, p. 712). They conclude that for small populations, the mean values from the stochastic model do not correspond with the prediction from the deterministic model, because the latter model does not account for the chance loss of alleles in small breeding groups (1960, p. 719).

Thus, with the application of a stochastic model to small breeding groups, it is possible to produce simulation results that fit the actual results better than the deterministic model. Information is also gained regarding the exact inadequacies of the deterministic model for the particular phenomenon being modelled. In this case, the assumption of infinitely large effective population size, N, led to inadequacy of the model containing that assumption.

3. The Structure of Population Genetics Theory. Having presented a few particular examples of population genetics models which highlight the presence and utility of certain facets of model description, I would like to discuss details of the description of the theory according to the semantic view. Formalization of any theory T, according to the semantic view, involves defining the class of models of T. The theory is conceived as defining a kind of ideal system. The main items needed for this description are the definition of a state space, state variables, parameters, and a set of laws of succession and coexistence for the system (see Section 1.1). Section 3.1 discusses the most common state space for the representation of genetical phenomena of populations, and its theoretical

disadvantages. Choice and evaluation of parameters, and the relations between parameters and the structures are discussed in Section 3.2. Section 3.3 contains some general comments regarding the laws or rules of the models. Finally, I discuss very briefly the interrelationship among the models.

3.1 State Spaces. Choosing a state space (and thereby, a set of state variables) for the representation of genetic states and changes in a population is an important part of population genetics theory. As Lewontin notes

> The problem of constructing an evolutionary theory is the problem of constructing a state space that will be dynamically sufficient, and a set of laws of transformation [i.e. laws of succession] in that state space that will transform all the state variables (1974, p. 8).

Paul Thompson has suggested that the state space for population genetics would include the physically possible states of populations in terms of genotype frequencies. The state space would be "a Cartesian-space where '*n*' is a function of the number of possible pairs of alleles in the population" (Thompson 1983, p. 223). We can picture this geometrically as *n* axes, the values of which are frequencies of the genotype pair. The state variables are the frequencies for each genotype. Note that this is a one-locus system. That is, we take only a single gene locus, and determine the dimensionality of the model as a function of the number of alleles at that single locus.

Another type of single-locus system, used less commonly than the one described by Thompson, involves using single gene frequencies, rather than genotype frequencies, as state variables. Debates about "genic selectionism" center around the adequacy of this state space for representing evolutionary phenomena (see Sober and Lewontin 1982 for discussion of this issue). With both genotype and gene frequency state spaces, though, treating the genetic system of an organism as being able to be isolated (meaningfully) into single loci involves a number of assumptions about the system as a whole. For instance, if the relative fitnesses of the genotypes at a locus are dependent upon *other* loci, then the frequencies of a single locus observed in isolation will *not* be sufficient to determine the actual genotype frequencies. Assumptions about the structure of the system as a whole can thus be incorporated into the state space in order to reduce its dimensionality. Lewontin (1974) offers a detailed analysis of the quantitative effects on dimensionality of various assumptions about the biological system being modelled. It is made clear in his discussion that, although a state space incorporating the most realistic assumptions is desirable from a descriptive point of view, it is mathematically and

theoretically intractable. For instance, a total genetic description (with no implicit assumptions) of a population with only two alleles at two loci would have a dimensionality of nine, while three alleles at three loci would be described in a 336-dimensional space (1974, p. 283). Most organisms have thousands of loci; the one-locus system is much more manageable, e.g. for the formulation of laws of succession for the system.

A number of objections to the single-locus system have been raised by biologists. These objections, reviewed below, can be understood in terms of the descriptive inadequacy of the dimensions of the state space.

Michael Wade, in his discussion of group selection models, objects to the use of the single-locus model in calculations of the strength of group selection vs. individual selection. Since some of the processes important to the operation of group selection (e.g. genotype-genotype interaction and interactions between loci) *cannot* be represented by a single-locus model, results of comparison of the forces of individual selection vs. group selection within the context of such models is inevitably skewed (Wade 1978, pp. 103–104).

