Drift and "Statistically Abstractive Explanation"*

Mohan Matthen†‡

A hitherto neglected form of explanation is explored, especially its role in population genetics. "Statistically abstractive explanation" (SA explanation) mandates the suppression of factors probabilistically relevant to an explanandum when these factors are extraneous to the theoretical project being pursued. When these factors are suppressed, the explanandum is rendered uncertain. But this uncertainty traces to the theoretically constrained character of SA explanation, not to any real indeterminacy. Random genetic drift is an artifact of such uncertainty, and it is therefore wrong to reify it as a cause of evolution or as a process in its own right.

1. Introduction. In this article, I outline what I claim is a legitimate, though hitherto neglected, form of explanation. I call it "statistically abstractive" (or SA) explanation. In SA explanation, certain factors that are probabilistically relevant to an outcome are excluded because these factors are extraneous to the theoretical project under way. (The suppression of these factors is what I call "statistical abstraction.") When relevant factors are excluded in this manner, the explanandum becomes predictively uncertain; that is, it has a probability less than one. But this uncertainty need not lie in any real indeterminacy: it traces, at least in part, to the deliberate suppression of relevant factors.

Though I shall sketch a fictitious example from economics for illustrative purposes (Section 3), my main motivation for introducing SA explanation (in Section 4) comes out of an attempt to understand the concept

*Received July 2009.

†To contact the author, please write to: Department of Philosophy, University of Toronto, 170 St. George St., Toronto, ON M5R 2M8, Canada; e-mail: mohan.matthen@utoronto.ca.

‡Thanks to Andre Ariew, Roberta Millstein, Michael Strevens, and (especially) Denis Walsh for patient discussion. Thanks also to Thomas Basbøll, Joseph Berkowitz, Margie Morrison, and Eric Schliesser for comments on parts of this essay and for sometimes trenchant criticism. Michael Dickson and two anonymous referees (one reporting to a journal that did not accept this piece) provided valuable advice.

Philosophy of Science, 76 (October 2009) pp. 464–487. 0031-8248/2009/7604-0001$10.00
Copyright 2009 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/648063 Published online by Cambridge University Press

of *random genetic drift* in population genetics. In Section 2, I set the scene by adumbrating some infelicities that arise from treating drift as a *cause* of evolutionary change. I concede that there may be other philosophical approaches to these infelicities; indeed, some prefer to think of them not as counterintuitive but as interesting peculiarities concerning a uniquely biological kind of process. However that might be, my aim here is to show how SA explanation throws light on drift without generating conceptual strain. My hope is that this will motivate other applications of this explanation schema.

2. Because-of-Drift Claims. Biologists sometimes make claims of the following sort:

BD. *X* occurred *because of drift* (and not because of natural selection).

The following quotation from Stephen Jay Gould contains a because-of-drift or BD claim: "When we turn to the species level, we find an interesting partnership among the three *causal forces* of drive, selection, and drift" (2002, 743; emphasis added). As we shall see, Gould means this claim robustly, for he suggests that drift operates more strongly at the level of species, and selection more so at the level of individuals. In this section, my aim is to show that taken literally, such claims sit poorly with intuitions about the compositionality of causes.

Drift is supposed to be connected to the uncertainty of evolutionary outcomes. But when a biologist makes a *because-of-natural-selection* (BNS) claim[1]—that is, that an outcome *Y* occurred because of natural selection (as opposed to drift)—she may well acknowledge that *Y* too was uncertain. How then does a BD situation differ from a BNS situation? In a classic article, John Beatty (1984) poses this problem well: "it is difficult to distinguish between random drift on the one hand, and the *improbable results of natural selection* on the other hand. Wherever there are fitness *distributions* associated with different types of organisms, there will be *ranges* of outcomes of natural selection—some of the outcomes within those ranges will be more probable than others, but all of the outcomes within the range are possible outcomes of natural selection" (196). Beatty's question: If *O* is a possible outcome of selection, when would you say that it is because-of-drift? I want here to pose a prior question: If *O* is a "possible outcome of natural selection," why is it

1. For a discussion of whether selection should be regarded as a cause, see Matthen and Ariew 2002, 2009.

https://doi.org/10.1086/648063 Published online by Cambridge University Press

necessary to invoke drift as a cause?[2] Does it make sense to regard drift as an additional cause in such a circumstance? Broadly speaking, BD claims are found in two kinds of contexts.

> First, drift is said to be a cause of an improbable evolutionary outcome. Consider, for example, a population P divided into two types, T and an alternative T'. Suppose that despite strong selection for T, T' winds up becoming fixed in P.[3] When such an improbable event takes place, it is sometimes said that it occurred because of drift. I shall call this the *improbabilist* application of drift.

Evolutionary biologists apply drift in this way when they suggest, for example, that improbable evolutionary changes occur in small populations, where (as we shall see in a moment) drift is supposed to be strong.

> Second, drift is said to be a cause of evolutionary outcomes in circumstances in which alternative outcomes are more or less equal in probability. Theorists will sometimes say that drift is operating in P if the probability of T becoming fixed in P is very close to the same as that of T' becoming fixed. I shall call this the *neutralist* application of drift.

The neutralist application most often occurs in the context of the claim that most molecular mutations have near-zero (incremental) adaptive value and that molecular evolution in such circumstances is not driven by variation in fitness.[4] Of course, it is possible to believe that drift is

2. It is likely that Beatty was not fully cognizant of the distinction, explained below, between drift-as-product and drift-as-cause. Note, however, that his term "outcome of natural selection" implies a distinction between natural selection itself and the distributions of types that may result from it.

3. Strictly speaking, it is alleles that get fixed, not traits. An allele becomes fixed, or goes to fixation, in a population relative to an alternative allele when its frequency rises to 100%, so that it is impossible (barring mutation) for the other allele to reemerge. I won't worry about this here.

4. Beatty (1984) makes a very similar distinction. One form of drift identified by him arises out of "indiscriminate sampling"—sampling in which "physical differences between organisms in one generation might be irrelevant to differences in their offspring contributions" (189). This echoes an equal-fitness or neutralist scenario. A second form is implicated when "outcomes are *less* representative of the physical abilities of . . . organisms to survive and reproduce in the environment in question" (196)—in other words, when the actual value departs from expected values given fitness differences. This corresponds to what I am calling the improbabilist scenario. (I am ignoring Beatty's typology of sampling processes here—parent sampling, gamete sampling, etc.—as they are not relevant to my present concerns.)

https://doi.org/10.1086/648063 Published online by Cambridge University Press

operative in both sorts of circumstances and in others as well.[5] What I want to show in this section, however, is that these two scenarios bring out certain peculiarities of BD claims.

