

# The Pomp of Superfluous Causes: The Interpretation of Evolutionary Theory\*

D. M. Walsh<sup>†‡</sup>

---

There are two competing interpretations of the modern synthesis theory of evolution: the dynamical (also known as ‘traditional’) and the statistical. The dynamical interpretation maintains that explanations offered under the auspices of the modern synthesis theory articulate the causes of evolution. It interprets selection and drift as causes of population change. The statistical interpretation holds that modern synthesis explanations merely cite the statistical structure of populations. This paper offers a defense of statisticalism. It argues that a change in trait frequencies in a population can be attributed only to selection or drift against the background of a particular statistical description of the population. The traditionalist supposition that selection and drift are description-independent causes of population change leads the dynamical interpretation into a dilemma: it must face a contradiction or accept the loss of explanatory power.

---

**1. Introduction.** The modern synthesis is an elegant, powerful, and highly successful theory of evolutionary change. But if the current debate about its interpretation is anything to go by, it is far from clear what it tells us about the workings of the biological world. There are two competing interpretational approaches to the modern synthesis—the dynamical (also known as the ‘traditional’) and the statistical—and the dispute between them has been receiving a considerable degree

\*Received April 2005; revised August 2007.

<sup>†</sup>To contact the author, please write to: IHPST/Department of Philosophy, University of Toronto, 91 Charles Street West, Toronto, ON, CANADA M5S 1K7; email: denis.walsh@utoronto.ca.

<sup>‡</sup>I wish to thank audiences at Stirling, Toronto, and CSHPS for helpful questions and comments. I have benefited greatly from discussion with Sorin Bangu, Anjan Chakravartty, Ronnie de Sousa, Ronpal Dosanjh, Margie Morrison, Elliott Sober, Kyle Stanford, Chris Stephens, and Michael Strevens. I am particularly indebted to my fellow ‘statisticalistas’ André Ariew and Mohan Matthen.

Philosophy of Science, 74 (July 2007): 281–303. 0031-8248/2007/7403-0007\$10.00  
Copyright 2007 by the Philosophy of Science Association. All rights reserved.

of attention of late.<sup>1</sup> Both approaches accept that the modern synthesis theory explains evolutionary change by allowing us to differentiate certain phenomena of populations—selection, drift, mutation, and migration. Dynamical and statistical interpretations part significantly, however, on the metaphysical implications of these explanations. The dynamical interpretation contends that population change results from the combined actions of discrete, proprietary causal processes: selection, drift, mutation, and migration. By ‘discrete’ I mean that these processes act independently, alone or in concert, and that the modern synthesis theory of evolution allows us to identify, resolve, and quantify their respective causal contributions. By ‘proprietary’ I mean that what it is for a change in the structure of a population to constitute selection (or drift, migration, mutation) is for it to be *caused* by the process of selection (drift, migration, mutation). Sober’s definitive statement of the dynamical interpretation sums this up nicely:

In evolutionary theory, the forces of mutation, migration, selection and drift constitute causes that propel populations through a sequence of gene frequencies. To identify the cause of a population’s present state . . . requires describing which evolutionary forces impinged. (Sober 1984, 141)

This is just what the statistical approach denies. According to the statistical interpretation, modern synthesis explanations do not account for population change by citing the actions of specific causal processes of selection, drift, mutation, and migration. They explain merely by citing the statistical properties of populations. On this view, what it is for a change in relative trait frequencies to constitute selection (or drift) is merely for it to be susceptible to a certain kind of statistical description.

The dynamical interpretation has an immediate intuitive appeal, and a popularity, that its statistical adversary lacks. If only for this reason the burden of proof weighs more heavily on the statistical interpretation. In this essay I attempt to offer it some support. My defense of statisticalism takes its cue from two quite different but equally robust challenges, one raised by Bouchard and Rosenberg (2004) and the other by Stephens (2004), Reisman and Forber (2005), Millstein (2006), Abrams (2007), and Shapiro and Sober (2007), *inter alia*. In what follows, as has become customary, I shall concentrate on selection and drift explanations; mutation and migration are special cases and will be discussed elsewhere.

1. Recent discussions of this issue include Bouchard and Rosenberg (2004), Stephens (2004), Matthen and Ariew (2005), Reisman and Forber (2005), Rosenberg and Bouchard (2005), Brandon (2006), Millstein (2006), Rosenberg (2006), Brandon and Ramsey (2007), and Shapiro and Sober (2007).

**2. Mere Statistical Effects.** The statistical interpretation incorporates a positive thesis about modern synthesis explanation and a negative thesis about modern synthesis metaphysics. They are:

S1. Explanations that invoke selection and drift cite *only the statistical properties of populations*.

S2. Selection and drift are not causes of population change; they are *mere statistical effects*.

S1 is argued for on the grounds that the property that explains population change in the modern synthesis theory is variation in trait fitness (see Matthen and Ariew 2002; Walsh, Lewens, and Ariew 2002; Walsh 2003, 2004; Ariew and Lewontin 2004). Trait fitness is a statistical parameter, usually measured as the mean and variance of the fitnesses of organisms with a given trait (Gillespie 1977; Sober 2001).<sup>2</sup> S2 is supported by the observation that changes in trait frequencies are realized by the differential survival and reproduction of individual organisms. Yet as Walsh et al. (2002) and Walsh (2004) demonstrate, none of the causes of individual survival and reproduction, and no aggregate of these causes, qualify as selection or drift. S2 is the crux of the issue, yet the arguments offered so far in its support are not wholly compelling.

For one thing, the notion of a ‘mere statistical effect’ needs some fleshing out. I have no definition to offer, but an (admittedly fanciful) example might serve to make the idea a little more vivid. Among three chosen varieties of apples, Braeburns on average have the largest mass (10.60 standard apple units), Jonagolds are slightly smaller (10.55 sau), and Pippins are smaller still (10.50 sau). Whereas Braeburns vary only little in mass (standard deviation = 0.2), Jonagolds vary rather more (s.d. = 0.7) and Pippins vary a lot (s.d. = 1.0). Suppose that we conduct an experiment in which we select at random and weigh 10 apples of each variety from five fruit stalls in each of five markets. (In all, 250 apples from each variety are weighed from 25 fruit stalls.) A simulation of this experiment was run in which the apples in the stalls were drawn randomly from the population and the stalls in each market were randomized.<sup>3</sup> Here are the results. The sample means preserve the order of mean mass for the varieties:  $\hat{X}_{\text{BRAEBURN}} = 10.60 > \hat{X}_{\text{JONAGOLD}} = 10.53 > \hat{X}_{\text{PIPPIN}} = 10.47$ . If we count, for each variety, the number of treatments in which *it* is on

2. “[F]itness is not a cause of evolution, but rather a measure of growth” (Matthen and Ariew 2005, 355).