Interactions between genotypes, and between one locus and another, cannot (except in the case of frequency dependent selection) be represented in a single-locus model, for the simple reason that they involve more than one locus. The trajectory of the frequency of a gene involved in these processes in a single-locus model will not follow a law-like pattern, and will be thus inexplicable. Lewontin offers an example involving two polymorphic inversion systems whose frequencies are dependent on one another. The actual frequencies are inexplicable in a one locus model, which does not allow for the interaction of the two polymorphisms in their determination of fitness. Models of higher dimensionality (or using different state variables) are necessary, because of the "dimensional insufficiency" of the single-locus models (1974, pp. 273–281; since this example has been discussed at length by Wimsatt 1980, pp. 226–229, I shall not go into detail here).

At this point, I would like to introduce an additional category. Although all single-locus models should, in some sense, be grouped together, they are not all exactly the same model—each particular model has a different number of state variables, depending upon the number of alleles at that locus. Van Fraassen has suggested calling the general outline for each model its "model type" (1980, p. 44). Since a model type is simply an abstraction of a model, constructed by abstracting one or more of the model's parameters, a single model can be an instance of more than one model type; the model types themselves are therefore not hierarchically arranged. Along similar lines, I suggest that each model type be associated with a distinctive *state space type*. In the preceding

example, the single-locus model is to be taken as an instance of a general state space type for all single-locus models, i.e. the different single-locus model types are conceived as using the same state space type. Alternatives, such as two-locus models (see Lewontin 1971 for an example), must be taken as instances of a different state space type.

Lewontin, dissatisfied with the theoretical results afforded by the use of single-locus and even multi-locus state space types, has suggested an entirely different state space type. The intention is to treat the entire genome as a whole, rather than as a collection of independently segregating, non-interacting genotypes of single loci. Ernst Mayr has stressed the importance of the interaction of genes and the homeostasis of genotypes (i.e. large amount of linkage) in evolutionary processes. The genome will respond to selection pressures *as a whole* says Mayr, instead of as an aggregate of individual loci (Mayr 1967, p. 53). In our terms, since evolution works this way, any accurate model of evolution cannot employ the single-locus state space type. Following up on his claim that the construction of a dynamically sufficient theory of a genome with many genes is "the most pressing problem of theory", Lewontin suggests an alternative approach using a completely different set of state variables (1974, p. 318).

According to the semantic view, a description of a theory's structure involves the description of the family of models for the theory. An essential part of this description of the family of models consists in describing the specific types of state spaces in terms of which the models are given. In this section, I have presented a general sketch of the types of state spaces associated with various model types, i.e. a description of the class of state space types.

3.2 Parameters. Values which appear in the succession and coexistence laws of a system that are the same for all possible states of the defined system are here called *parameters*. For instance, in the modification of the Hardy-Weinberg equation which predicts the frequencies of the genotypes after selection, the selection coefficient, s, appears as a parameter in the equation:

$$p^2AA + 2pqAA + (1 - s)q^2aa = 1 - sq^2$$

There are a variety of methods of establishing the value at which a parameter should be fixed or set in the construction of models for a given real system. Simulation techniques, like those presented in Section 2.3, can be used to obtain estimates of biologically important parameters. In some contexts, maximum likelihood estimations may be possible. Parameters can also be set arbitrarily, or ignored. This is equivalent to incorporating certain assumptions into the model for purposes of simplification

(see Section 3.1 on state space assumptions). (See also Levins 1968, pp. 8, 89; Bailey 1966, pp. 42, 220; Suppes 1967, pp. 62–63.)

Parameters can play roles of varying importance in the determination of the system represented by the theory. In this section, I shall discuss cases of the differing effect of the *values* of the parameters on the model outcome. The *choice* of parameters itself can also be theoretically important, as seen in the group selection example below.

One expects the values of parameters to have impact on the system being represented; but variations in parameter values can make a larger or smaller amount of difference to the system. For instance, take the deterministic model which incorporated a parameter for mutation, μ (see Section 2.1). The outcomes of this model are virtually insensitive to variations in the value of μ. Yet the selection parameter plays a crucial role in this same model; a very small amount of selection in favor of an allele will have a cumulative effect strong enough to replace other alleles (Lewontin 1974, p. 267).