Let us begin by examining improbabilist scenarios. Here, the centrally relevant phenomenon is that small sets of trials are *proportionately* more unpredictable than large sets. In any series of fair coin tosses, the expected proportion of heads is half the number of tosses. But in a short series, it is more probable that the proportion of heads will deviate from this expected value than in a long series. To put it in another way, if you specify a proportionate-error band around the expected value—say 25% of the total number of tosses—then the probability of the final proportion of heads falling within this error band increases as the number of tosses increases. The probability distribution gets more and more peaked around the expected proportion as the number of trials increases.

By analogy, imagine two populations of moths of variant coloring, all subject to predation by sharp-sighted birds in a forest predominantly consisting of dark-colored trees. The darker moths are better camouflaged and, hence, less subject to predation. Consequently, there is strong selection for dark coloring. Nonetheless, the lighter moths would increase in numbers if disproportionately many darker moths happened (for whatever reason) to sit on the few light-colored trees. Suppose that the only difference between the two populations is that one is large and the other is small. In the small population, the fixation of the dark color is less probable and that of the light color more probable than in the large. This is simply because the deviant outcome is more probable if there are just a few moths than if there are many. So biologists say that drift is greater in the small population: the unexpected outcome is more probable relative to the expected one. And if this unexpected outcome were to occur, they would say that it occurred "because of drift." (Beatty [1984, 196] chooses a comparative form of expression: "the *less* representative outcome is *more* a matter of random drift.")

What is the ontology of BD claims in improbabilist scenarios? Roberta Millstein (2002) has made an important clarificatory contribution to this discussion by distinguishing between drift-as-*product* and drift-as-*cause* (or as-process). When speaking of drift, biologists focus on *sampling error*, the difference $|E - A|$ between the *expected* outcome, E, of a set of trials

5. Robert Brandon (2006) suggests that drift is *always* present in natural selection. He is more or less correct if he means simply that natural selection *normally* allows for a range of possible outcomes (only more or less correct, since it does not *always* do so). It seems, however, that Brandon actually means to make a universal BD claim. (Recall that a BD claim offers drift as a cause of evolution.) The argument of this article is meant to cut against the latter interpretation of Brandon's claim.

https://doi.org/10.1086/648063 Published online by Cambridge University Press

and the *actual* outcome, *A*, as a proportion of the total number of trials. Sampling error is an *outcome* of the series of trials: let's call it drift-as-*product*. BD claims purport that drift is a *cause* of sampling error. Since it is a cause, it is *different* from its effect, that is, different from sampling error or drift-as-product. The sense of BD claims, then, is that there is something that causes sampling error. Call this something "drift-as-*cause*."

To illustrate drift-as-cause, consider Elliott Sober's claim that selection is a deterministic force (1984, 110) in the sense that *when it is acting alone*—that is, in infinite populations—future frequencies of traits are determined by their starting frequencies and fitness values. *Finite* populations may depart from expectation, and *small* populations afford "increasing scope" for such departures. Sober offers a causal analysis of this propensity of finite populations to stray. The shape of his analysis is prefigured by something he says earlier: "If genotype frequencies depart from Hardy-Weinberg equilibrium, some force must have been at work. For example, . . . 'sampling error' (*random genetic drift*)" (34).[6] Like expected trait frequencies under selection, the Hardy-Weinberg equilibrium is for Sober the deterministic outcome of an evolutionary process—deterministic when no other "force" is operating. In finite populations, he says, drift interferes with and opposes the drive to the deterministic outcome. This is why we get variant outcomes.

Now, Sober is clearly not saying just that genotype frequencies depart from expected values in small populations. He is positing a *cause* of this discrepancy. He makes this explicit in earlier work (Sober 1980, 370), where he contrasts two different causal analyses of variation within a population. In a "typological" analysis, groups of individuals of the same kind are impelled by the essence of their kind toward certain shared traits. However, because of certain *local disturbances*, these individuals do not manage to achieve these traits, but rather cluster around them. In a "populationist" analysis, by contrast, variation is fundamental: variation at one time is simply inherited from variation at an earlier time in accordance with measurable transition probabilities. Explaining variation does not, in this analysis, demand additional causes such as local disturbances.

In his discussion of selection and drift, Sober is dealing with the variability of trait distributions *across* possible populations (not *within* a population). Within the class of possible populations, there is variation in trait distributions because these distributions are only probabilistically predicted. Sober can be seen as taking a higher-order "typological," or metatypological, view with respect to this variation: all populations of the same kind (i.e., all those that display the same starting frequencies

6. Note that Sober here identifies sampling error as a "force." But, as noted earlier, sampling error is drift-as-product, not a "force" that produces this.

https://doi.org/10.1086/648063 Published online by Cambridge University Press

and fitness distributions) are driven by selection toward a predicted trait distribution. Infinite populations achieve the predicted result deterministically by means of the main force of selection, but local disturbances prevent some finite populations from achieving these distributions. These local disturbances constitute drift-as-cause.[7] A higher-order "populationist" view would take variation among possible populations to be fundamental; it would require no further cause to explain it. This, in effect, is the view I take in this article.

Following Sober, Christopher Stephens (2004, 557–558) suggests that drift-as-cause acts on populations, causing sampling error. The point that interests him is that there is, on this approach, a relatively greater impetus to the less probable outcomes in small populations. He attributes this to the greater strength of drift-as-cause in small populations; for him, drift is an equalizer that raises the probability of improbable outcomes. In short, drift interferes with selection, getting stronger as the population gets smaller. It is a force produced by small set size.

Now, notice an oddity of Stephens's argument. Suppose that in the large population of moths mentioned above, a freak accident occurs. Say a large gathering of dark moths is destroyed by fire, and as a result, the lighter moths get fixed in the population. This would be an improbabilist scenario, and so it could be said that drift brought about the result. But because of his reliance on the connection between small numbers and drift, Stephens would say that drift-as-cause was weaker in the large population. ("Drift plays a larger role in flipping a fair coin 10 times than it does in flipping a coin 10,000 times," he says [2004, 556].) So there is supposed to be less drift in the larger population to explain a larger departure from the expected value. Moreover, subpopulations of large populations, being smaller, are subject to stronger drift than the populations of which they are parts. Given that drift is nondirectional, this is odd. How can strong nondirectional (and therefore noncanceling) forces operating on parts of a population give rise to a weak force operating on the whole?[8] What kind of force is this?

It might be replied that while large populations tolerate larger *absolute*

7. Michael Strevens (2003, 10) correctly says that Sober's main-force/local-disturbance analysis is "in the spirit of [Adolphe] Quetelet." Ironically, Quetelet is a proponent of the first-order typological analysis. At this lower level, Sober appears to prefer the populationist way of thinking. So why does he take the typological path when dealing with variation among populations?