3. The random numbers for each mean and standard deviation were generated by Microsoft Excel. The stalls were randomized by sequences generated by <http://www.random.org>.

average the largest, we get the following results. Taking the relevant treatments to be markets ( $N = 5$ ), we get Braeburns = 4, Jonagolds = 1, and Pippins = 0. Taking the relevant treatments to be fruit stalls ( $N = 25$ ), we get Braeburns = 14, Pippins = 7, and Jonagolds = 4. The rank order of the three varieties, then (i.e., the rank according to the number of treatments in which each variety is on average the biggest), considered by *market* is Braeburns > Jonagolds > Pippins. The rank order considered by *fruit stall* is Braeburns > Pippins > Jonagolds. The apparent reversal of rank orders between markets and fruit stalls I shall call the ‘rank order effect’. It is an example of what I am calling a ‘mere statistical effect’.

Rank order effect demonstrates certain diagnostic marks of a ‘mere statistical effect’. Significantly for what follows, it has three salient features:

1. *Statistical explanation*: The effect is explained by the statistical properties of the setup and not by the causal properties of the apples. The reversal of rank order is explained by the fact that while the *mean* mass of Jonagolds is greater than that of Pippins, the *standard deviation* among Pippins is higher. As sample sizes decrease—that is, as we go from sampling markets to sampling stalls—the probability of Pippins outranking Jonagolds increases.
2. *Description dependence*: There is no absolute rank order in this population of apples. Whether the rank order is  $B > J > P$  or  $B > P > J$  depends on which distribution we choose to describe the population by (i.e., by stalls or by markets). Moreover, rank order ‘effect’ is simply the consequence of describing the population of sampled apples first one way and then another.
3. *No causal inferences*: The explanation of rank order effect sanctions no particular causal inferences. The fact that there are different rank orders within stalls and markets should not lead one to posit some causal process operating within stalls that is absent within markets, or vice versa. Nor should the ‘change’ in rank order as one moves from a distribution described by stalls to one described by markets be taken as evidence for a particular causal process at work.

Rank order effect demonstrates how the statisticalist theses S1 and S2 go nicely together, that is to say, how *statistical properties explain mere statistical effects*. The statistical interpretation of the modern synthesis maintains that natural selection and drift explanations are relevantly like explanations of rank order effect. Changes in the structure of a population are described by the statistical properties of the population (as per point 1 above). Whether these changes constitute selection or drift is simply a matter of how the statistical properties are described (as per point 2) and not a matter of how they are caused (as per point 3). Consequently,

selection and drift explanations do not articulate the underlying causes of population change. Selection and drift are mere statistical effects.

**3. Dynamical Responses.** There are two general strategies pursued by supporters of the dynamical interpretation in response to the statistical interpretation. Despite their differences, they share an objective: to undermine S2. The essence of the dynamical interpretation, after all, is that natural selection and drift explanations articulate the causes of population change. The first strategy, embodied in Bouchard and Rosenberg (2004), seeks to undermine S2 by casting doubt on S1. The second, articulated by Stephens (2004), Reisman and Forber (2005), Millstein (2006), and Shapiro and Sober (2007), accepts S1 but rejects S2.

*3.1. Fitness and Individual-Level Causes.* Bouchard and Rosenberg (2004) and Rosenberg (2006) argue that even though natural selection explanations cite the distribution of trait fitnesses, these distributions are fixed by the causal properties of individuals, namely, individual (ecological) fitnesses; and ultimately, it is these individual fitnesses that explain evolutionary phenomena:

In evolutionary theory, all we need to understand where fitness coefficients of populations come from is the ‘concession’ that there is such a thing as comparative differences in (ecological) fitness between pairs of organisms; and these differences can be aggregated into fitness differences between populations. (Bouchard and Rosenberg 2004, 703)

If we take selection to be differential survival and reproduction of individuals (by dint of their fitnesses), the negation of thesis S2 follows quite straightforwardly: “selection [is] a contingent causal process in which individual fitnesses are the causes and subsequent population differences are the effects” (710).

In making their case for the explanatory primacy of the causal properties of individual organisms, Bouchard and Rosenberg attempt to set aside one of the mainstays of the statisticalist position, the analogy between modern synthesis theory and the statistical interpretation of thermodynamics. They illustrate the putative analogy by use of the concepts of an *arrangement* and a *distribution*. An arrangement in thermodynamics comprises all the positions and momenta of the particles in a system. Each arrangement realizes a *distribution* of energy in the system. In thermodynamics, the explanatorily basic property is entropy, a measure of the unevenness of the *distribution* of energy within a system. Similarly, an arrangement in the modern synthesis comprises a specification of survival and reproduction rates of all individuals and the heritable traits they

possess. Each arrangement realizes a *distribution of growth rates* of trait types in a population. Within a biological population, it is the *distribution of trait fitnesses* that predicts and explains the rate and degree of change in population structure. This is the statisticalist thesis S1.

The statisticalist argument from analogy goes that if distributions of energy are explanatorily basic, statistical properties of ensembles in thermodynamics, then we ought to consider that distributions of trait fitnesses are explanatorily basic, statistical properties of populations in modern synthesis theory. Bouchard and Rosenberg, however, argue that there is a significant disanalogy between modern synthesis theory and thermodynamics, one that vitiates the appeal to thermodynamics as a model for modern synthesis explanations. In the modern synthesis theory the relation between an arrangement and a distribution makes clear why distributions explain, but not in thermodynamics.

It is easy to see that when growth rates among trait types are unevenly distributed, certain consequences follow: one trait type preponderates over others, concomitantly, variation in growth rates (trait fitnesses) decreases, and the average growth rate increases. Changes that decrease the variation in trait fitness and increase average trait fitness are obviously more likely than those that do not; this is Fisher's fundamental theorem (1930). Analogously, in thermodynamics, it would be clear why ensembles tend to move from those arrangements that realize lower-entropy distributions to higher if these changes were demonstrably more likely than the converse. But, according to Bouchard and Rosenberg, they are not. For each distribution, there are continuum many arrangements that realize it. So the class of arrangements that realize higher-entropy states is not larger than that of lower-entropy states. Consequently, it cannot be demonstrated from the relation between arrangements and distributions alone that in a thermodynamic system changes from low to high entropy are more likely than the converse. Entropy, the property that explains the trajectories of ensembles in thermodynamics, cannot be a simple function of the causal properties of the particles. According to Bouchard and Rosenberg, it is an irreducibly ensemble-level, statistical property (see also Rosenberg and Bouchard 2005).