Population size is another case in which the value assigned to the parameter has a large impact on the model results. As the case of polymorphism in the house mouse shows (discussed at length in Section 2.3), effective population size, N, can play a crucial role in some models, since selection results can be quite different with a restricted gene pool size (see Mayr 1967, pp. 48–50). In many of the stochastic models involved in calculating rates of evolutionary change, the resulting distributions and their moments can depend completely on the ratios of the mean deterministic force to the variance arising from random processes (Lewontin 1974, p. 268). This variance is usually proportional to $1/N$, and is related to the finiteness of population size. Thus, change in the value of the single parameter, N, can completely alter the structures represented by the theory.

The choice of parameters can also make a major difference to the model outcome. Theoreticians have choices about how to express certain aspects of the system or environment. The choice of parameters used to represent the various aspects can have a profound effect on the structure, even to the point of rendering the model useless for representing the empirical system in question. Group selection models provide a case in which choice of parameters not only altered the results of the models, but also led to the near disappearance (in the models) of the phenomenon being modelled, according to Wade (see 1978).

Wade analyzed the major group selection models and found that they contain a number of common assumptions about various ecological processes, including extinction rates, migration, dispersion, and colonization (1978, pp. 103, 112). He challenged the accuracy of several of these assumptions on grounds that they are not biologically realistic enough.

Take the assumption about colonization in the models—there are different *modes* of colonization, and the presence of these different modes has different effects on a number of factors affecting the existence and strength of group selection (1978, p. 103). Wade claims that the assumptions about colonization incorporated into existing models limit, automatically, several mechanisms for creating and maintaining genetic variation between populations (1979, p. 105; 1977, p. 150).

Since variation of group traits between groups is the analog in group level models to genetic variation in the genotype level models, and the operation of selection is dependent upon variation among the units of selection, the initial assumptions about colonization can have large effects on the selection results of the system modelled. Existing assumptions regarding colonization are not representative, Wade claims, and he challenges them empirically. He finally suggests an alternate model incorporating new assumptions about colonization and new values for the colonization parameters suggested by his empirical research (1978, pp. 109–110; 1979; 1976; see Wimsatt's discussion, 1980, pp. 238–248).

Some authors, when discussing genetical changes in populations, speak of the system in terms of a phenotype state space type (Eden 1967, p. 10; Lewontin 1974, pp. 9–13). This makes sense, since the phenotype determines the breeding system and the action of natural selection, the results of which are reflected in *some* way, in the genetic changes in the population. In his analysis of the present structure of population genetics theory, Lewontin traces a single calculation of a change in genetic state through both genotypic and phenotypic descriptions of the population. That is, according to Lewontin, population genetics theory must map the set of genotypes onto the set of phenotypes, give transformations in the phenotype space, and then map the set of phenotypes back onto the set of genotypes. We would expect, then, that descriptions of state in population genetics would be framed in terms of *both* genetic and phenotypic variables and parameters. But this is not the case—the description can be in terms of *either* genotype *or* phenotype variables, but not both. Dynamically, then, it seems as if population genetics must operate in two parallel systems: one in genotype state space; one in phenotype state space (Lewontin 1974, pp. 12–13).

Lewontin explains that such independence of systems is illusory, "and arises from a bit of sleight-of-hand in which phenotype and genotype variables are made to appear as merely parameters that need to be experimentally determined, constants that are not themselves transformed by the evolutionary process" (1974, p. 15). A prime example of such a "pseudo parameter" is the fitness value associated with the individual genotypes while computing the mean fitness value, $\bar{w}$. The mean fitness value appears in the equation which expresses the relative change in allele

frequency, Δq, of an allele at a locus after one generation, in terms of the present allele frequency, q, and the mean fitness, $\bar{w}$, of the genotypes in the population:

$$\Delta q = \frac{q(1-q)}{2} \frac{d \ln \bar{w}}{dq}$$

(Lewontin 1974, p. 13). Although $\bar{w}$ is used in computation in a genotype state space type, fitness is a function of phenotype, not genotype. Thus information regarding values of phenotype variables is smuggled into the genotype models through parameters.[4]

3.3 Laws. In line with my goal of providing a general approach for describing the models of population genetics (since the theory is being described in terms of a family of models), I discuss in this section a few particular aspects and forms of the laws used in these models. The most obvious differences, in laws as well as in state space types, are between deterministic and stochastic models. But, as discussed below, even laws having the common framework of the Hardy-Weinberg equilibrium can differ fundamentally.

