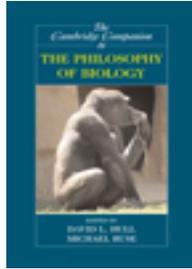# Cambridge Companions Online

The Cambridge Companion to the Philosophy of Biology

Edited by David L. Hull, Michael Ruse

Chapter

4 - What's Wrong with the Emergentist Statistical Interpretation of Natural Selection and Random Drift? pp. 66-84

Cambridge University Press

# 4    What's Wrong with the Emergentist Statistical Interpretation of Natural Selection and Random Drift?

Population-level theories of evolution – the stock and trade of population genetics – are statistical theories par excellence. But what accounts for the statistical character of population-level phenomena? One view is that the population-level statistics are a product of, are generated by, probabilities that attach to the individuals in the population. On this conception, population-level phenomena are explained by individual-level probabilities and their population-level combinations. Another view, which arguably goes back to Fisher (1930) but has been defended recently,[1] is that the population-level statistics are sui generis, that they somehow emerge from the underlying deterministic behavior of the individuals composing the population. Walsh, Lewens, and Ariew (2002) label this the *statistical interpretation*. We are not willing to give them that term, since everyone will admit that the population-level theories of evolution are statistical, so we will call this the *emergentist statistical interpretation* (ESI). Our goals are to show that (1) this interpretation is based on gross factual errors concerning the practice of evolutionary biology, concerning both what is done and what can be done; (2) its adoption would entail giving up on most of the explanatory and predictive (i.e., scientific) projects of evolutionary biology; and finally (3) a rival interpretation, which we will label the *propensity statistical interpretation* (PSI), succeeds exactly where the emergentist interpretation fails.

66

## 1. PROPENSITY AND EMERGENTIST INTERPRETATIONS

The propensity interpretation of fitness was introduced into the philosophical literature in 1978 (see Brandon 1978; also see Mills and Beatty 1979). The prime motivation was to make room for an explanatory theory of natural selection, which is tantamount to solving the so-called tautology problem. This problem arises from a casual inspection of the phrase "survival of the fittest," followed by the question of what defines the fittest. If the answer is those that reproduce the most, then it seems we are explaining a phenomenon, differential reproduction, in terms of itself, which is no explanation at all.

Brandon's approach was to think of fitness (or adaptedness) as a disposition. Just as it is not explanatorily empty to cite the water solubility of salt in explaining the behavior of a particular sample of salt when placed in water, so too it is not explanatorily empty to cite differences in adaptedness to a common environment when explaining a particular case of differential reproduction. Of course, in the case of water solubility we want, and indeed have, a deeper explanation of that disposition – a general explanation given in terms of molecular bonding.

The case of fitness differs in two important ways from that of water solubility. First, ceteris paribus, water-soluble substances dissolve when placed in water, period. That is, although we qualify the claim with a ceteris paribus clause – we want to exclude cases such as that when the water is frozen, or already saturated, and so on – the claim itself is not probabilistic. But we think chance can intervene in real biological populations so that higher fitness and higher levels of reproductive success can be dissociated. On the propensity interpretation, fitness (or adaptedness) is an explicitly probabilistic concept.[2] Thus it is a probabilistic propensity. Second, unlike in the case of water solubility, there is no general underlying explanation of differential fitness (see Brandon 1978 or 1990, 13–25; also see Rosenberg 1978, 1985; and Sober 1984). The underlying causal basis of fitness differences can be uncovered by detailed study of particular populations in particular selective environments, but it will not be general. It is our impression that this interpretation of fitness is widely accepted in both the philosophical and biological

communities. But some of its broader implications are probably not appreciated.

If the propensity interpretation is correct, then population-level probabilities are derivable from individual-level probabilities in a familiar way. For example, if a coin and tossing device yields a probability of heads of .5, then we can calculate the probability of various results in an ensemble of tosses, say four, by the laws of probability theory. For those of you who have trouble thinking of coin tossing as genuinely stochastic, substitute the following example. Oxygen-15 has a half-life of two minutes. Take four atoms of that isotope. What is the probability that exactly two of them will decay during a two-minute time interval? To answer that question we do the exact same calculation as in the coin example and get the same answer. The probability of that outcome is .375. So the propensity interpretation of fitness yields a familiar and natural way of understanding the population-level probabilities that are essential to evolutionary theory. The only sticking point here is that, taken literally, the propensity interpretation of fitness is committed to the fundamental indeterminacy of the lives, deaths, and ultimately reproductive successes of individual organisms (see Brandon and Carson 1996). Some people find that a difficult ontological commitment.