8. Another peculiarity of Stephens's account: Denis Walsh (2007) argues that since drift depends on population size and populations can be divided up in different ways, the subpopulation selection (and drift) values that result will depend on the mode of division. Walsh argues that this constitutes an intensionality incompatible with selection and drift being causes.

https://doi.org/10.1086/648063 Published online by Cambridge University Press

departures from expected values, they do not tolerate larger proportional variation. Drift, however, is thought of as bringing about *proportionate* sampling error, which is indeed smaller in large populations. This is correct. Nevertheless, the example highlights the sharp difference between physical forces and those alleged to be operating in drift. It takes more physical force to move 1% of a whale than it does to move 100% of a flea; evidently drift is not this kind of cause since, as we have just noted, a smaller amount of drift is supposed to effect larger absolute sampling error in large ensembles. Again: What kind of force, or cause, is it?

Gould (see the above quotation) also seems to be interested in the improbabilist conception. He argues that "when we turn to the species level," we encounter smaller population sizes than at the individual level and hence greater drift. Suppose that 10 different species of some plant genus occupy distinct geographical regions. There is relatively little drift at the level of individuals, Gould says, because the 10 species together might comprise millions of individual organisms—large numbers, small drift. But at the level of species there is quite a lot of drift because there are relatively few species (2002, 736). Thus, suppose that five of the 10 regions are randomly attacked by a parasite and that the five occupying species go extinct. Here, Gould says, we have species extinction due to drift. He means this as a BD claim. Because of the small number of species, the probability of differential extinction was significant.

Again, notice a strikingly odd implication of Gould's argument. His claim is that when you view the parasite as attacking a large collection of individuals, it would be improbable that five types out of 10 would be singled out for destruction, given that all the individuals are equally vulnerable to it. On the other hand, he wants to say, if you view the parasite as acting on a small collection of species, then it is more probable that it would destroy five out of 10 types, given drift in a small ensemble. Yet these are surely two different ways of characterizing the same occurrence: the parasite's attacking the species is nothing other than its attacking the individuals that constitute the species. Right away, we should be alert to an ontology of forces very different from that which we find in physics. What sort of "causal force" (Gould's words) is it that can act in this noncompositional (and intensional) manner? For in physics, forces that act on wholes are vector sums of those that act on parts. One could not, for example, have a force that acts on a planetary system differently than it does in sum on the individual planets.

Now, let us turn to neutralist scenarios. Here is Anya Plutynski's (2007) admirably clear account of how this application of drift emerged:

> In 1968, Motoo Kimura submitted a note to *Nature* entitled "Evolutionary Rate at the Molecular Level," in which he proposed what

https://doi.org/10.1086/648063 Published online by Cambridge University Press

> has since become known as the neutral theory of molecular evolution. This is the view that the majority of evolutionary changes at the molecular level are caused by random drift of selectively neutral or nearly neutral alleles. . . . According to Kimura, most changes at the *molecular level* from one generation to the next do not affect the fitness of organisms possessing them. (129)

It was a neutralist scenario that was in the minds of Jack King and Thomas Jukes (1969) when they wrote:

> Evolutionary change at the morphological, functional, and behavioral levels results from the process of natural selection, operating through adaptive changes in DNA. It does not necessarily follow that all, or most, evolutionary change in DNA is due to the action of Darwinian natural selection. There appears to be considerable latitude at the molecular level for random genetic changes that have no effect upon the fitness of the organism. Selectively neutral mutations, if they occur, become passively fixed as evolutionary changes through the action of random genetic drift. (788)

The contrast here is between an equal fitness scenario and one in which there are "adaptive changes."

Now, however close or far apart two alleles are to one another in fitness, one or the other will eventually go to fixation in a finite population. When allele A is much fitter than a, population size makes a difference to the chances of each getting fixed: in small populations, a has a reasonable chance of getting fixed despite its lower fitness; in large populations it does not. When the fitness of A gets closer to that of a, population size becomes increasingly irrelevant. In the limit, if the substitution of A for a makes no difference to fitness, then the probability of A becoming fixed is equal to that of a becoming fixed, regardless of the size of the population. (For a graphical treatment of this point, see Beatty 1984, 202–203.) Plutynski sums this up by saying that in neutralist situations, "if Kimura is to be believed, the effects of drift . . . are independent of population size" (2007, 138).

In neutralist scenarios, there is no privileged outcome—either allele can go to fixation. Thus, Sober's deterministic main force is reduced to zero, and drift is supposedly *acting on its own*. This ought to give us a good way of observing the character of the disturbing force in Sober's account—the force that disrupts the deterministic action of selection. However, as implied by the preceding paragraph, the strength of drift is intertwined with the strength of selection. When selection is strong, drift is weak but

https://doi.org/10.1086/648063 Published online by Cambridge University Press

strongly affected by the size of populations. When selection is weak, drift is strong but weakly affected by the size of populations.[9]

Again, consider now how long it takes for a particular trait to get fixed in a population, or rather consider a probability distribution over time-to-fixation measured in number of generations. If selection for a trait is strong, this probability distribution will be more peaked—time-to-fixation will be relatively predictable—and less dependent on population size. If selection is weak, time-to-fixation will be relatively unpredictable and more dependent on population size.[10] Such interdependencies undermine the idea that drift is something distinct from selection, as Sober's two-process model might suggest. Rather the two look like aspects of a single nexus.

3. An Economic Parable. The effects we have been considering occur in the context of probabilistic connections between cause and effect. In probabilistic causation, exactly the same cause will sometimes bring about one effect, sometimes another. No additional cause is needed to explain this variation: to say that *X probabilistically* causes *Y* is just to say that sometimes *Y* will not ensue from *X*. But how does probability enter the picture? I want now to argue that sometimes probabilities come not from the world, but from the theoretician's principled omission of certain relevant factors. This makes the "no additional cause" intuition just articulated even stronger: the theoretician's stance cannot very well create real-world "forces."

Imagine an economist interested in the price elasticity of demand. She finds that price changes have different effects on demand for different kinds of commodities. For necessities such as staple food items, price changes have very little effect; for small conveniences such as dry cleaning, the effect of price increases is linear; for luxury items with pretensions of exclusivity, the effect might be inverted—the higher the price, the greater the demand—and so on. The economist also discovers that hitherto overlooked transaction factors are relevant: elasticity takes on different values in business-to-business transactions, in online purchases, and so forth. Such additional determinants would interest our economist: though she may not have been aware of or interested in different categories of transactions until they turned up in her surveys, such factors are relevant to her question. She is interested in a complete economic account of price elasticity, not just in some preselected set of factors. Her discipline defines

9. It is, however, unclear (at least to me) why three variables are needed here: selection coefficient, size of population, and drift. The first two completely determine the value of the third.

10. I am indebted to Denis Walsh for the discussion of time-to-fixation.

https://doi.org/10.1086/648063 Published online by Cambridge University Press

the relevant features by its methodology; she does not define them by idiosyncratic interests.