Coexistence laws describe the possible states of the system in terms of the state space. In the case of evolutionary biology, these laws would consist of conditions delineating a subset of the state space which contains only the biologically possible states. Changes in the state of the system are described by laws of succession. In the case of evolutionary theory, dynamical laws concern changes in the genetic composition of populations[5] (see Lewontin 1974, pp. 6–19).

The laws of succession select the biologically possible trajectories in the state space, with states at particular times being represented by points in the state space (this is simplified—see discussion on time variables below). The law of succession is the equation of which the biologically

[4]A referee for this journal points out that the situation is even more complicated than this passage suggests. That is, this phenotype information amounts to "the average effect of the phenotypes in fact produced by the relevant genotype in the present generation"; but changes elsewhere in the genome or any other changes in the genetic environment may yield a different 'average' phenotype for the same genotype. Furthermore, the same average phenotype could yield very different fitnesses in slightly different environments.

[5]The concept of a system changing over time, where the system is usually interpreted as a single population or species, is peculiar. David Hull has suggested that a more appropriate interpretation of such systems would be as *lineages,* which have the desirable qualities of being necessarily spatiotemporally localized and continuous (personal communication). Note, however, that such interpretive problems are a separate issue from description of the models, which simply represent ideal systems. Clearer understanding of the ideal systems and their interrelations should shed light on the advantages and disadvantages of the various possible empirical interpretations of the systems (e.g. see Hull 1980).

possible trajectories are the solutions (van Fraassen 1970, pp. 330–331).

The Hardy-Weinberg equation, of which several variations were presented in Section 2, is the fundamental law of both coexistence and succession in population genetics theory. As Lewontin has noted, even the dynamical laws of the theory appeal to only the equilibrium states and steady-state distributions, which are estimated from the Hardy-Weinberg equation or variations thereof (1974, p. 269). The Hardy-Weinberg law is a very simple, deterministic succession law which is used in a very simple state space. As parameters are added to the equation, we get *different* laws, technically speaking. For example, compare the laws used to calculate the frequency p' of the A allele in the next generation. Including only the selection coefficient into the basic Hardy-Weinberg law, we get $p' = p/(1 - sq^2)$. Addition of a parameter for mutation rate yields a completely different law, $p' = p - ps + \mu q$. We could consider these laws to be of a single type—variations on the basic Hardy-Weinberg law—which are usually used in a certain state space type. The actual state space used in each instance depends on the genetic characteristics of the system, and not usually on the parameters. For instance, the succession of a system at Hardy-Weinberg equilibrium and one which is *not* at equilibrium but is under selection pressure, could both be modelled in the same state space, using different laws.

In the discussion in Section 2 involving equilibrium and dynamical models employing the Hardy-Weinberg equilibrium, the distinction between stochastic and deterministic models loomed large. Examination of the general features of the deterministic and statistical laws which appear in these models should help clarify the structure of the theory itself.

A theory can have either deterministic or statistical laws for its state transitions. Furthermore, the states themselves can be either statistical or non-statistical. In population genetics models, gene frequencies often appear in the set of state variables; thus the states themselves are statistical entities.

In general, according to the semantic view, a law is deterministic if, when all of the parameters and variables are specified, the succeeding states are uniquely determined (this definition of determinism, and its advantages over other definitions are discussed in detail by van Fraassen, 1972, pp. 306–321). In population genetics, this means that the initial population and parameters are all that is needed to get an exact prediction of the new population state (Lewontin 1967, p. 87).

Statistical laws are constructed by specifying a probability measure on the state space. The example presented in Section 2.2 entailed assigning probabilities (frequencies) to each distinct possible value of gene frequency. Thus, the probability measure is constructed by taking a certain value for the gene frequency, obtaining the joint distribution (in this case,

through simulation), and making a new state space of probabilities on the old state space of gene frequencies.

Sometimes it is possible, depending on the variables and parameters in the laws, to translate a stochastic law on determinate states into a deterministic law on statistical states (van Fraassen 1970, pp. 333–334). In the case of population genetics models containing statistical states, this particular translation may not be possible, and the laws might remain statistical laws on statistical state variables. Consider, for example, the case of the polymorphism discussed in Section 2.3. The stochastic model actually contained more *relevant* information, (i.e. about population size) but less information in general, since it did not yield determinate values. Stochastic and deterministic models can thus contain more or less information, depending on the question being asked, and the aspects of the system or environment being included.