So perhaps one motivation for the emergentist statistical interpretation is that it is not committed to the indeterminacy of individual lives and deaths. Indeed it seems that the inspiration for this interpretation is the relation between statistical thermodynamics and Newtonian mechanics. Here, *supposedly*, the underlying mechanics of the molecules in a gas are deterministic, but at the macrolevel we get the explicitly probabilistic second law of thermodynamics.[3] Perhaps then there is an analogue of the emergentist statistical interpretation in physics. But we do not think one should be much impressed by that, since no one understands the relationship between mechanics and statistical thermodynamics (Sklar 1999). Looking to physics will not help us understand the emergence of population-level probabilities.

Another possible motivation for the emergentist interpretation is that we are stuck with it. That is, we can clearly see that there are population-level probabilities governing the evolutionary trajectories of populations, but we have *no access* to the individual-level

probabilities postulated by the propensity interpretation – either because they do not exist, or because we have epistemic limitations. We will deal with this possibility in the next section.

Finally, one might find the emergentist interpretation attractive primarily because one thinks that evolutionary theory deals solely with population-level probabilities, and therefore has no need for individual-level probabilities.[4] So ontological parsimony suggests we do without them.[5] This point will be dealt with in Section 3.

## 2. THE ARGUMENT FOR THE EMERGENTIST STATISTICAL INTERPRETATION

Basing fitness on type- or population-level effects has precedent, especially among biologists. Fisher (1930) took the fitness of a type to be the objective representation of that type in the next generation. Similarly, a standard evolutionary biology textbook (Futuyma 1986) defines fitness as ''the average contribution of one allele or genotype to the next generation or succeeding generations, compared with that of other alleles or genotypes.''

Despite drawing support from a number of biologists, until recently this position has received little support from philosophers. In our discussion of the ESI, we will focus on recent defenses of this view by Walsh et al. (2002) and Matthen and Ariew (2002). Walsh et al. ask whether evolutionary theory is a statistical theory or a dynamical theory. A statistical theory is phenomenological, not causal, and a dynamical theory is a theory of forces, à la Newtonian mechanics. Thus their question is more or less equivalent to this: is evolutionary theory like the kinetic theory of gases or Newtonian mechanics? This seems an impoverished range of options. Why should it be relevantly similar to either one? Are those the only two types of scientific theories? Although they present us with a false choice the logic of their argument is clear: Sober's (1984) description of evolutionary theory as a theory of forces is, they claim, wrong; therefore the emergentist statistical interpretation is correct.[6]

Sober's description of evolutionary theory as a theory of forces has flaws, some more serious than others. As Endler (1986) has pointed out, there are a number of disanalogies between natural selection and the concept of force in physics.[7] But this is a quibble compared to the most important problem with Sober's analogy, which is the

fact that selection and drift are not opposing forces, but rather two copossible outcomes of the same process – the process of sampling from a population where the probabilities of being sampled for each member of the population do not all equal 1 or 0 (see Brandon 2005). That is a serious flaw in Sober's account, but it does not mean we have to settle for a purely phenomenological account.

Why do Walsh et al. think that it does? Given the impoverished range of options they present, logic seems to force this choice on them. But we think there is more to it than that.

In a related article Matthen and Ariew (2002) present another argument for what amounts to the same conclusion. Again Sober's comparison of evolutionary theory to Newtonian mechanics is the target. Matthen and Ariew (2002, 67) argue that different components of fitness are not comparable, and in particular that there is nothing like vector addition that would allow us to combine different ''forces'' of selection. So although we know that, everything else being equal, it is best to produce the minority sex in a population with a skewed sex ratio:

we have no way of calculating whether a given sex-selection strategy interacts with a given parental-care-strategy, and how the fitness produced by variants of these strategies combine. This inability to add the ''forces'' of fitness is even more pronounced when the source laws are in unrelated domains. Suppose a certain species undertakes parental care, is resistant to malaria, and is somewhat weak but very quick. How do these fitness factors add up? We have no idea at all. The theory of probability has no general way to deal with such questions. (2002, 67)

(This last sentence of this quote is quite odd. Why should the theory of probability tell us how different components of fitness interact in biology? Should that not be a matter of biology?)