But the counterpart of entropy in the modern synthesis—variation in trait fitnesses—is wholly determined by the arrangement of causal properties of organisms (their individual fitnesses). Trait fitness supervenes on individual fitness. And this supervenience makes a substantial difference.

[T]he features that make entropy an emergent property in the second law of thermodynamics are largely absent from the foundations of the theory of natural selection. The emergent character of the second law is generated by the fact that entropy is not a property of the

individual components of the ensemble, but of the ensemble as a whole. (Rosenberg and Bouchard 2005, 349)

[T]he *fitness* of an ensemble is nothing like the *entropy* of an ensemble, just because *unlike* entropy, fitness is a calculable value of the properties of the components of the ensemble. (Bouchard and Rosenberg 2004, 704; emphasis in original)

The thermodynamics analogy, then, offers no support for the statistical interpretation of modern synthesis theory, because, unlike ‘entropy’, ‘trait fitness’ is definable in terms of the causal properties of elements of the ensemble. “Once we understand the differences between entropy and fitness, the temptation to treat natural selection theory solely as a claim about ensembles disappears” (Bouchard and Rosenberg 2004, 702).

Bouchard and Rosenberg use the putative disanalogy as part of a strategy to undermine the importance of thesis S1. S1, while true, they suppose, does not suggest that statistical (distributional) properties play an *irreducible* explanatory role. Their argument goes as follows:

T1. Distributions of trait fitnesses explain changes in trait frequencies in a population.

T2. Distributions of trait fitnesses are wholly determined by the arrangement of causal properties of individual organisms (individual fitnesses).

T3. Therefore, individual fitnesses explain the changes in trait frequencies.

T1 is simply just S1. T2 is simply the thesis that trait fitnesses supervene on individual fitnesses. But T3 entails, contra statisticalism, that natural selection explanations ultimately appeal to the individual-level causes of population change.

One line of statisticalist response might be to attempt to reinstate the thermodynamics analogy (Matthen and Ariew 2002, 2005). One might resist T2 on biological grounds. There is underdetermination of the ensemble-level explanatory properties by the individual-level causal properties for fitness, just as (allegedly) there is for entropy. This strategy gets its impetus from an argument offered by Ariew and Lewontin (2004) to the effect that there is no general criterion of trait fitness that is both explanatorily adequate for the purposes of modern synthesis theory and specifiable exclusively in terms of the causal properties of individual organisms. Nor should we think that there ought to be; trait fitness and individual fitness “are distinct concepts coming from distinct explanatory schemes” (347).

The differential growth rates of trait types in a population, they point

out, depend not just on the causal properties of individuals, but also on demographic features of the population. For example, if a population grows from size  $N$  to  $N + \delta$  in the time interval  $t_0$  to  $t_1$ , an organism,  $\alpha$ , that contributes  $n$  individuals to the population by reproduction early in the cycle (at  $t_0$ ) will increase its representation in the population by  $n/N$ . Had  $\alpha$  contributed  $n$  individuals by reproduction late in the cycle (at  $t_1$ ), it would have increased its representation in the population by a lesser amount:  $n/(N + \delta)$ . So the individual-level capacity to produce  $n$  offspring realizes different trait fitnesses in different demographic contexts. Examples of this sort can be elaborated to demonstrate that there is no general criterion of trait fitness to be given in terms of the causal properties of organisms.

One might be tempted to conclude from Ariew and Lewontin's argument that just as an explanatorily adequate conception of entropy cannot be specified in terms of the causal properties of the elements of an ensemble, so an explanatorily adequate conception of trait fitness cannot be specified in terms of the ecological fitnesses of individuals. So then, it is not true that "*unlike* entropy fitness is a calculable value of the components of the ensemble" (Bouchard and Rosenberg 2004, 704; emphasis in original). The reasons for accepting that entropy is an irreducible ensemble-level statistical property should apply equally to trait fitness, after all.

I am inclined to believe that this conclusion would be a misappropriation of Ariew and Lewontin's line of reasoning.<sup>4</sup> All their position demonstrates is that the fitness consequences of an organism's causal properties are sensitive to demographic contexts. But it does not follow that, for any given demographic context, the distribution of trait fitnesses is not wholly fixed by the causal properties of individual organisms. After all, the demographic contexts (e.g., population size, growth rate) are also fixed by the properties of individual organisms. So far, the Bouchard and Rosenberg argument for T2 looks to be on fairly solid ground.

The second, to my mind more trenchant, response to Bouchard and Rosenberg is just to accept the premises T1 and T2 and simply deny the conclusion: T3 does not follow. Consider the analogue of the Bouchard and Rosenberg argument applied to the rank order effect discussed above. In our apple experiment, the *distribution* of causal properties of the apples (i.e., mass) explains the rank order effect (cf. premise T1). The distribution of apple masses is fixed by the causal properties of individual apples (cf. T2). So, by parity, it should follow that the causal properties of the individual apples explain the rank order effect (cf. T3). But they do not;

4. André Ariew (personal communication) is inclined to believe that it is not.



the analogue of T3 is false in the rank order effect example. The reversal in rank order between Jonagolds and Pippins is explained by the fact that the *variance* among Pippins is higher than that of Jonagolds. So as sample sizes decrease, the probability of Pippins outranking Jonagolds increases. Bouchard and Rosenberg have simply assumed that if the arrangement of causal properties fixes the distribution, then the arrangement explains everything the distribution explains. But, as rank order effect demonstrates, where the explanandum is a mere statistical effect, the assumption does not hold.

The Bouchard and Rosenberg argument that, ultimately, the arrangement of causal properties in a system explains everything that the distribution explains fails to countenance the possibility of ‘mere statistical effects’. In this sense, Bouchard and Rosenberg beg the question against the statisticalist interpretation.

3.2. *Population-Level Causes.* Stephens (2004), Reisman and Forber (2005), Millstein (2006), Abrams (2007), and Shapiro and Sober (2007) take another line. They accede to S1, the explanatory autonomy of statistical properties of populations, but reject S2, their noncausality. Stephens says:

While it is true that we can explain how a population can be expected to change by citing trait fitnesses (which are statistical properties). If we want to know *why* the trait fitnesses have the values they have, however, we need to appeal to the causal notion of selection. (2004, 562; emphasis in original)

Selection and drift are *population-level* causes:

[N]atural selection is neither a *purely* statistical (acausal) population-level summation, nor is it a process of individual-level causation. Natural selection is, properly understood, a process that exhibits population-level causation. (Millstein 2006, 651; emphasis in original)

This interpretation originates with Sober’s (1984) influential reading of the modern synthesis.