Let us suppose that at a certain point in her investigation, the economist is satisfied that she has discovered all the economic determinants of elasticity—even those that she did not anticipate at first. But she has also stumbled on a *noneconomic* factor involved in some sales transactions. Say, for instance, that for reasons having to do with color perception, some women's willingness to tolerate a higher cost varies with the color in which the price of an item is displayed and with the kind of light shining on the price ticket. (Many women are more sensitive to color differences than men.) In different stores in which the price happens to be displayed differently, price increases affect demand differently on average because women are differently affected.[11]

I conjecture that the economist will not count this as an admissible factor. The effect arises situationally; it may vary simply with the prevailing light at different times of day and with the purely accidental use of different price tickets in different stores. Such accidental variation has nothing to do with the kinds of interaction that the economist is trying to understand. It affects elasticity, but to include its effects would distort the economist's understanding of the transactional influences she is investigating. Because these data are irrelevant from her point of view, the economist would average over the perceptual variations rather than build them into her model. Of course, she may come to change her mind about this: she, or economists in general, may come to think that perceptual oddities are actually of economic interest (just as they have recently come to believe that the psychological oddities of consumers are of economic interest). However this may be, the point that I wish to make here is that there are usually some factors of this sort in the "special sciences"—factors that change the probability of the explanandum but are deemed outside the concern of the investigator. I entitle the exclusion of such factors in special-science explanation *statistical abstraction*.

I think that what I have just said is defensible as an account of how an economist would proceed in such an instance, but there is one reason why intuition might tell against the appropriateness of excluding relevant factors in this manner. When one seeks to explain an individual event, one is normally looking for an account of everything that made a difference to this event, not simply an account within the bounds of a particular theory. Suppose that Sherry buys a tchotchke even after its price has risen but forgoes equivalently priced opera tickets when they go up

11. I am imagining that the difference in elasticity is wholly attributed to the perceptual difference: obviously, I do not mean even to imagine an economic rationality deficit among women! (The latter would be economically relevant, I take it.)

https://doi.org/10.1086/648063 Published online by Cambridge University Press

by the same amount. Suppose that one wants to know why Sherry behaved in this way. Now let it be the case that tchotchkes and opera tickets are indistinguishable from an economic point of view. Then economics has nothing to say about Sherry's behavior. But suppose that the price of tchotchkes is displayed in a way that happens to make women more likely to buy them even after the price has gone up. This changes the probability of a woman behaving as Sherry did. Shouldn't this figure in the explanation of Sherry's behavior?

This kind of example might suggest to some that SA explanation is incomplete. But I don't think that this is the right conclusion to draw. My claim would be that the *default context* of why-individual-event questions is theory unrestricted. Our economist is, however, theory restricted: she is not interested in Sherry as such; she is investigating price-demand patterns as economic phenomena. Though there are a number of different factors that are relevant to Sherry's behavior, the economist is justified in refusing to disaggregate those that are inadmissible in her model of elasticity. This kind of consideration will be relevant to our discussion of explanations in the theory of evolution later.

When I discuss drift in Section 5 below, I shall offer two possible reasons why certain factors are excluded in the context of the theory of evolution. In the present section, I have deliberately chosen a controversial and open-ended example in order to suggest that some factors may be intuitively outside the scope of certain kinds of explanation, even though it may not be clear why. I have no good account of why the economist should disregard perceptual factors, but assuming that I am correct in saying that she would do so, there is, doubtless, some such account. (Perhaps the reason is that the perceptual variation is a subliminal, hence extrarational, determinant of behavior.) Not all the issues that arise from my economic parable are relevant to random genetic drift or to other specific phenomena. What I want to emphasize here is that statistical abstraction by factor exclusion may be appropriate in a variety of situations for a variety of reasons.

Let me make one more comment about the scope of factor exclusion before I leave the example. Most accounts of explanation are *explanandum*-oriented. We ask what kinds of things make a difference to phenomena such as price elasticity. But there are contexts in which a theorist might take up an *explanans*-oriented perspective. Our fictional economist wants to know what difference economic factors make. Thus, she may not say "Here is a price increase: what caused it?" Instead, she might say "Here is my general conception of an economic factor: what will factors of this type explain concerning the price increase?" There may be instances in which this explanans-oriented economist will say "Look, I know that color will help me predict, control, and diagnose demand; nevertheless,

https://doi.org/10.1086/648063 Published online by Cambridge University Press

it doesn't provide me with the kind of understanding I am looking for. It may be relevant to a marketer, but not to me." This is how statistical abstraction works.

A moment ago, I mentioned "averaging over" perceptual variation. Here is what I mean. The economist is trying to predict demand: this is the outcome that interests her. Her model predicts how demand varies with price, given certain other characteristics of transactions. Let *C* be a type or class of transactions that are indistinguishable from an economic point of view. Every transaction in *C* is, in other words, the same as every other with respect to the factors that interest our economist. Within *C*, demand will vary, depending (in part) on how transactions happen to be influenced by the perceptual factors just mentioned. The economist ignores this variation in *C*: she simply lumps *C* transactions together and calculates elasticity for *C* as an average over different modes of display, men and women, and different lighting conditions. In this manner, the economist tolerates a source of variation in instances of *C*. Consequently, she predicts elasticity only probabilistically. Given an increase of price, there is a probability function associated with various values of demand for *C*.

Now, suppose that the perceptual factor happens to be unbiased: sometimes it decreases a woman's sensitivity to price; sometimes it increases it. Under such circumstances, proportional variation due to color of display might decrease as the number of transactions increases. In infinite sets of transactions, elasticity might act "deterministically"; at least this might be so if these sets accurately represent the range and probability of situations that occur in the real world. The point made at the start of this section is that the economist's decision to ignore certain factors—whether for bad reasons or good—cannot create a real-world cause responsible for the nondeterminacy of elasticity in finite sets of transactions.

4. Statistically Abstractive Explanation. It has been held that full explanations properly invoke probabilities relative only to a specification of all relevant initial conditions. Say the chances of a Toronto resident contracting influenza is .04. You might think that this explains why I got influenza. But if Toronto residents who work desk jobs have a .06 (or .01) chance, the explanation is incomplete: my working a desk job is relevant but has been left out. All probabilistically relevant factors must be included in an explanation of this type. Carl Hempel (1965, 397–400) called this the requirement of *maximal specificity*.

The example of the economist investigating price elasticity introduces a clarification that needs to be added to this notion of full explanations in probabilistic contexts. The economist's model deliberately disregards certain factors probabilistically (but not economically) relevant to the outcome. But from the point of view of economics, the explanation is still

https://doi.org/10.1086/648063 Published online by Cambridge University Press

complete. The reason is that the disregarded factors are, from this point of view, *ultra vires*.

Let us say that a *reference class H* is homogeneous in theory *T* (or *T*-homogeneous) for some outcome *O* if for every predicate *F intra vires* of *T*

$$\text{Probability } (O \text{ given } H) = \text{Probability } (O \text{ given } (H \text{ and } F)).$$

In other words, *H* is *T*-homogeneous for an outcome if *H* cannot be subdivided by *T-admissible* factors into subclasses in which the probability of the outcome is different. The probabilities cited in a full explanation in *T* must be relative to reference classes that are homogeneous in *T*. These probabilities arise from uncertainty that cannot be eliminated by citing additional *intra vires* factors.