In the last part of this section, I would like to discuss briefly a related problem regarding the flexibility available in representing a given system.

In the representation of a system, a state can be conceived as a function of time, or not. That is, the state vector itself can be a function of time; the state is represented as a point, while the history of the system can be represented as a curve. On an alternative approach, the operator representing the magnitude can be a function of time; the history of the system would be represented as point in this state space, the different points representing different "possible worlds" or world histories (van Fraassen 1970, p. 329–35).

Lewontin is interested in the biological usefulness of each of these possible ways to represent systems. He claims that although the usual mode of presentation is done (in our terms) by employing an instantaneous state space, the information presented thereby is not very interesting to the biologist. A description of the "time ensemble of states of a given population" would be much more useful, he claims (1967, p. 82). We might interpret this as a claim that a "possible worlds" representation would represent the information in a more useful way. But Lewontin seems to be saying more than this.

The case he is considering involves the following problem: In one case, the gene frequency, Q, of a certain allele is calculated using a series of randomly fluctuating, uniformly distributed values of the selection coefficient. In the other case, the same procedure is performed using the exact same set of selection coefficient values, except in reverse temporal order. The resulting values of Q are *different* for the two cases. In other words, in general, if the curves representing the paths of the selection coefficients of each population through time are not identical, *even though* they have the same mean, variance, and any other statistical measurement, the model outcomes will *not* necessarily be identical, because of the difference in

temporal order of the values (Lewontin 1967, p. 84). Thus, if a possible worlds representation were possible, it would seem to contain more information about the system, since the time histories are preserved in a certain sense. If this is so, then there would probably be problems translating between the two possible types of system, i.e. possible worlds and instantaneous state space (analogous to Heisenberg and Schrödinger pictures, respectively, in quantum mechanics). Are biological systems different from physical systems in that the descriptions of the systems, conceived as both a function of time and independent of time, are *not* both represented as two aspects of the same system in a Cartesian space? Lewontin explicitly claims that the gene frequencies of populations do not follow the law of large numbers (1967, p. 84). In any case, this poses an intriguing problem for future foundational research.

3.4 Interrelation of Models. The issue of the exact interrelations among the different model types of population genetics is the topic of another entire paper. Here I wish to make a few preliminary remarks.

According to the semantic view, the structure of a theory can be understood by examining the family of models it presents. In the case of population genetics theory, the set of model types—stochastic and deterministic, single-locus or multi-locus—can be understood as a related family of models. The question then becomes defining the exact nature of the relationships among them.

One rather nice example of a detailed analysis of a relation among models was discussed in Section 3.2. There, parameters of genotype fitness were found to be versions of information about phenotypes, condensed into genetic form. The model types constructed on phenotype and genotype state space types can thus be understood as overlapping through the specific parameter of fitness.

It can also be useful to examine models of the same phenomenon which have different degrees of complexity. Some loss of information occurs in all models when the parameters are set. By fixing the value of or ignoring a factor which is known to be important in some contexts, assumptions are made which simplify the model.

Sometimes the incorporation of simplifying assumptions reduces the usefulness of the model. In the example presented in Section 2.3 the assumption, present in the deterministic model, that the effective population size had no bearing on the outcome of the model, turned out to render the model inferior to a model which omitted such an assumption. Lewontin and Dunn concluded that the latter model "more nearly explains what is observed in nature" since it is "closer to the real situation" (1960, p. 707).

It might seem that the inclusion in a model or set of models of as-

sumptions that obviously do not correspond to the observed phenomena would necessarily detract from the usefulness or accuracy of the models involved. But the biologist Richard Levins has suggested a method of eliminating detrimental effects of arbitrary assumptions on a theory as a whole. He recommends replacing the unrealistic assumptions in a given model or model type with other (perhaps equally unrealistic) assumptions. A theorem which is supported by means of different models "having in common the aspects of reality under study but differing in other details, is called a robust theorem" (1968, p. 7). The actual operation and usefulness of such theorem testing is in the realm of theory confirmation, and will not be discussed here. The important point is that, through construction and comparison of alternative models and model types of a given phenomenon for purposes of confirmation, the interrelation among the various model types and parameters is made explicit.