Stephens urges us to recognize a distinction between selection and drift the *processes* and selection and drift the *product* of those processes.<sup>5</sup> The Stephens and Sober version of the dynamical interpretation maintains that drift and selection are equally both effects *and* population-level

5. Shapiro and Sober (2007) also make this claim. Millstein (2002) argues for the product/process distinction with respect to selection and drift. Hodge (1987) states that selection and drift are differentiated by the fact that they are different kinds of causes of population change.

causes. To date, statisticalists have only argued that selection and drift are not *individual-level* causes, but as yet they have offered no reason to believe that selection and drift are not *population-level* causes. If the Stephens and Sober position is correct, statisticalism, in particular thesis S2, must be wrong.

**4. Drift and Selection as Population-Level Causes.** It is widely held that drift is sampling error; in fact it is something of an analytic truth. The term was introduced by Wright (1931) to mean precisely this. Sampling error, in turn, is a statistical effect. It results when the distribution within a sample diverges from the structure of the population as a whole. This stipulative definition of drift-the-effect, however, leaves open the question whether there is a proprietary cause, or causal process, responsible for sampling error. Sober (1984) and Stephens (2004) claim that there is, and they tell us how to measure it (see also Beatty 1984; Millstein 2002; Reisman and Forber 2005; Abrams 2007). Sampling error—the effect—has a systematic relationship to certain statistical parameters of populations. There is a *prima facie* plausibility to the thought that these statistical parameters are the *causes* of sampling error.

To demonstrate this, we can introduce an experimental setup, a series of independent trials, where there are two possible outcomes for each trial,  $p$  and  $q$ . Where  $\Pr(q) = 1 - \Pr(p)$ , the chance of getting  $i$   $p$ 's in a series of  $n$  trials is given by the Bernoulli theorem:

$$\Pr(i) = (n!)/i!(n-i)! \Pr(p)^i \Pr(q)^{n-i}.$$

Statistical setups like this are often used as an illustration of evolution in populations of traits. Suppose that  $\Pr(p)$  and  $\Pr(q)$  each stand for the fitness of an allele at a given locus and  $n$  is the population size. Where  $\Pr(p) \neq \Pr(q)$ , there is selection; where  $\Pr(p) = \Pr(q)$ , there is none. Sampling error plays the role of drift. The probability of the frequency of allele  $p$ 's increasing in the population to a specified degree is a function of the fitnesses of  $p$  and  $q$  and the size of the population,  $n$ .

Sampling error, the putative cause, is easily picked out in this sort of setup. For any determinate values of  $\Pr(p)$  and  $\Pr(q)$ , the probability of getting a significant departure from the ratio  $p : q$  predicted by  $\Pr(p) : \Pr(q)$  in a series of  $n$  trials is an inverse function of  $n$  (sample size). Where  $n$  is very large, deviation from expectation—sampling error—tends to be small. Where  $n$  is small, large deviations from the expected outcome are common. The relation of sample size to sampling error allows the Stephens and Sober view to introduce the distinction between product and process that statisticalists neglect to make. Sampling-error-the-effect (product) is constituted by the actual amount of deviation from expectation in a given sequence of trials. Sampling-error-the-cause (process) is captured by the

systematic relation between sample size and the probability of deviation from expectation. Just as one would hope, sampling-error-the-cause explains sampling-error-the-effect.

Analogously, drift is both a statistical effect—actual deviation from the expected outcome—and a cause operating within a population. The magnitude of the cause is measured by the way the probability of a given deviation changes as a function of *population size*. Stephens says of drift:

It is a *population-level* cause. One sees the differential causal impact of drift *only* by comparing populations of different size. Drift plays a larger role in flipping a coin 10 times than it does in a coin 10,000 times. (2004, 556; emphasis in original)

In the same way, one supposes, the distinction between process and product can be applied to selection. Selection-the-process is measured by the way the probability distribution of an outcome (for a given  $n$ ) changes as a function of the probabilities of  $p$  and  $q$  (i.e. their fitnesses). Actual change in gene frequencies is the product of a process whose efficacy is measured by the difference between  $\Pr(p)$  and  $\Pr(q)$ .

Interpreted in this way, the statistical parameters of a population identify certain causal propensities (Abrams 2007). The relative values of  $\Pr(p)$  and  $\Pr(q)$  (for a given value of  $n$ ) identify a propensity of populations—to produce a preponderance of one allele over the other. The value of  $n$  (for given values of  $\Pr(p)$  and  $\Pr(q)$ ) identifies an independent propensity of populations—to *diverge* from the outcome predicted by the values of  $\Pr(p)$  and  $\Pr(q)$ .

The case for these statistical properties being genuine causes has been bolstered by Reisman and Forber (2005). These authors claim that, according to the interventionist criterion for individuating causes (Woodward 2003), these statistical parameters ought to count as causes. They point out that in an experimental setup such as this, there are two independently manipulable parameters:  $n$  and the ratio  $\Pr(p) : \Pr(q)$ . Interventions on these parameters have distinct, systematic effects on the change in a population's structure. Abrams concurs:

The sense in which natural selection and drift causally interact is this . . . . [D]rift is the aspect of those distributions that is controlled by population size, and selection is the aspect of the distributions that is controlled by fitness differences. (2007, 14)

Casting selection and drift in the role of probabilistic, population-level causes gives impetus to the intuitions informing the dynamical interpretation. Selection and drift turn out to be discernible, independent, quantifiable explanatory parameters: variations in trait fitness on one hand and the size of the population  $n$  on the other. Each has distinctive ex-

planatory consequences, and each can be varied independently of the other. They may ‘operate’ singly, or together, antagonistically or cooperatively. To some degree their respective effects are decomposable, in the way that Newtonian forces are (Sober 1984). Most important for our purposes, casting selection and drift as population-level causes allows the dynamical interpretation to endorse thesis S1—the statistical nature of modern synthesis explanation—while denying thesis S2—the *noncausal* nature of selection and drift.

One problem raised by this population-level twist on the dynamical interpretation is that it converges very strongly on the statistical interpretation. Both positions make a distinction between the statistical properties of populations and the outcomes they explain. On either view selection-the-effect is explained by the inequality of  $\Pr(p)$  and  $\Pr(q)$  and drift-the-effect is explained as a function of  $n$  (population size). The only apparent difference is that the dynamical interpretation applies the term ‘cause’ to the statistical parameters  $\langle \Pr(p), \Pr(q) \rangle$  and  $n$ , whereas the statistical interpretation withholds that particular honorific. The differences may be nothing more than terminological.