Now, there are three different ways that a reference class might be homogeneous relative to a theory.

1. A reference class might be *T*-homogeneous because the phenomena are genuinely *undetermined*. A theory may, in other words, disallow subdivision because it recognizes that, objectively speaking, there are no subdividing predicates that make a probabilistic difference. It has been claimed that this is so for certain outcomes in quantum mechanics. Here, exactly the same initial conditions will result in different outcomes on different occasions, even though there is simply no additional factor that differentiates the occasions. Let us call these reference classes *objectively* homogeneous.

2. A reference class might be *T*-homogeneous because the phenomena are *messy*. When a coin is tossed, it spins in the air, and with each half turn a different face is upward. Here, the final result depends on three things: which face was up when the coin was flipped (F), the angular velocity of the coin (ω), and the total length of time in the air (t, which is determined by the height of the toss). Now, since the coin spins rapidly, the input values of F, ω, and t oscillate rapidly between heads-producing values and tails-producing values. Given a specific value of F, the graph representing possible values of ω and t is thus divided into alternating thin bands of heads outcomes and tails outcomes (Keller 1986). The class of coin tosses is not, therefore, objectively homogeneous: the outcome varies with exact values of F, ω, and t. Now, in coin toss gambling situations (as opposed to lab situations) the alternating heads and tails bands in ω-t space are much thinner than the limits of the gambler's knowledge of these input values. Given the limitations of knowledge, the probability one-half constitutes for the gambler a significant uniformity across all possible values of ω and t. (It would be different for a scientist who possessed an instrument to measure these very accurately.) Thus, although

https://doi.org/10.1086/648063 Published online by Cambridge University Press

objective homogeneity is not in play, there is good reason to use the value one-half and ignore variations in ω and t.

3. Finally, a reference class might be homogeneous because the differentiating condition F is *theoretically inadmissible*. The example from economics illustrates this. Here the reference classes display variability that is traced to the omission of theoretically inadmissible features.

I call an explanation "statistically abstractive"[12] if it invokes reference classes of the second or third kind (whether or not it also invokes reference classes of the first kind). Thus:

> **Definition**. A *statistically abstractive explanation* is one that omits some probabilistically relevant factors because these are theoretically inadmissible.

The second kind of homogeneity may overlap with the third kind; often, scientists employ statistical abstraction just because the phenomena are messy. In fact, both the reasons I shall offer for statistical abstraction in the theory of evolution are arguably of the second kind. I do want to emphasize, however, that it is sometimes legitimate to ignore *nonmessy* factors. Philip Kitcher (1984) has argued that there are many contexts in which it is misleading to add to classical genetics biochemical factors that are relevant to reproductive outcomes. If he is right, this is a good example of how it might be permissible to disregard a nonmessy relevant predicate. (See Section 5, point b below.) Robert Batterman (2002, 2) similarly argues that "scientific understanding often requires methods which eliminate detail, and in some sense, precision. . . . Detailed accounts simply provide explanatory 'noise'." This, again, is an argument for omitting factors that may not be messy.

5. Statistics and Natural Selection. Let us move now to a consideration of population genetics, the mathematical form of the theory of natural selection with the additional assumption that inheritance is Mendelian. Population genetics assumes that the traits of organisms are determined in a given environment by its genes. Traits interact with the environment to produce different reproduction rates for the genes responsible for them. Given frequencies of genes within a population at one time, population genetics seeks to predict frequencies at subsequent times. A gene frequency at a given time is an "evolutionary outcome."

12. In philosophy of science, the term 'statistical' has standardly been used for all explanations involving probabilities, even those in class 1 above (cf. Hempel's contrast between deductive-nomological and inductive-statistical explanation). Hence, the somewhat cumbersome term.

https://doi.org/10.1086/648063 Published online by Cambridge University Press

Explanations in population genetics are statistically abstractive. Here are some examples.

a. First, consider the predominance of dark-colored moths in highly polluted conditions. Darker specimens are better camouflaged, but it is possible for them to go extinct because a catastrophic event (such as a forest fire) wipes them out, while sparing the lighter-colored ones. Here, a condition relevant to the outcome is easy to specify—distance from the fire. But the theory of natural selection is uninterested in this condition.

b. It may be determinate, in any given process of meiosis, which of an offspring's genes will come from its male parent and which from its female parent. But the physical or chemical factors that determine this may be circumstantially determined in individual reproduction events, and quite *messy* in the sense outlined above. Thus, the theory might consider it a matter of chance which of an offspring's genes came from which parent, averaging over the different kinds of situations that determine this. In this context, "chance" once again arises from the fact that certain factors are being ignored.

c. Finally, consider that most unromantic idea, "random mating." The point here is not that mating is random in the sense that males as a whole constitute an objectively homogeneous reference class as far as any given female's mating choices go, and vice versa. Rather, the claim is that the basis for mate choice involves attributes that are *ultra vires* of the theory. For example, it might be conceded that individuals who live close to one another are more likely to mate, but if organisms or lineages are relatively mobile relative to other organisms or lineages, place of residence is not the kind of heritable factor that is considered relevant to the theory of natural selection. On the other hand, if organisms choose each other for a heritable attribute such as size, then mating is assortative (i.e., nonrandom).

Now, why are these factors ruled *ultra vires* of population genetics or theory of natural selection?

5.1. Microconstancy Exclusion. For one possibility, I return to the example of coin tosses. As discussed earlier, the coin alternates rapidly between a face-up and face-down position while in the air. Thus, though differences in an input condition—say angular velocity—make a difference to the outcome, there is a rapid alternation of outcomes in the interval between two values of this variable. Thus, for any large enough interval between values of the input variables, the probability of the coin coming up heads is one-half. And because the coin rotates rapidly, quite small is large enough.

Strevens (2003, Chapter 2) calls this condition *microconstancy*. Simpli-

https://doi.org/10.1086/648063 Published online by Cambridge University Press

fying his discussion considerably, a probability distribution is *microconstant* if the space of input conditions can be divided into "many small, contiguous regions" within each of which the probability distribution over outcomes is the same as over the whole space of input conditions (53). Clearly, the probability distribution of coin toss outcomes is microconstant in Strevens's sense. Since this is so, then *assuming that one can have only approximate information about input conditions*,[13] the probability of a given outcome (heads/tails) does not depend on input conditions. Given feasible (not exact) specification, input conditions constitute a contiguous region of a certain minimum size. Under these circumstances, the probability is a constant function of input conditions. The input conditions "wash out."

Strevens argues that this condition (together with an independence-of-microlevel-events condition that need not concern us here) ensures the possibility of macrolevel laws. Here is a biologically relevant example that he discusses.