The conclusion of this is the following:

The disanalogy is that, while force affords Newtonian mechanics the means to compare and add up the consequences of these diverse causes, fitness does not add up or resolve. *This is why population geneticists are forced to estimate fitness by measuring population change.* (2002, 68, emphasis added)

The logic of this argument is, we think, clear enough. Its conclusion is false, and we want to focus on that. But let us briefly examine the major premise.

First, population genetic models regularly do combine different factors of evolutionary change in straightforward ways. One could, for instance, write down a simple model that tracks the evolution of two alleles, $A$ and $a$, in a haploid population with discrete generations. Here the frequency of $A$ in generation 2 is simply the product of the fitness of $A$, $w_A$, and its frequency in generation 1, $p_1$. Thus (where $p_2$ is the frequency of $A$ in generation 2):

$$p_2 = w_A p_1$$

and similarly for change in $a$ (where $q_i$ is the frequency of $a$ in generation $i$):

$$q_2 = w_a q_1.$$

Given this very simple model we can easily add the effects of migration and mutation (where $\mu$ is the mutation rate from $A$ to $a$, $\upsilon$ is the mutation rate from $a$ to $A$, and $m_{1A}$ is the rate of loss of $A$ due to emigration, $m_{2A}$ the gain in $A$ due to immigration, $m_{1a}$ the rate of loss of $a$ due to emigration, and $m_{2a}$ the rate of gain of $a$ due to immigration):

$$p_2 = w_A p_1 + p_1(1 - \mu) + (1 - p_1)\upsilon - m_{1A} + m_{2A}$$
$$q_2 = w_a q_1 + q_1(1 - \upsilon) + (1 - q_1)\mu - m_{1a} + m_{2a}$$

The frequency of $A$ in generation 2, $p_2$, equals $w_A p_1$ plus the mutation rate from $a$ to $A$, minus the mutation rate from $A$ to $a$, minus the emigration rate of $A$, plus the immigration rate of $A$. Mutation, migration, and selection are fully comparable. This, of course, is not a discovery by us but is simply elementary population genetics. Thus, if Matthen and Ariew's claim were that different factors of evolutionary change, such as selection and mutation, are not comparable, their claim would be contrary to standard practice in population genetics and would be wrong.

But that is not their claim; rather they claim that different components of fitness are not comparable. Again this claim seems to be contradicted by standard population genetics. As Michod (1999, 12) points out, "Almost all models of natural selection involve some kind of fitness decomposition in one form or another." Perhaps the examples most familiar to philosophers are group selection models for the evolution of altruism. In such models different components

of fitness are separated in fitness equations. The following equations are representative:

$$w_s = 1 + bx$$
$$w_a = 1 + bx - c$$

where $w_s$ is the fitness of a selfish type and $w_a$ the fitness of an altruist, $c$ is the cost of altruism, $b$ is the benefit, and $x$ is the number of altruists within the particular group; $c$ represents the component of selection due to the selective disadvantage of altruism within a group, while differences in the value of $x$, and therefore of $bx$, represent the component of selection due to the selective advantage of groups with a larger number of altruists. Thus, in this case at least, different components of selection are comparable, just as selection, mutation, and migration are fully comparable. And so, it would seem, the major premise of Matthen and Ariew's argument is simply wrong.

We think it is wrong, but we are not sure that the preceding example fully addresses their point. Their point, we think, is that we have no *general* theory that would allow us to compare different components of fitness, that is, nothing like vector addition in Newtonian mechanics. In their example, we have no *theory* that allows us to combine the components of selection due to sex-ratio differences and those due to parental-care differences. This interpretation may explain their odd remark about the failure of probability theory to provide a framework for such a comparison. So our group selection example would be atypical in that in this case we do have an explicit theory of how to compare the individual- and group-selection components. If this interpretation of their remarks is correct, then we agree with them, but then the conclusion of their argument does not follow.

It is hardly surprising that we have no general theory that would allow us to predict the fitness of every possible combination of every character state. Any such theory we develop is likely to be local and post hoc. In a population where various sex-ratio strategies are extant as are various parental-care strategies we can *measure* the fitnesses of the extant combinations. With sufficient study we may be able to offer an ecological explanation of these fitness values. But the resulting generalizations do not derive from any general theory.