If this is a problem, though, it seems to afflict statisticalism rather more than its dynamical opponents. The statistical parameters that explain appear to qualify as genuine causes on most of the prominent and plausible criteria for individuating causes (Millstein 2006). They are (evidently) difference makers on which experimental intervention is possible (Reisman and Forber 2005). Changes in population structure are (allegedly) counterfactually dependent on them. They also count as causes on the chance-raising criteria of causation. Selection, the inequality of  $\Pr(p)$  and  $\Pr(q)$ , systematically raises the chances of a particular kind of outcome, that is, the preponderance of one trait over the other. Drift (sample size) systematically raises the chances of divergence from, or convergence on, the outcome expected given the values of  $\Pr(p)$  and  $\Pr(q)$  (Sober and Shapiro 2007). These parameters probabilify their respective outcomes and hence *explain* them. It seems perverse, then, for advocates of the statistical interpretation to deny that selection and drift are causes.<sup>6</sup> S2 must be mistaken.

**5. Statistics and Causes.** However, I think that there are good reasons to believe that S2 is true: the statistical parameters that explain selection and drift are not causes. My argument requires a substantive assumption about causation: *causal relations are description-independent*. By this I mean that

6. I am of the opinion that the criteria for being causes under the interventionist, counterfactual-dependence, and chance-raising accounts are being misapplied in these cases. But that claim must await another paper.

if  $x$  causes  $y$ , then this relation holds no matter how  $x$  and  $y$  are described. If so, the dynamical interpretation is committed to the following thesis:

DI. If selection and drift occur within a population, they do so *no matter how the population is described*.

DI, I think, captures a commitment of genuine importance for the dynamical interpretation.<sup>7</sup> It also identifies the most significant difference between the dynamical and statistical interpretations. The statistical interpretation is not committed to DI. If selection and drift are mere statistical effects, like the rank order effect, and the statistical parameters that explain them are not causes, there is no reason to suppose that DI must hold. One way to adjudicate the debate is to test DI.

*5.1. Drift as a Statistical Cause.* Suppose that we perform an experiment in which two fair coins are tossed 50 times each. The experiment is run as follows. Experimenter 1 flips coin 1 10 times. Simultaneously, experimenter 2 flips coin 2 10 times. Then each hands her coin to a new experimenter who in turn tosses it 10 times. Here again, the probability of heads and tails can be seen as the analogue of the fitnesses of two alleles at the same locus, and  $n$  is the analogue of the size of a population (or subpopulation). A simulation of this experiment is given in Table 1.

There are at least three different, equally legitimate ways to describe this experiment:

1. A single series of 100 tosses of a fair coin. All 100 tosses have the same probability of coming up heads or tails. The results for this experiment are 49 heads and 51 tails.
2. Two series of 50 tosses of a fair coin. There are two coins being tossed. It is reasonable to think of each coin as a treatment. In this case, the results are as follows:
  - a. series 1: 20 heads and 30 tails
  - b. series 2: 29 heads and 21 tails
3. 10 series of 10 tosses of a fair coin. There are 10 experiments. It is reasonable to consider each experimenter as a treatment. In this case the results are given in the bottom line of the table and summarized in Figure 1.

Allow sampling error to be the analogue of drift, as is customary. Observed error (drift-the-product) varies according to the way the experiment is characterized. In the population of 100 flips taken as a whole, there is only a slight departure from the expected outcome of 50 heads

7. Waters (1991) develops a similar notion in his account of the strong realism about the forces of evolution inherent in the traditional interpretation.

TABLE 1. DRIFT IN A POPULATION WITH NO SELECTION.

Expt. 1 Coin 1	Expt. 2 Coin 2	Expt. 3 Coin 1	Expt. 4 Coin 2	Expt. 5 Coin 1	Expt. 6 Coin 2	Expt. 7 Coin 1	Expt. 8 Coin 2	Expt. 9 Coin 1	Expt. 10 Coin 2
H	T	H	H	H	T	H	T	H	T
T	H	H	T	H	T	H	H	T	H
H	H	T	H	T	H	H	T	T	H
T	T	T	H	H	H	T	H	H	T
H	H	T	T	T	T	H	T	T	H
T	H	H	T	T	H	T	H	H	H
T	H	T	T	T	T	T	H	H	H
T	H	T	T	T	H	T	H	T	T
H	T	T	H	H	T	T	H	T	H
4/6	7/3	3/7	5/5	4/6	5/5	5/5	5/5	4/6	7/3

NOTE.—Experiment: 100 tosses of two fair coins, 50 tosses each, 10 experimenters, 10 tosses each of either one coin or the other. Randomization is generated by <http://www.random.org>.

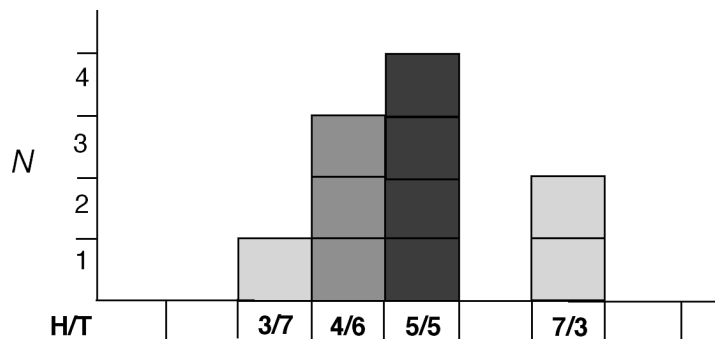


Figure 1. Results of coin tossing for 10 series of 10 tosses of a fair coin.

and 50 tails (49 : 51). There is a considerably greater degree of departure from expectation, however, in the population considered as two series of 50 flips (expected outcome 25 : 25; observed 20 : 30 and 31 : 29). It would be unlikely that we would find this degree of departure from an equal number of heads and tails in the population of 100 flips.<sup>8</sup> The degree of error observed in the population is greater still when we consider it as 10 series of 10 flips. Sixty percent of the series have results that diverge at least as far from an equal distribution of heads and tails as does the outcome for  $N = 50$ .

Error-the-(putative)-cause varies with each setup too, just as Stephens tells us. We can assess the relative strength of error-the-cause in each case by calculating the standard deviation, a measure of dispersion around the mean.<sup>9</sup> In a population of this kind, we expect that ~68% of all samples will fall within one standard deviation on either side of the mean. In a population of 100 flips of a fair coin, we expect that ~68% of all samples will fall within the interval between 45% heads and 55% heads. In a population of 50 flips, the standard deviation is slightly larger: ~68% of all samples will be expected to fall between 43% heads and 57% heads. For a population size of 10, the standard deviation is even larger: approximately 68% of all samples fall within an interval between 30% heads and 70% heads.<sup>10</sup>

8. Each of these outcomes is in itself reasonably unlikely for  $N = 50$ , as each falls more than one standard deviation from the mean.