> We have an ecosystem of rabbits. The rabbits do not reproduce, but they are occasionally eaten by foxes. What law governs the change in the rabbit population over time? . . . To trace the change in population, one might think, it is necessary to trace the life history of each rabbit in the population. . . . But suppose you have the following information: the probability of any rabbit's [surviving] over the course of a month is 0.95, and the deaths are stochastically independent. It is easy to calculate from these facts a probability distribution over the possible values for the rabbit population a month from now. The calculation is especially straightforward if the population is very large, for the law of large numbers implies that, with very high probability, the population in a month's time will be about 0.95 of the size that it is now. (2003, 14; misprint corrected)

Strevens's point is that given microconstancy (and stochastic independence), the probabilities are "simply passed up from the microlevel to the macrolevel." Given only the information that a rabbit belongs to this ecosystem, the probability of its survival is 95%. Further information is irrelevant to this probability, just as further information about angular velocity is irrelevant to the probability of a coin coming up heads.

How can this be? For example, wouldn't information about how the rabbits are distributed over the region relative to the foxes be relevant? Suppose that they are a bit closer, on average, to foxes at time t than at time t'. Wouldn't the probability of a rabbit surviving at t be lower than at t'? Strevens (2003, Chapter 4.4) argues that population genetics is at

13. This is my condition, not Strevens's, though it is implicit in some things he says.

https://doi.org/10.1086/648063 Published online by Cambridge University Press

least plausibly microconstant. His claim, in other words, is that information about the distribution of rabbits can be dispensed with because within the regions of input space defined by large enough variations of input variables, the whole-ecosystem probability is preserved.

At first sight, this strikes one as implausible. Consider the example of random mating above. Suppose that a city is divided into two ethnic enclaves and that marriage is more likely within these enclaves than across them. This will have an influence on evolution: for instance, if people in one enclave are congenitally larger on average than those in the other, the result will be a bimodal distribution of size, whereas if mating is random across the enclaves, a normal distribution will ensue in time. This kind of possibility seems to argue for the relevance of input conditions. However, if place of residence stays constant only for a small number of generations, this would be a temporary effect.[14] Consequently, the proximity/marriage effect may be disregarded, provided that evolution is thought of as stable change over time periods significantly in excess of n generations. In such circumstances, the input condition—geographical distribution of heritable size—is probabilistically relevant (over short time periods) but is disregarded because evolution is redefined as changes in gene frequency over a longer period of time.

Microconstancy implies that even though values of input conditions are probabilistically relevant to outcomes, *approximate values* of input conditions, or conditions defined over short periods of time, may be ignored altogether. Under approximate knowledge conditions, or conditions of long-term change, the probability over the entire range of input conditions may be treated as constant. Conditions that define tiny subranges are beyond the reach of the theory, even though they may be probabilistically relevant.

5.2. Metaconstancy Exclusion. I worried earlier that it is prima facie implausible that microconstancy holds for evolution. I emphasize the qualification "prima facie" since Strevens's argument in favor of this condition is deep. I find the argument surprisingly persuasive, but I don't want to completely rely on it here.

Beatty, however, offers us a much simpler alternative. Considering a case like that mentioned in point *a* above, Beatty says:

> The environmental circumstances relative to which fitness values are ascribed to members of a population are, ideally, all the environ-

14. What if the gametic separation of ethnic enclaves is maintained indefinitely? This is an example of "fencing" (Strevens 2003, 267–269), a condition under which microconstancy fails.

https://doi.org/10.1086/648063 Published online by Cambridge University Press

> mental circumstances relevant to determining differences between the reproductive successes of those organisms. Some of these factors discriminate between the organisms on the basis of fitness differences between them. For instance, the combination of a dark background and color-sensitive birds favors the reproductive success of dark moths over light moths, *because of the difference in their color*. Some other factors among the specified environmental circumstances may be responsible for differences in reproductive success, but *not* in connection with any fitness differences between the organisms in question. Forest fires, for instance, might kill and spare moths without regard to any fitness differences between them. (1984, 193)

Beatty's discussion suggests the following reflection. Consider all the heritable kinds into which members of a population may fall. We say that event type E is selection-wise *neutral* if the probability of a particular organism becoming involved in an event of type E is unaffected by its falling under any of these heritable kinds. For example, forest fires are selection-wise neutral because the probability of a moth becoming involved in a forest fire is unaffected by its color (and by other heritable characteristics). Dark moths may disproportionately become involved in forest fires, and this may affect the probability of their being selected. However, the probability of their being disproportionately involved in fires is the same as that of light moths being so. I'll call this *metaconstancy* condition. It is a reason for excluding information about the involvement of different types of organisms in neutral events.

Now, different kinds of investigation would render different factors relevant. A study of risky forest locations might well want to consider location. A real estate agent would not regard location as irrelevant. Genetic engineers might be interested in the conditions under which offspring inherit characteristics more from one parent than from another. It might well be, for instance, that certain chemical states of the vaginal tract could predict whether certain characteristics would be inherited from the male or female parent. But these factors may still be irrelevant to the theory of natural selection for exactly the same reasons that were given above. Relative to the outcomes studies by evolutionary theory, these factors might be microconstant or metaconstant.

This means that sampling error does not occur only where there is objective indeterminacy. Deviations from expected values in statistical sciences arise out of variation within reference classes constructed by suppressing differences that are of no interest to a particular theory, in this case, population genetics. They are not unexpected as such; they are unexpected only relative to the factors of interest to S. Since the uncertainty of evolutionary outcomes arises in this way, it makes no sense to

https://doi.org/10.1086/648063 Published online by Cambridge University Press

explain it by reference to a particular cause or process ("drift") or to a particular kind of cause or process.

Now, statistical explanations may have to be supplemented when individual events are the targets of explanation. A well-known example is that of the evolution of mammals in the late Cretaceous period. It is probable that this was brought about by a particular event that caused climate change and caused the extinction of dinosaurs. In this case, the scope of population genetics extends only to modeling the evolution of mammals given that their fitness changed after the event relative to dinosaurs. Obviously, though, the nature of the event that caused climate change is of considerable interest. As I said earlier, the default context of why-individual-event questions is theory unrestricted, and in such a context, we take into account all the probabilistically relevant factors. However, evolutionary theory suppresses some of these factors.[15]

The study of evolution is both a general and a historical study. Evolutionists want to know both why certain kinds of events take place and how certain individual events came to pass. While both kinds of investigation might invoke and ultimately be grounded in statistics, there are factors that are theoretically admissible in the theory-unrestricted investigation of individual occurrences or admissible in other studies and disciplines, though they are inadmissible in population genetics. Results that are unexpected when viewed as types from the point of view of the theory are not objectively unexpected.

6. Models of Selection. We have drawn some negative conclusions about the nature of probabilities in the theory of natural selection. Now we need to say something positive about how certain types emerge in natural selection.