The fitness values of various types are among the basic parameters of models in population genetics. They are like other basic parameters, such as mutation rates, migration rates, and effective population size, in that they need to be measured empirically and cannot be predicted from some general theory. There is here a difference from, and a similarity to, Newtonian mechanics. The difference is the locality of these parameter values – that they apply to particular populations in particular environments – and the resultant need to remeasure them time and again (see Brandon 1994). The similarity is that the basic parameters of Newtonian physics need to be empirically measured as well, for example, the value of $G$, the gravitational constant.

But have we not just conceded Matthen and Ariew's point, namely, that fitness must be measured in terms of its consequences? No. First, we can develop, through detailed ecological investigations, local theories of organism-environment interactions that would allow us to measure fitness indirectly. Second, and much more importantly, when looking at the effects of fitness we do not have to look at evolutionary, or transgenerational, change.[8] We can, and biologists often do, look at something else.

## 3. METHODS FOR DETECTING SELECTION

John Endler (1986) in his comprehensive overview of studies of natural selection in the wild describes ten methods for detecting natural selection. For present purposes we do not need an analysis at that fine a grain, although we would recommend his account to any philosopher who would pronounce on how biologists *must* measure natural selection. A simpler classification results from first distinguishing between those methods that detect selection in terms of its effects versus those that detect selection in terms of its causes. Let us label the second category *CF*, for *c*auses of *f*itness. The first category needs to be further subdivided. The first subdivision – methods that detect selection in terms of *e*volutionary *c*onsequences – will be labeled *EC*. We will label the second subdivision – methods that detect selection by *d*irect *m*easurement of (parts of) the process of natural selection, that is, measurement of differential survivorship, mating ability, fertility, fecundity, and so forth – *DM*.

Studies using method *CF* are difficult in that they require detailed knowledge of organism-environment relations. As we will see they are

rare. But it is important that they are not impossible (Lewontin 1978; Brandon 1990, chap. 1). They are not. However, on the basis of our knowledge of evolutionary biology, we would say that we are close to being able to apply method *CF*, usually, if not inevitably, through repeated applications of method *DM*. We will return to this shortly.

Method *EC* is the method that Matthen and Ariew (2002) claim biologists *must* use.[9] What is common to all the cases we lump under *EC* is that patterns of variation, either extant (horizontal) or over time (vertical), are used to compare models of selection to a null model of no selection. The models may be informal and qualitative (Endler's I–III and some cases of V), or they may be formal and quantitative (Endler's IV and some cases of V). But the essential feature of all of these cases is that the past or present existence of selection is *inferred* from data that eliminate the null (no-selection) hypothesis. A few examples will clarify just how *EC* works.

If one observes a consistent correlation between some environmental feature and character state, then one can hypothesize that these environmental differences lead to different selective environments that result in the observed distribution of character states. For example, one might, as Kettlewell (1955, 1956) did, observe a correlation between the darkness of tree trunks (due to industrial pollution) and the frequency of the dark morph in *Biston betularia*. The selection hypothesis is that in woods affected by pollution the dark morph is selectively favored over the light form, and vice versa, in nonpolluted woods.[10] Of course, Kettlewell did not stop with this hypothesis; he went on to demonstrate experimentally, using method *DM*, that selection was indeed operating in accordance with the selection hypothesis. We think all will agree that this was a good thing. Although the observed patterns of variation were not consistent with the null hypothesis (which in this case would be that the different color morphs were distributed randomly about the different areas Kettlewell studied), they are consistent with still other hypotheses. For instance, the hypothesis that air pollution has a developmental effect on moths that darkens wing color is not eliminated by the observed patterns of variation. This sort of problem seems to be quite general when the models in question are informal and qualitative. If, as in the preceding case, there are multiple nonselection hypotheses, then the elimination of one of them will not automatically support the hypothesis of selection.

This is less of a problem, but still a problem, when the models are quantitative. To illustrate how *EC* works with such models we will describe two examples, the first a simple "toy" example, and the second a genuinely interesting piece of contemporary biological research.