9. I am assuming here that sufficiently large binomial distributions, such as these, approximate a normal distribution.

10. This quantity is slightly exaggerated for  $n = 10$  for two reasons. Binomially distributed populations of this size approximate the normal distribution less closely than populations of large  $n$ , and the standard deviation (1.6) has been rounded up to the nearest integer.

The dynamical interpretation takes standard deviation to be a measure of the causal power of drift. So it is committed to the claim that drift-the-cause is strong in the aggregate of 10 sequences of 10 tosses. This of course explains the large amount of observed error. In the single sequence of 100 tosses, however, drift is not very strong at all. This explains why drift-the-effect is minimal. But these are not two populations; *they are different ways of describing the same population*. Given DI, the dynamical interpretation is committed to the conclusion that, in this population, the force (or causal process) of drift exists and is both strong and not strong.

The traditionalist could avoid this contradiction by designating one of the descriptions as the canonical one. A natural choice might be to take this experiment to be *really* 10 series of 10 flips. This allows the traditionalist to preserve the intuition that drift really is strong, but that its effects, spread across the series of 10 samples, cancel one another out. The resultant of these 10 distinct error forces is small.

There are at least a couple of problems with this attempt to salvage the dynamical interpretation. One is that it is arbitrary. This is no more obviously an experiment comprising 10 series of 10 flips than it is one series of 100 flips, or for that matter 50 pairs of flips. No one description can lay claim to carving the experimental setup at its joints.

A more serious problem with this strategy is that in pursuing it the dynamical approach forfeits explanatory power. For example, if we describe the experiment as 10 series of 10 flips, it will be noted (as it was above) that the resultant force of drift in the population as a whole is small. This is no fluke, but the choice to designate this experiment as *really* 10 independent experiments of 10 flips each offers no explanation of this tendency for error effects to cancel one another out. Considering the experiment as a single series of 100 flips, however, gives us an explanation. Each of the 10 series of 10 flips is drawn at random from a larger, normally distributed population. The central limit theorem entails that the means of samples of the same size drawn from a normal distribution will themselves be normally distributed about the population mean (Zar 1974). So it is to be expected that the degree to which the outcomes in these subpopulations depart from the grand mean will tend to offset one another. By the same token, this does not privilege the description of this setup as a single series of 100 flips. Such a description would offer no account of why the errors in the subsamples of 10 flips are regularly so large. Privileging one description over the others will not work: each of these ways of describing the distribution is equally legitimate (or real) because each is explanatorily indispensable.

So the Stephens and Sober variant on the dynamical interpretation faces a dilemma with respect to drift. Either it must accept a contradiction—drift is objectively both strong and weak in the population—or it



must choose a canonical description and forfeit explanatory power. The statistical interpretation faces neither of these consequences. It accepts that drift explains an outcome only with respect to a statistical description. There are many distinct descriptions of this experimental setup. Each explains something that the others do not. According to some descriptions, drift is ‘strong’; according to others it is not. But there is no contradiction because, according to statisticalism, there is no description-independent fact of the matter about the occurrence of drift. Because the statistical approach can accept all legitimate descriptions of the experimental setup without incurring a contradiction, it does not forfeit explanatory power in the way that the dynamical approach must. The cost of this explanatory pluralism is a little metaphysical relativism (but just a little): DI is false of drift.

Like rank order effect, drift manifests the diagnostic features of a mere statistical effect: (i) it is explained by the statistical properties of a population, (ii) there is no description-transcendent fact of the matter whether it occurs, and (iii) drift explanations do not articulate the causes of population change.

5.2. *Selection as a Statistical Cause.* Similar considerations apply to selection. A variation on the previous experiment demonstrates this. For this experiment we allow the analogues of the selection coefficients ( $\Pr(p)$ ,  $\Pr(q)$ ) to vary randomly. Suppose that the two coins in our previous experiment are biased. Coin 1 has a probability of coming up heads of 0.6. Coin 2 has a probability of coming up heads of 0.4. Each of our 10 experimenters throws a coin 10 times, as in the first experiment, but in this version, each experimenter chooses a coin at random for each throw. On any given throw,  $\Pr(\text{coin 1}) = \Pr(\text{coin 2}) = 0.5$ . So overall in the population,  $\Pr(H) = \Pr(T) = 0.5$ . But within a series of 10 flips, the frequencies of coins 1 and 2, and hence the probabilities of H and T, may vary at random. Table 2 gives the outcome for one such experiment. Take the series designated ‘experiment 6’. Here the experimenter chose coin 1 twice and coin 2 eight times. Given that distribution, the expected relative frequency of heads in this subpopulation is 0.44. How should we characterize experiment 6 with respect to selection?

There seem to be two distinct ways, one in which there is *no* selection and one in which selection is strong. The ‘no selection’ characterization casts experiment 6 as simply the consequence of drawing a random sample from a population that, overall, is unbiased. This is a perfectly reasonable description. Significantly, the probabilities  $\Pr(H) = \Pr(T) = 0.5$  capture the way in which experiment 6 is identical to all the other series of 10 tosses. The ‘no selection’ description explains why we should expect a ratio of heads : tails (2 : 8) as extreme as this to be rare. This in turn

TABLE 2. RANDOMLY VARYING SELECTION COEFFICIENTS.

	Expt. 1	Expt. 2	Expt. 3	Expt. 4	Expt. 5	Expt. 6	Expt. 7	Expt. 8	Expt. 9	Expt. 10
	1	1	1	1	2	2	1	1	1	2
	1	1	2	1	1	2	1	2	2	1
	1	2	1	1	2	2	1	2	2	2
	2	2	2	2	1	2	1	2	1	2
	2	1	2	1	1	2	2	1	2	1
	2	2	2	1	2	2	1	2	1	1
	1	1	2	2	2	1	2	2	2	2
	2	1	2	1	2	1	2	1	1	2
	1	1	1	1	1	2	2	1	2	1
	2	1	2	1	2	2	1	1	1	2
Coin 1/2	5/5	7/3	3/7	8/2	4/6	2/8	6/4	5/5	5/5	4/6
Pr(H)	.5	.54	.46	.56	.48	.44	.52	.5	.5	.48
H/T	8/2	7/3	3/7	3/7	4/6	2/8	6/4	6/4	9/1	5/5

NOTE.—Two coins, 10 experimenters, 10 tosses each. Each experimenter chooses a coin at random then flips it, repeating the procedure up to 10 tosses. Biases: for coin 1,  $\Pr(H) = .6$ ; for coin 2,  $\Pr(H) = .4$ . Overall result: .53 H, .47 T. Randomization is generated by <http://www.random.org>.

explains why the results in experiment 6 are offset by the results of other experiments in the population as a whole: the subpopulation means are normally distributed around the grand mean. The ‘no selection’ account quite correctly attributes the result in experiment 6 wholly to drift.