Consider a finite population of asexually reproducing organisms. There are two types of organisms in this population, *A* and *B*. The two types are equally viable: the expected number of *A*-deaths, in proportion to the total number of *A*'s, is the same as the proportion of *B*-deaths to *B*'s. However, *A*'s reproduce slightly more than *B*'s—on average, *A*'s have slightly more offspring than *B*'s do. Reproduction occurs without mutation: *A*'s always produce *A*'s and *B*'s *B*'s. Finally, population size is constant: for every birth, there is a death. What will happen?

Here is one way to model the dynamics of such a situation. Imagine a

15. Thus Sober (1980, 370): "Darwin and Galton focused on the population as a unit of organization. The population is an entity, subject to its own forces, and obeying its own laws. The details concerning the individuals who are parts of this whole are pretty much irrelevant. Describing a single individual is as theoretically peripheral to a populationist as describing the motion of a single molecule is the kinetic theory of gases."

https://doi.org/10.1086/648063 Published online by Cambridge University Press

series of time steps. In each step, two events occur. First, two fair coins are tossed. The first toss is used to pick a removal: if heads (H), an *A* is removed; if tails (T), a *B*. The second toss models a reproduction event: if H, an *A* is added to the population; if T, a *B*. Further, if (and only if) this second toss was H, a six-sided die is rolled—if it comes up 6, then an additional *A* is added to the population, and another organism is removed in accordance with the first toss. We assume that this set of events reflects probabilities relative to admissible factors.

In each time step, there are four possible outcomes:

- in half of the kinds of trials (HH or TT), the number of each type stays the same;[16]
- in one-quarter (HT), *B* gains over *A* by 2;
- in five cases out of 24 (TH, 1–5), *A* gains over *B* by 2;
- in one case out of 24 (TH, 6), *A* gains over *B* by 3.

Thus, averaging over 24-trial cycles, *A* gains by 1 over *B* each cycle.

This kind of model, slightly modified from what is known as a Moran process (Moran 1958; Nowak 2006, Chapter 6), allows us to figure out the probability of a mutant *A*-type going to fixation without further mutation: we let the *B*-types initially constitute $N - 1$ individuals in a population of N individuals and work out the probabilities from there. It turns out that where r is the fitness of *A*-types relative to that of *B* and N is large, this probability is $(1 - 1/r)$. It can be seen that given the advantage enjoyed by *A*-types, a mutant *A* has an approximately 4% chance of becoming fixed in a large population (Nowak 2006, 102). In small populations, its chances are somewhat larger.

Moran processes and other models use series of random events such as coin tosses, die rolls, and blind picks from urns to model natural selection. The "randomness" of the individual trials reflects the suppression of micro details responsible for outcomes and the use only of probabilities of various different outcomes. The series consists of trials that are unconnected with one another. This lack of connection illustrates how evolution by natural selection occurs by a simple accumulation of births and deaths. Each time step in the above process brings two or three independently determined events; the "process" has a cumulative impact on gene frequencies in the population *by virtue of simple arithmetic*. There is evidently no unitary or connected process involved here—nothing like the trajectory of a particle moving under the influence of gravity. Rather,

16. In the (HH, 6) case, an *A* is removed in accordance with the first H, and then two *A*'s are added because of the second H and the subsequent 6. At this point, an *A* is removed because of the first H. So the net gain is zero.

https://doi.org/10.1086/648063 Published online by Cambridge University Press

changes of gene frequencies in the population take place as probabilistically independent births and deaths add up.

I want to make three points about Moran processes.

i) What is drift in such a model? Notice first that each trial in a Moran process is uncertain. But the randomness of these events represents the uncertainty that results from suppressing certain factors. It is not a reflection of any real "disturbance" or "force."

ii) Natural selection is a statistical trend in the accumulation of trials in a Moran process. Such statistical trends admit of a variety of actual outcomes. In an infinite series of trials, the outcome is certain; however, in a finite series of trials, there is a spread of possible outcomes, each with its own probability. Drift is simply the uncertainty associated with this spread of possibilities. Natural selection and drift are not distinct processes working on a finite population, but mathematically connected aspects of the same accumulation. There is nothing in this accumulation that we would be entitled to call drift-as-cause, or natural-selection-as-cause for that matter.

iii) Biased Moran processes are no different in principle from unbiased ones. Consider the Moran process that consists only of the toss of two coins—no roll of the six-sided die. This is an unbiased process, but there is no difference with regard to the point just made in point ii—predictability in infinite sequences, uncertainty in finite sequences. Thus, if the biased Moran process models natural selection, you would think that the unbiased Moran process models something similar. Considering a model very similar to the unbiased Moran process, Beatty (1984, 192) says, disapprovingly: "if this is a proper explication of selection, then the problem of the relative evolutionary importance of random drift vs natural selection is a pseudo-problem—there is no difference between them." I agree that on this kind of explication, there is no difference between natural selection and drift *as causes*. And this is indeed the conclusion I draw. It does not follow that the debate about neutralism is a pseudo-problem: it is a debate about whether selection was strong or weak.

7. The Abstractness of the Theory. The theory of natural selection is abstract. It is a mathematical description of the accumulation of reproduction events, and as such it is applicable to domains that have very different causal structures. This can be illustrated as follows.

Consider a set S that consists of two completely diverse kinds of organisms, drawn at random from two different populations—say 50 moths from a population M of 100 moths and 50 rats R from a population of 100 rats. The set S is constructed simply by designating certain organisms as members, not by removing them from M or R. The organisms themselves stay where they are.

https://doi.org/10.1086/648063 Published online by Cambridge University Press

Now let us say that the moths come in two colors, *D* and *L*, the former of which offers better camouflage in their environment, and the rats have two kinds of digestive system, *E* and *M*, the former of which is better at absorbing a particular kind of grain that is relatively abundant in their environment. Stipulate that *D* and *L* are wholly determined by homologous alleles, and similarly *E* and *M*. Let us suppose that the fitness advantage enjoyed by the well-camouflaged moth is the same as that enjoyed by the rat with better digestion.

Now define kinds *F* and *G* as follows:

> *x* is *F* if and only if *x* is a moth in *S* and *x* is *D* OR *x* is a rat in *S* and *x* has *E*.
>
> *x* is *G* if and only if *x* is a moth in *S* and *x* is *L* OR *x* is a rat in *S* and *x* has *M*.

Many of the characteristic phenomena of population genetics will obtain just as much in the causally bisected set of rats and moths as they do with respect to camouflage outcomes in the causally unified population of moths taken by themselves, or digestion outcomes in that of rats. For instance, since *F*-organisms tend to reproduce in greater numbers than *G*-organisms, the former will tend to grow in numbers relative to the latter. This shows that the causal connectedness of populations is incidental with regard to many natural selection phenomena. The arithmetical accumulation of births and deaths is what matters. Neither the statements nor the proofs of theorems such as Fisher's Fundamental Theorem and others mention causal connectedness as a necessary condition of the validity of their results. Fisher's theorem does not describe a phenomenon that takes place exclusively in connected populations. Drift does not either.