Consider a simple example of heterozygote superiority. Suppose that there are two alleles, *A* and *a*, at a locus and that the locus is in linkage equilibrium with all other loci. The fitness of *Aa* is normalized to 1, and selection coefficients are assigned to the two homozygotic genotypes. A simple population genetic model shows how a population satisfying the preceding will settle into an equilibrium. The equilibrium frequencies of *A* and *a* can be mathematically derived given the values of the selection coefficients associated with the two homozygotes. That is, the fitness values mathematically determine the equilibrium frequencies of the two alleles. But the relevant equations work in both directions, so given the equilibrium frequencies we can determine the fitness values. How could we use this fact to infer not just the existence of selection, but the quantitative strength of selection?

If we observed stable allele frequencies at the locus over a number of generations, then we could show that the likelihood of the null hypothesis (in this case that the alleles are selectively neutral) is considerably lower than that of the selection hypothesis (see Brandon 2005 and Brandon and Nijhout forthcoming). And given the support of the selection hypothesis we could then go on to infer, in the manner outlined, the fitness values of the three genotypes. Of course these inferences are based on the assumptions mentioned, so one's confidence in the inferences should be proportional to one's confidence in the truth, or approximate truth, of the assumptions.

Consider this slightly different scenario. In this case we have no access to vertical data, but we can observe strong selection against one of the two homozygotes (e.g., we observe the negative effects of sickle cell anemia associated with a known homozygotic genotype at the hemoglobin locus). *If* we assume that the population is at equilibrium, then we can again eliminate the null hypothesis and estimate the fitness values of the other two genotypes.

The preceding "toy" examples are meant to illustrate clearly the inferential character of the *EC* method. Let us now turn to the work of Marty Kreitman and others, which we consider to be the most interesting use of *EC* method.[11] Kreitman has developed an elegant

method for determining where in the genome selection is acting, with drift being the explicit null hypothesis. Because of the redundancy of the genetic code, substitutions in the third position of a codon often produce synonymous codons (i.e., codons that code for the same amino acid). Given this fact, one can compare the behavior of the first two codons with that of the third. If selection is acting at the relevant genomic region, the first two positions (substitutions in which will not produce synonymous codons) should behave differently than the third. In contrast, if selection is not acting on the region, then the first two positions should be as free to drift as the third, and thus no difference is expected in the behavior of the third position. This is exciting work and is certainly a powerful way of investigating the selectionist/neutralist debate. But notice that this sort of work tells us nothing about the "why" of selection. It offers no ecological explanation of selection.

According to Darwinian theory, small differences in organisms can result in differences in various abilities and capacities, such as the ability to survive, the ability to attract mates, fertility, or fecundity. Although Matthen and Ariew (2002) complain that we have no way to combine or compare these different capacities, nature surely does, since at the end of the day, at the end of the generation, they combine to produce a given level of reproductive success. And this capacity is exactly what the propensity interpretation of fitness defines. As empirical biologists we should not be surprised that there is no general *theory* about how these various capacities combine to produce fitness, but there is a general *method* to investigate this. It is called *fitness component analysis* (see Endler 1986, 84–86, for discussion and references). Basically, we want to sample the population under study at as many life history stages as possible. The idea is that different capacities will manifest themselves at different stages of life history and we can then see *empirically* how they combine. The ideal, though rarely attainable, would be to observe every member of the population throughout its entire lifetime. In practice, biologists typically look at the *effects* of a small number of these capacities, getting a *direct measure* of some component of fitness. This is the method we are labeling *DM*.

Let us first describe a couple of familiar studies that have used the *DM* method; then we want to explore some important philosophical differences between it and the *EC* method.

The most famous studies of natural selection in the wild are those conducted by H. B. D. Kettlewell (1955, 1956). As we have already seen, the correlational data supported (perhaps only weakly) the hypothesis that selection was responsible for the increase in the melanic form of *Biston betularia* in woods downwind of large industrial areas. But Kettlewell did a series of experiments to support that hypothesis more strongly, and in his best known experiments he marked individual moths, released them into the woods, and then recaptured them several days later. He then compared the relative frequency of the two color morphs in the recaptured group to that in the released group. In the experiments conducted in woods downwind of industrial areas, he observed an increase in the relative frequency of the dark form in the recaptured group compared to the released group. On the basis of many auxiliary studies, he attributed this change in frequency to selection (by birds), and thus had a measure of *one component of fitness* of the two morphs in that environment.