4. Conclusion. In this paper, I have taken the semantic view of theory structure as a general framework for foundational studies of population genetics. A brief review of the semantic approach provided a basis on which to discuss the theory. I examined specific problem areas in this particular program of foundational studies, using examples of population genetics models presented in Section 2.

According to the semantic view, to present a theory is to present a related family of models. In the simplest case, these models are related by having a common state space and common law of succession. Differences between the models lie in different initial conditions (i.e. initial location of the system in the state space), and hence different successions of states satisfying the same law. Generally, however, the theory allows for greater essential variety in the system it deals with (for instance, different degrees of freedom). Representation of these systems then requires models with different state spaces (for instance, of different dimensionality) and different laws.

In Section 2, I reviewed a variety of models employed in population genetics. Using these as a starting point, in Section 3 I approached the task of describing the models of population genetics in terms of the classification of state space types (such as single-locus state spaces), and law types (such as the general form of the Hardy-Weinberg laws). In general, instances of a type differ by having different values for parameters, while different types result from the choice of parameters. Further research is indicated, especially on the relations between state spaces (such as in genotype and phenotype modelling) and the relations between different sorts of laws of succession (such as in deterministic and stochastic modelling).

REFERENCES

Bailey, N. T. J. (1967), *The Mathematical Approach to Biology and Medicine*. London, New York: Wiley.

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

Beatty, J. (1981), "What's Wrong with the Received View of Evolutionary Theory?", *PSA 1980*: Vol. Two. East Lansing, Michigan: Philosophy of Science Association.

Beatty, J. (1982), "The Insights and Oversights of Molecular Genetics: the Place of the Evolutionary Perspective", *PSA 1982*: Vol. One. East Lansing, Michigan: Philosophy of Science Association.

Bennett, D. (1975), "The T-locus of the Mouse", *Cell 6*: 441–454.

Bruck, D. (1957), "Male segregation ratio advantage as a factor in maintaining lethal alleles in wild populations of house mice", *Proceedings of the National Academy of Sciences USA 43*: 152–158.

Dobzhansky, T. (1970), *Genetics of the Evolutionary Process*. New York: Columbia University Press.

Kimura, M. and Ohta, T. (1971), *Theoretical Aspects of Population Genetics*. Princeton: Princeton University Press.

Hull, D. L. (1980), "Individuality and Selection", *Annual Review of Ecology and Systematics 11*: 311–332.

Lewontin, R. C. and Dunn, L. C. (1960), "The Evolutionary dynamics of a polymorphism in the house mouse", *Genetics 45*: 701–722.

Lewontin, R. C. (1962), "Interdeme Selection controlling a polymorphism in the House Mouse", *The American Naturalist* Vol. 46 No. 887: 65–78.

Lewontin, R. C. (1967), "The Principle of Historicity in Evolution", in *Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution*, P. S. Moorehead and M. M. Kaplan (eds.). Philadelphia: Wistar Institute Press, pp. 81–88.

Lewontin, R. C. (1974), *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

Levins, R. (1968), *Evolution in Changing Environments*. Princeton: Princeton University Press.

Maynard Smith, J. (1968), *Mathematical Ideas in Biology*. Cambridge, England: Cambridge University Press.

Mayr, E. (1967), "Evolutionary Challenges to the Mathematical Interpretation of Evolution", in *Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution*, P. S. Moorehead and M. M. Kaplan (eds.). Philadelphia: Wistar Institute Press, pp. 47–54.

Moulines, C. U. (1975), "A Logical Reconstruction of Simple Equilibrium Thermodynamics", *Erkenntnis 9*: 101–130.

Sneed, J. (1971), *The Logical Structure of Mathematical Physics*. Dordrecht: Reidel.

Sober, E. and Lewontin, R. C. (1982), "Artifact, Cause, and Genic Selection", *Philosophy of Science 49*: 157–180.

Starfield, A. M. et al. (1980), "An Exploratory Model of Impala Population Dynamics", in *Mathematical Modelling in Biology and Ecology*, Vol. 33 of *Lecture Notes in Biomathematics*, S. Levin (ed.). Berlin, New York: Springer-Verlag.