The ‘strong selection’ characterization, on the other hand, emphasizes the bias toward tails *within* experiment 6. The fact that there is selection for tails ( $\Pr(H) = 0.44$ ;  $\Pr(T) = 0.56$ ) shows the actual outcome (H2 : T8) to have a higher likelihood than the ‘no selection’ description does. More to the point, these probabilities explain why if the very same series of flips was repeated a number of times the result would converge on a value more like 44% heads than 50% heads. This description also allows us to distinguish experiment 6 from experiment 1. In experiment 1, we find an outcome as strongly divergent from the predicted ratio of H : T (H8 : T2) as we do in experiment 6. But there is an important difference between them. Within experiment 1,  $\Pr(H) = 0.5$ . In contrast to experiment 6, if we were to repeat experiment 1 a large number of times, the outcome would converge toward 50% heads. The fact that there is selection within experiment 6 and no selection within experiment 1 explains the differences in their long-term prospects.

So we have two statistical descriptions of the probabilities in experiment 6. In one description  $\Pr(H) = 0.5$  and in the other  $\Pr(H) = 0.44$ . Neither description is dispensable. Each explains something the other cannot. This raises a problem for the dynamical interpretation. It holds that the ratio  $\Pr(H) : \Pr(T)$  constitutes a measure of the causal power of selection operating over experiment 6. So, if  $\Pr(H) = 0.5$  and  $\Pr(H) = 0.44$  are both acceptable descriptions, then by DI, the dynamical interpretation is committed to the conclusion that selection both does and does not occur in this experiment.

In the face of this contradiction, the dynamical interpretation must choose one statistical description over the other. It might, for instance, legitimately insist that  $\Pr(H) = 0.44$  captures what is really happening in experiment 6 and hence there is selection. But when the statistical descriptions are legislated in this way, two problems arise. The choice is arbitrary and, more important, it forfeits explanatory power. As we have seen, each description (i.e., the ‘no-selection’ and the ‘selection’ descriptions) explains some feature of this experiment that the other cannot. Again, the commitment to DI imposes a dilemma on the dynamical interpretation: it must accept a contradiction—selection both does and does not occur in experiment 6—*or* forfeit explanatory power.

The statistical approach faces no such problem, as it repudiates DI. In order to give a full explanation of experiment 6, we need to observe two distinct statistical distributions: one in which selection occurs (i.e.,  $\Pr(H) \neq \Pr(T)$ ) and one in which it does not (i.e.,  $\Pr(H) = \Pr(T)$ ). But,

as the statistical interpretation holds that there is no description-transcendent fact of the matter whether selection occurs in a population, it can accept the consistency of these descriptions. DI is false of selection too.

Like drift and rank order effect, then, selection bears the hallmarks of a mere statistical effect: (i) it is explained by the statistical properties of a population, (ii) there is no description-dependent fact of the matter whether selection occurs, and (iii) selection explanations do not advert to the causes of population change.

*5.3. Description Dependence.* The statistical interpretation is committed to the claim that selection and drift are mere statistical *descriptions* of population change. There is no objective, description-independent fact of the matter whether a population is undergoing selection, drift, or both. The description dependence of selection and drift has (perhaps) a salutary analogy in the phenomena of electromagnetism as they are treated by classical and relativistic physics.<sup>11</sup> In classical physics, electric and magnetic fields are considered to be distinct, objective, force-generating entities, just as, in the dynamical interpretation, drift and selection are distinct, objective change-generating processes. Yet, in classical physics, a body may appear to be experiencing different combinations of electric and magnetic forces depending on the frame of reference from which it is observed. “If the [electric and magnetic] fields are interpreted as real entities, then the question arises: Applied from which frames of reference do [we] describe the fields *as they really are?*” (Lange 2001, 181). In contrast, the relativistic treatment of electromagnetic phenomena faces no such issue: “Einstein’s theory reveals that . . . various combinations of electric and magnetic fields, as seen by different observers, are really the same object (the electromagnetic field) . . . . If a given force is electromagnetic, there is no fact of the matter whether it is caused by the electric field or the magnetic field” (175–176). The electric and the magnetic forces, we might say, are ‘mere perspectival effects’.

Analogously, when measured from different *statistical* perspectives, a population may appear to be experiencing different combinations of selection and drift for a given change in trait frequency. Experiment 6 in Table 2 illustrates this: the outcome is two heads and eight tails. Where the statistical perspective is that of the entire population— $\Pr(H) = \Pr(T) = 0.5$ —this subpopulation appears to be experiencing a significant amount of drift, but no selection. Whereas, from the perspective of the subpopulation taken in isolation— $\Pr(H) = 0.44$  and  $\Pr(T) = 0.56$ —the

11. I thank Sorin Bangu for drawing this analogy to my attention.

outcome of two heads and eight tails suggests that this subpopulation appears to be experiencing a combination of strong selection and moderate drift.

Like electric and magnetic forces, selection and drift are specifiable only relative to a 'frame of reference'. Changes in trait frequencies in a population—like the electromagnetic field—are objective and description-invariant, but the status of these changes as selection or drift (or both) varies from one statistical 'frame of reference' to another. There is no need to privilege a particular frame of reference and, as we have seen, plenty of reason not to. So, there is no objective fact of the matter how much selection or drift a population is experiencing. This is the sense in which selection and drift are description-dependent, mere statistical effects.