As a measure of lifetime fitness, the sort of fitness that ultimately matters for evolution, Kettlewell's study is incomplete. It tells us nothing about how the two forms perform in the larval stage; it tells us nothing about any differences that might exist between them in mating ability, fertility, fecundity, and so on. Its power to explain the existing patterns of variation depends on the truth of the assumption that the two forms are more or less equivalent with respect to their other components of fitness. But given the fit between the observed selection differentials and the observed patterns of variation, that assumption is not at all unreasonable. And, most importantly, Kettlewell's work is a direct demonstration of the existence of selection in the areas he studied, during the life history stage that he studied.

Less familiar, but more complete, are the studies of the evolution of heavy metal tolerance in grasses conducted by Janis Antonovics and others (Antonovics, Bradshaw, and Turner 1971). Here fitnesses were measured more directly and more completely. As is that of Kettlewell's moths, this too is a study of adaptation to an environmental perturbation. Mining activities produce soils with high levels of heavy metals. These metals are typically toxic to most plants. When this contaminated soil is piled by the side of a mine, there is often a sharp boundary between metal-contaminated soils

and normal soils. This dramatic difference in a factor in the external environment leads to a dramatic difference in selective environments. Antonovics was able to show this by monitoring the lives of large numbers of individuals, both in the contaminated soil and in noncontaminated soil. Looking at differences in lifetime survivorship the strength of selection was measured in both selective environments. Genetically tolerant plants were strongly selected over nontolerant types in the contaminated soil, and vice versa in the normal soil. Surprisingly, selection was so strong that genetically differentiated subpopulations were produced over very short distances in spite of considerable gene flow between them.

Antonovics's measures of fitness values were more complete than Kettlewell's in that he was able to look at lifetime survivorship, rather than survivorship during a small portion of the life cycle. They were more direct in that differential deaths were actually observed, rather than inferred from differences between released and recaptured groups. But this still is a measure of a component of fitness, not complete fitness. For instance, there was no attempt to measure potential differences in fertility and fecundity. But given the strength of selection observed, and the remarkable genetic differentiation associated with the different selective environments, we can be reasonably confident that Antonovics's measures of fitnesses captured a crucial part of the causal story.

As we said previously, the ideal application of the *DM* method would be to observe every stage of the life histories of the organisms in the population under study, and so to measure every component of fitness and to then see how, in this particular situation, those components combine to produce overall fitness. There is absolutely no philosophical or conceptual difficulty in doing this. The difficulties are of a practical nature. And so biologists using the *DM* method almost always measure some component or components of fitness. Such measures provide good (but not complete)[12] explanations of evolutionary change to the extent that the measured components of fitness dominate the unmeasured components.

One might think that the difference between the *EC* and *DM* methods is rather minor, that they both detect fitness values in terms of effects, and that the only real difference is that the *EC* method looks for effects over an evolutionary time scale while the *DM* method looks at effects of traits such as differential mating

ability that are observable over the time scale of a single generation. That is a difference, and an important difference, but it is not the only difference.

Before discussing the less obvious differences between these two methods, let us briefly comment on the major consequence of the preceding difference. It is not minor. Since *EC* looks at trans-generational effects, it necessarily confounds the effects of the ecological process of selection (which is basically what *DM* studies) with multiple effects of the genetic system. For instance, in cases of heterozygote superiority, once the population reaches equilibrium there is no evolutionary change. Thus there are no *EC* effects. With the appropriate vertical data (many generations with no change at that locus) or horizontal data (the exact same system found in a number of related species) the *EC* method could eliminate the null hypothesis of neutrality, but it could not estimate the selection differentials. Without those data the *EC* method cannot even differentiate a case of strong selection, say that both homozygotes are lethal, from no selection. Combining the results of two separable processes results in a loss of information. And, unfortunately for the *EC* approach, that information is crucial to evolutionary explanation and prediction. We will return to this shortly.

At least as important philosophically is a less obvious difference between the two methods. Although we will need to add a little nuance to this, the *EC* method is model based and inferential; the *DM* method is not model based and is appropriately described as *measurement*.

Let us focus on how we can come to know fitness values by means of these methods. It should be clear from our discussion of *EC* examples that data about patterns of variation are used to support a model of selection, from which fitness values can be estimated. Without the model, there could be no estimation of fitness values, since the data are simply patterns of horizontal and/or vertical variation. Philosophically speaking, the inference to fitness values in these cases is *abductive*.