Stegmuller, W. (1976), *The Structure and Dynamics of Theories*. New York: Springer-Verlag.

Suppe, F. (1972), "What's Wrong with the Received View on the Structure of Scientific Theories?" *Philosophy of Science 39*: 1–19.

Suppe, F. (1973), "Theories, Their Formulations, and the Operational Imperative", *Synthese 25*: 129–164.

Suppe, F. (1974), "Theories and Phenomena", in *Developments of the Methodology of Social Science*, W. Leinfellner and E. Kohler (eds.). Dordrecht: Reidel, pp. 45–92.

Suppe, F. (1974a), "Some Philosophical Problems in Biological Speciation and Taxonomy", in *Conceptual Basis of the Classification of Knowledge*, Wojcieckowski (ed.). Munich: Verlag Dokumentation, pp. 190–243.

Suppe, F. (1976), "Theoretical Laws", in *Formal Methods in the Methodology of Empirical Sciences,* M. Przelecki, et al. (eds.). Dordrecht, Boston: Reidel, pp. 247–267.
Suppe, F. (ed.) (1977), *The Structure of Scientific Theories* (2nd ed.). Urbana, Illinois: University of Illinois Press.
Suppe, F. (1979), "Theory Structure", in *Current Research in Philosophy of Science,* East Lansing, Michigan: Philosophy of Science Association, pp. 317–338.
Suppes, P. (1957), *Introduction to Logic.* New Jersey: D. Van Nostrand and Co.
Suppes, P. (1961), "A Comparison of the Meaning and Use of Models in Mathematics and the Empirical Sciences", in *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences,* H. Freudenthal (ed.). Dordrecht: Reidel, pp. 163–177.
Suppes, P. (1962), "Models of Data", in *Logic, Methodology, and the Philosophy of Science. (Proceedings of the 1960 International Congress,* Vol. 1), E. Nagel, P. Suppes, and A. Tarski (eds.). California: Stanford University Press, pp. 252–261.
Suppes, P. (1967), "What is a Scientific Theory?" in *Philosophy of Science Today,* S. Morgenbesser (ed.). New York: Meridian Books.
Thompson, P. (1983), "The Structure of Evolutionary Theory: a Semantic Approach", *Studies in History and Philosophy of Science 14*: 215–229.
van Fraassen, B. C. (1969), "Meaning Relations and Modalities", *Nous 3*: 155–167.
van Fraassen, B. C. (1970), "On the Extension of Beth's Semantics of Physical Theories", *Philosophy of Science 37*: 325–339.
van Fraassen, B. C. (1972), "A Formal Approach to the Philosophy of Science", in *Paradigms and Paradoxes,* R. Colodny (ed.). Pittsburgh: University of Pittsburgh Press.
van Fraassen, B. C. (1974), "The Labyrinth of Quantum Logic", in *Logical and Empirical Studies in Contemporary Physics,* R. S. Cohen and M. Wartofsky (eds.). Boston Studies in the Philosophy of Science, Vol XIII.
van Fraassen, B. C. (1980), *The Scientific Image.* Oxford: Clarendon Press.
van Fraassen, B. C. (forthcoming), "Aim and Structure of Scientific Theories", in *Proceedings of the Seventh International Congress of Logic, Methodology, and Philosophy of Science.*
Wade, M. J. (1976), "Group Selection among Laboratory populations of *Tribolium*", *Proceedings of the National Academy of Sciences USA* Vol. 73, No. 12: 4604–4607.
Wade, M. J. (1977), "An Experimental Study of Group Selection", *Evolution 31*: 134–153.
Wade, M. J. (1978), "A Critical Review of the Models of Group Selection", *Quarterly Review of Biology 53*: 101–114.
Wessels, L. (1976), "Laws and Meaning Postulates (in van Fraassen's View of Theories)", in *PSA 1974,* R. S. Cohen et al. (eds.). Dordrecht: Reidel, pp. 215–235.
Wimsatt, W. C. (1980), "Reductionist Research Strategies and their biases in the units of selection controversy", in *Scientific Discovery: Case Studies,* T. Nickels (ed.). Dordrecht: Reidel, pp. 213–259.