Larry Shapiro and Elliott Sober (forthcoming) urge a causal interpretation of natural selection on the grounds that "no biologist would treat two individuals as part of the same (token) selection process if they were at opposite ends of the universe" (quoted by Matthew Haug [2007, 437n], who adds: "The reason no biologist would do so is that such individuals are not part of the same biological population").

Shapiro, Sober, and Haug are undoubtedly right. Many evolutionary phenomena are dependent on the causal connectedness of populations. Consider, for example, the fixation conditions for causally connected populations. Suppose we get to a point where there are no *L*-allele bearing moths left in *M*. Then, this gene cannot reappear (barring a mutation). Suppose, however, that *L* disappears from *S*, the bisected ensemble whose members are drawn from distinct populations. The *L*-allele can reappear here, provided that it is available in *M*. In causally connected populations,

https://doi.org/10.1086/648063 Published online by Cambridge University Press

there is gene flow to all parts, and this makes a difference to what kinds of individual events are possible.[17]

Haug (2007, 443) observes that on the statistical interpretation, population genetics "leaves out interactions between mice, predators, and their environment." He is right: the abstract theorems of population genetics omit the nitty-gritty of biological reality. And this nitty-gritty is indeed relevant to evolutionary biology. This is not what I wish to contest. My point is simply that the increase of fitter varieties, and the uncertainty attending that increase, resides in the statistical treatment of births and deaths. We need not seek for a cause of that uncertainty. The cumulative toting up of proportions is merely bookkeeping (cf. Sterelny and Kitcher 1988). By my account, God the Creator is simply the accountant in charge.

8. Conclusion. The theory of natural selection is, I have argued, a statistical theory that disregards certain factors probabilistically relevant to the outcomes that it is interested in explaining. This generates uncertainty with respect to the outcome, which accumulates over the multiple births and deaths that constitute natural selection. There is no process that accounts for this uncertainty. Since the uncertainty lies not in the events but in the theoretician, none is needed.

Note Added in Proof. "Drift is a term designating a set of *physical processes*, arguably, indiscriminate sampling processes," write Roberta Millstein, Robert Skipper, and Michael Dietrich (my emphasis) in a paper published online after the above was copyedited ("(Mis)Interpreting Mathematical Models: Drift as a Physical Process", *Philosophy and Theory in Biology* 1, http://hdl.handle.net/2027/spo.6959004.0001.002). Obviously, their view is radically at odds with mine, for they view drift as physical, whereas I take it to be an artifact of the theoretician's acts of statistical abstraction. Now, even where there is strong selection for a trait, that trait can disappear. I take it that Millstein et al. agree that this uncertainty constitutes drift-as-outcome. What is responsible for this outcome? According to Millstein et al., natural selection is discriminate sampling, whereas drift is indiscriminate sampling. To accommodate drift-in-selection, then, Millstein et al. would have to say that selection processes are accompanied, in finite populations, by certain physical processes—indiscriminate sampling processes. So (1) Millstein et al. are committed to a version of the Quetelet-Sober main-force/disturbing-influences model. I have argued in the main text that drift is intertwined with selection in a way that is unfriendly to this model's assumptions. But (2) since there is

17. The gene-flow condition makes a difference to species and speciation too (see Ereshefsky and Matthen 2005).

https://doi.org/10.1086/648063 Published online by Cambridge University Press

no drift-in-selection in infinite populations, the indiscriminate physical processes they posit would be absent. In other words, they are committed to infinite models being physically different from finite models of the very same selection pressures. This seems implausible to me.

REFERENCES

Batterman, Robert (2002), *The Devil in the Details*. New York: Oxford University Press.

Beatty, John (1984), "Chance and Natural Selection", *Philosophy of Science* 51: 183–211.

Brandon, Robert (2006), "The Principle of Drift: Biology's First Law", *Journal of Philosophy* 103: 319–335.

Ereshefsky, Marc, and Mohan Matthen (2005), "Taxonomy, Polymorphism, and History: An Introduction to Population Structure Theory", *Philosophy of Science* 72: 1–21.

Gould, Stephen Jay (2002), *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press.

Haug, Matthew C. (2007), "Of Mice and Metaphysics: Natural Selection and Realized Population-Level Properties", *Philosophy of Science* 74: 431–451.

Hempel, Carl G. (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press, Macmillan.

Keller, Joseph B. (1986), "The Probability of Heads", *American Mathematical Monthly* 93: 191–197.

King, Jack Lester, and Thomas H. Jukes (1969), "Non-Darwinian Evolution", *Science* 164: 788–798.

Kitcher, Philip (1984), "1953 and All That: A Tale of Two Sciences", *Philosophical Review* 93: 335–373.

Matthen, Mohan, and André Ariew (2002), "Two Ways of Thinking about Fitness and Natural Selection", *Journal of Philosophy* 99: 55–83.

——— (2009), "Selection and Causation", *Philosophy of Science* 76, forthcoming.

Millstein, Roberta (2002), "Are Random Drift and Natural Selection Conceptually Distinct?", *Biology and Philosophy* 17: 33–53.

Moran, P. A. P. (1958), "Random Processes in Genetics", *Proceedings of the Cambridge Philosophical Society* 54: 60–71.

Nowak, Martin A. (2006), *Evolutionary Dynamics: Exploring the Equations of Life*. Cambridge, MA: Harvard University Press.

Plutynski, Anya (2007), "Neutralism", in Mohan Matthen and Christopher Stephens (eds.), *Handbook of the Philosophy of Science*, vol. 3, *Philosophy of Biology*. Amsterdam: Elsevier, 129–140.

Shapiro, Larry, and Elliott Sober (forthcoming), "Epiphenomenalism—the Do's and the Don'ts", in G. Wolters and P. Machamer (eds.), *Studies in Causality: Historical and Contemporary*. Pittsburgh: University of Pittsburgh Press.

Sober, Elliott (1980), "Evolution, Population Thinking, and Essentialism", *Philosophy of Science* 47: 350–383.

——— (1984), *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Cambridge, MA: Bradford Books, MIT Press.

Stephens, Christopher (2004), "Selection, Drift, and the 'Forces' of Evolution", *Philosophy of Science* 71: 550–570.

Sterelny, Kim, and Philip Kitcher (1988), "The Return of the Gene", *Journal of Philosophy* 85: 339–361.

Strevens, Michael (2003), *Bigger than Chaos: Understanding Complexity through Probability*. Cambridge, MA: Harvard University Press.

Walsh, D. M. (2007), "The Pomp of Superfluous Causes: The Interpretation of Evolutionary Theory", *Philosophy of Science* 74: 281–303.

https://doi.org/10.1086/648063 Published online by Cambridge University Press