Now we do not want to claim that in using the *DM* method no inferences are made. Remember that Kettlewell inferred that the differences in relative frequency of the two forms in the recaptured class compared to the released group reflected differential predation. He, in fact, had a lot of evidence to back up that inference. We do not

think, however, we should describe that evidence as a model. Antonovics's more direct measurements avoided that particular inference. We are not particularly concerned here with the curse of post-Kuhnian philosophy of science – the view that all observation is theory laden. Whatever one thinks about that, we would hope one would still be able to distinguish measurements of some quantity in nature from model-based inferences. Surely much *DM* work is done because the investigator is interested in some hypothesis or model. And one can always describe any parameter measurement as a hypothesis test (see Brandon 1994). But that possibility does not mean that that is the most perspicuous description. The *DM* method is appropriately named; it is a method of parameter measurement, one that is more or less direct.

## 4.    APPLYING THE METHODS OF SELECTION DETECTION

When Matthen and Ariew claim that "population geneticists are forced to estimate fitness by measuring population change," we take this to be both a descriptive and a prescriptive claim. If biologists are forced to use the *EC* method, then they do use it, and presumably use it exclusively. That is descriptive. Furthermore if they must estimate fitness that way, then they ought to do it that way. That is prescriptive.

Our own prescriptive views may well have come through in the last section, but for the record, let us be explicit. We think that it is important that it be possible to apply the *CF* method, in principle at least. But we think that the only way we will have the biological knowledge required to apply that method is through repeated applications of the *DM* method. In this way the *DM* method has priority over the *CF*. We have also said that some really interesting work has been done by using the *EC* method. Were we in charge, we would certainly fund more of it. But the *EC* method is really a method of last resort. In Krietman's studies it is used because he is looking over vast expanses of evolutionary time and is looking at genomic regions where the function is often unknown. It would be impossible to apply the *DM* method here. But we can imagine no situation in which both the *DM* and *CF* methods were applicable and the *CF* method preferable. In this sense, the *DM* method is a better way of doing evolutionary biology.

Prescriptive disputes are not easily settled, though we hope we have given some good reasons in support of our views. But the descriptive implications of Matthen and Ariew's claim can be easily dismissed. There are real data here.

Table 5.1 in Endler (1986) lists all of the published demonstrations of natural selection in the wild that Endler could find. Surely he missed one or two, but this is by far the most comprehensive survey of the literature in existence.[13] Endler lists the studies by species and then by the traits studied in that species. For instance, in *Homo sapiens* there are entries for tooth size, birth weight and gestation time, height, body shape, and haemoglobin S. For each species trait he records the method(s) of demonstration. We mapped Endler's ten methods onto our three methods as follows: Endler's I–V are our *EC*, his VI–VIII are our *DM*, and his IX–X equate to our *CF*. When Endler listed more than one method we counted more than one of our methods *if* his listed methods crossed our categories (as happened only once). The results are as follows:

| Method | Number of studies |
| --- | --- |
| *EC* | 1 |
| *DM* | 172 |
| *CF* | 2 |

I think it is fair to conclude from this that Matthen and Ariew's descriptive claim is false. And if their prescriptive claim is true, then evolutionary biologists are certainly not behaving as they ought.

## 5. CONCLUSIONS

If cavers were to race from one cave entrance to another, the winner would surely owe her success to such characteristics as her ability to navigate, her swiftness at making vertical ascents and descents, and her ability to squeeze through narrow apertures. The aboveground observer will recognize that these skills are necessary but would not be able to say how the caver's skills combined to lead her to victory. Those who argue for the ESI apparently think that biologists are in the same epistemological cul-de-sac.

Those who argue for the PSI would argue that the case of the cavers is disanalogous to natural selection in two ways. First, unlike the aboveground observer, scientists are able to observe not just the *outcome* of natural selection, but also the *process*. This would be like being able to track the cavers and see, for example, how many navigational errors they make or how fast they are able to make ascents and descents. From these data we could (1) *explain* why the winner won the race and (2) *predict* how the cavers would fare in different caves. Analogously, as we saw earlier, we can use *CF* to measure the variety of factors that lead to the success of an individual organism or type of organism. We can use these data to explain the success of organisms and predict how they would fare in different environments.