**6. Conclusion.** Selection and drift, according to the statistical interpretation, are part of the theoretical apparatus of the modern synthesis that seeks to explain changes in gene frequencies by appeal to the statistical properties of populations. While selection and drift *explain* population change, there is no reason to reify them as objective features of the biological world; we cannot treat them as “forces . . . [or causes] . . . that propel a population through a sequence of gene frequencies” (Sober 1984, 141). As far as hypostatizing selection and drift goes, statisticalism holds that “more is in vain when less will serve; for [statisticalism] is pleased with simplicity, and effects not the pomp of superfluous causes.”<sup>12</sup>

Let there be no misunderstanding: the statistical interpretation does not legislate against causal talk in evolutionary biology, much less the causal study of evolutionary processes.<sup>13</sup> Differences in individual propensities to survive and reproduce cause changes in the structure of populations, right enough. One may even apply the term 'natural selection' to this process, in keeping with Darwin's coinage. Much of biological research is legitimately engaged in the study of the individual-level causes of population change. Nevertheless, the statistical interpretation insists that in deploying the *modern synthesis* theory of evolution we are *not* engaged in the project of articulating causes. We are engaged in the project of explaining changes in population structure by appeal to statistical properties of populations only. As Marjorie Grene enjoins, “[we] must . . . distinguish between ‘genetical selection,’ which is purely statistical, and Darwinian selection which is environment-based and causal” (1961, 30).

12. With apologies to Newton ([1729] 1963, iii).

13. Brandon and Ramsey (2007), e.g., seem to labor under the misapprehension that the statistical interpretation is inimical to the causal study of evolutionary processes. Nothing could be further from the truth.

In reifying statistical selection and drift as causes of population change, the dynamical interpretation conflates the causal study of evolutionary processes with the statistical study of their effects. The statistical interpretation calls for a clearer demarcation of these two kinds of study. The causal study of evolution involves an investigation of those mechanisms that cause differential death, survival, and reproduction, and crucially those that secure the high fidelity of inheritance, and the capacity of individuals to produce, sustain, and pass on adaptively significant phenotypes. If the statistical interpretation is correct, the concepts of *selection* and *drift* embodied in the modern synthesis theory of evolution play no role in the causal study of evolution.<sup>14</sup> The statistical interpretation, in drawing the distinction between causal theories and statistical theories, calls for a reassessment of how the modern synthesis theory and Darwinian thinking fit together.

## REFERENCES

- Abrams, Marshall (2007), "How Do Natural Selection and Random Drift Interact?", *Philosophy of Science* (forthcoming).
- Ariew, André, and Richard C. Lewontin (2004), "The Confusions of Fitness", *British Journal for the Philosophy of Science* 55: 347–363.
- Beatty, John (1984), "Chance and Natural Selection", *Philosophy of Science* 51: 183–211.
- Bouchard, Frederic, and Alexander Rosenberg (2004), "Fitness, Probability and the Principles of Natural Selection", *British Journal for the Philosophy of Science* 55: 693–712.
- Brandon, Robert (2006), "The Principle of Drift: Biology's First Law", *Journal of Philosophy* 103: 319–336.
- Brandon, Robert, and Grant Ramsey (2007), "What's Wrong with the Emergentist Statistical Interpretation of Natural Selection and Random Drift", in Michael Ruse and David L. Hull (eds.), *The Cambridge Companion to Philosophy of Biology*. Cambridge: Cambridge University Press.
- Fisher, Ronald A. (1930), *The Genetical Theory of Natural Selection*. New York: Dover.
- Gillespie, James (1977), "Natural Selection for Variances in Offspring Number—a New Evolutionary Principle", *American Naturalist* 111: 1010–1014.
- Grene, Marjorie (1961), "Statistics and Selection", *British Journal for the Philosophy of Science* 12: 25–42.
- Hodge, M. J. S. (1987), "Natural Selection as a Causal, Empirical, and Probabilistic Theory", in L. Krüger (ed.), *The Probabilistic Revolution*. Cambridge, MA: MIT Press, 233–270.
- Lange, M. (2001), *Introduction to the Philosophy of Physics: Locality, Fields and Mass*. London: Routledge.
- Matthen, Mohan, and André Ariew (2002), "Two Ways of Thinking about Fitness and Natural Selection", *Journal of Philosophy* 119: 55–83.
- (2005), "How to Understand Causal Relations in Natural Selection: Reply to Rosenberg and Bouchard", *Biology and Philosophy* 20: 355–364.
- Millstein, Roberta (2002), "Are Random Drift and Natural Selection Conceptually Distinct?", *Biology and Philosophy* 17: 33–53.
- (2006), "Natural Selection as a Population-Level Causal Process", *British Journal for the Philosophy of Science* 57: 627–653.
- Newton, I. ([1729] 1963), *Sir Isaac Newton's Mathematical Principles of Natural Philosophy*:

14. The relation between the causal and statistical study of evolution is discussed in Walsh (2003).

- The System of the World*. Translated by A. Motte; revised by F. Cajori. Berkeley: University of California Press.
- Reisman, Kenneth, and Patrick Forber (2005), "Manipulation and the Causes of Evolution", *Philosophy of Science* 72: 1113–1123.
- Rosenberg, Alexander (2006), *Darwinian Reductionism, or How to Stop Worrying and Love Molecular Biology*. Chicago: University of Chicago Press.
- Rosenberg, Alexander, and Frederic Bouchard (2005), "Matthen and Ariew's Obituary for Fitness: Reports of Its Demise Have Been Greatly Exaggerated", *Biology and Philosophy* 20: 343–353.
- Shapiro, Lawrence A., and Elliott Sober (2007), "Epiphenomenalism—Some Do's and Don'ts", in G. Wolters and P. Machamer (eds), *Studies in Causality: Historical and Contemporary*. Pittsburgh: University of Pittsburgh Press.
- Sober, Elliott (1984), *The Nature of Selection*. Cambridge, MA: MIT Press.
- (2001), "Two Faces of Fitness", in R. Singh, D. Paul, C. Krimbas, and J. Beatty (eds.), *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*. Cambridge: Cambridge University Press, 309–321.
- Stephens, Christopher (2004), "Selection, Drift and the 'Forces' of Evolution", *Philosophy of Science* 71: 550–570.
- Walsh, Denis M. (2003), "Fit and Diversity: Explaining Adaptive Evolution", *Philosophy of Science* 70: 280–301.
- (2004), "Bookkeeping or Metaphysics? The Units of Selection Debate", *Synthese* 138: 337–361.
- Walsh, Denis, Tim Lewens, and André Ariew (2002), "The Trials of Life: Natural Selection and Drift", *Philosophy of Science* 69: 452–473.
- Waters, C. Kenneth (1991), "Tempered Realism about the Force of Selection", *Philosophy of Science* 58: 553–573.
- Woodward, J. (2003), *Making Things Happen*. Oxford: Oxford University Press.
- Wright, S. ([1931] 1986), "The Statistical Theory of Evolution", *Journal of the American Statistical Association* 22: 201–208. Reprinted in William Provine (ed.), *Evolution: Selected Papers*. Chicago: University of Chicago Press, 89–96.
- Zar, Jerrold (1974), *Biostatistical Analysis*. Englewood Cliffs, NJ: Prentice-Hall.