The second disanalogy is that unlike the singular cave race, biological phenomena are repeatable. Even if we could not observe how the cavers act underground, we could race them in a number of different caves, some without alternate routes (to eliminate navigational errors) and others with many, some with few vertical drops and others with many. Through this comparison, we could see which cavers fare well in which kinds of caves. This would allow us to learn which individual cavers (or caver type) do well in which kind of cave. We could use these data to predict which caver would win in a particular cave race and to explain why the winner won and the loser lost. Analogously, we could use *DM* to see how different organisms fare in different environments. We could even clone organisms and raise the clones in a diversity of environments. This would give us data to understand how different components of the organism's fitness combine to prove successful in a particular environment. As we saw in the previous section, *DM* is commonly employed by biologists.

In sum, the arguments for the ESI have been thoroughly refuted. This is a good thing. It is incapable of explaining differential reproduction, a key part of the process of evolution by natural selection. It can merely posit the existence of population-level statistical distributions – they emerge in mysterious ways. And it can make no sense of the way biologists actually measure fitness in the wild. In contrast, the PSI does these things easily and naturally. The ESI does have the advantage of allowing one to hang on to a philosophical prejudice, namely, that phenomena at the level of individual

organisms are deterministic. But, as we have seen, hanging on to this particular prejudice is quite costly.

NOTES

1. Walsh, Lewens, and Ariew (2002), and Matthen and Ariew (2002); see also Sterelny and Kitcher (1988).
2. In 1978 Brandon defined adaptedness as follows: for an organism $O$ in environment $E$ there is a range of possible offspring numbers, $Q_1^{OE}$, $Q_2^{OE}$, ..., $Q_n^{OE}$, and for each number there is an associated probability, $P(Q_I^{OE})$. The adaptedness of $O$ in $E$, $A(O, E)$, then is the *expected value* of $O$'s reproductive success in $E$. That is, $A(O, E) = \Sigma\ P(Q_I^{OE})Q_I^{OE}$. Later, drawing on the work of John Gillespie (1973, 1974, 1977), it was discovered that this expected value needed to be discounted by some function of the variance in offspring number (see Brandon 1990, 18–20). For further discussion of this point, see Beatty and Finsen (1989) and Sober (2001).
3. We emphasize the word ''supposedly'' because it seems to us that our confidence that the underlying mechanics is really deterministic should be much lower than our confidence in the second law.
4. Sterelney and Kitcher (1988, 345) argue that ''evolutionary theory, like statistical mechanics, has no use for such a fine grain of description: the aim is make clear the central tendencies in the history of evolving populations.''
5. We will not be able to deal with the general issue of ontological commitment here, but let us simply assert our view that parsimony is not an ontological virtue; rather accuracy is. The world either is or is not a simple place. Our job is to describe it as it is, not as we wish it were.
6. See Stephens (2004) for a recent endorsement of the Newtonian option.
7. But see Brandon (2005).
8. Matthen and Ariew (2002) do not fully understand the implications of their position. As they define it ''*predictive fitness* (as we shall call it) is a statistical measure of evolutionary change, the *expected* rate of increase (normalized relative to others) of a gene, a trait, or an organism's representation in future generations.'' Thus in the conclusion quoted when they speak of population geneticists' being forced to estimate fitness by ''measuring population change,'' we must interpret ''population change'' as transgenerational change. Their conception of fitness is not novel, it is called Fisherian fitness. It is unsuited for explanatory purposes (see Brandon 1990, chap. 1; Ramsey 2006). Unfortunately for Matthen and Ariew, they do realize that they are committed to this (see footnote 30, 74).

9. Walsh and colleagues (2002) are less explicit on this point, but it seems they also are committed to this view.

10. The vice versa hypothesis is not really necessary here but was part of Kettlewell's explicit experimental research. See Brandon (1999) and Rudge (1999).

11. See Yang and Bielawski (2000), Bamshad and Wooding (2003), and Hamblin, Thompson, and DiRienzo (2002).

12. See Brandon (1990, chap. 5) for an account of ideally complete adaptation explanations.

13. Of course, it is now seventeen years old. For a more recent (1984–97) list of studies of natural selection in the wild, see Kingsolver et al. (2001).