The Difference Between Selection and Drift:
A Reply to Millstein

ROBERT N. BRANDON

Departments of Biology and Philosophy, Duke University, Durham, NC 27708, USA (e-mail: rbrandon@duke.edu)

Abstract. Millstein [Bio. Philos. 17 (2002) 33] correctly identifies a serious problem with the view that natural selection and random drift are not conceptually distinct. She offers a solution to this problem purely in terms of differences between the *processes* of selection and drift. I show that this solution does not work, that it leaves the vast majority of real biological cases uncategorized. However, I do think there is a solution to the problem she raises, and I offer it here. My solution depends on solving the biological analogue of the reference class problem in probability theory and on the reality of individual fitnesses.

Introduction

Millstein (2002) correctly identifies a serious problem with the view that natural selection and random drift are not conceptually distinct (a view seemingly defended by Beatty 1984). The problem is that one of the major disputes in 20th, and now 21st, century evolutionary biology – the selectionist/neutralist debate – is incoherent if that view is correct.[1] Biologists do make mistakes and do, sometimes, pursue blind alleys, but Millstein is loath to say that this huge area of work is all based on a conceptual mistake. I completely agree. But then it is incumbent on one to clearly draw this distinction. That is the primary goal of her article. She thinks that once we distinguish between process and outcome, that the *processes* of natural selection and drift can be sharply distinguished, even though the *outcomes* may not be clearly distinguishable. Unfortunately, her solution to this very real problem does not work. She identifies another "solution" to this problem; one that she thinks fares less well than hers and that she attributes to Brandon and Carson (1996). I agree with her that this "solution", which treats the theory of evolution by natural selection as deterministic, is unattractive for a number of reasons. Fortunately

---

[1]Also a considerable amount of theoretical work showing that the evolution of certain traits requires the interaction of selection and drift would be misguided. See, e.g., Rausher and Englander (1987) on the evolution of evolutionarily stable strategies for habitat selection under soft selection, or Eshel and Feldman (1982) on sex-ratio evolution.

for me, I do not now hold, nor have I ever held, a position remotely close to that attributed to me (unless 180° away counts as close). More fortunately still, there is a real solution to the problem.

Exegetical matters

The primary point of Brandon and Carson (1996) is that the process of evolution by natural selection is autonomously indeterministic, i.e., indeterministic in a way that follows directly from the theory of evolution by natural selection.[2] That is a strong conclusion and the reader of this discussion is not asked to accept it in total.[3] The argument for it is complicated, but the form of the argument is familiar to philosophers of science. It is the standard argument for the reality of theoretical entities. When the positing of such entities plays a crucial role in the development of a body of a predictive and explanatory theory, and when all available experimental evidence supports the posit, then (if, of course, you are a realist) you should accept the reality of the posited entity. The better confirmed the theory, the stronger this argument.

Our conclusion, one that we state very clearly, is that the probabilistic propensities attributed by the theory of natural selection are real. In short, fitness values are real propensities, not useful instruments. So our conclusion is not about the *theory* of evolution by natural selection, but rather about the *process*. However the major premise of this argument is that the *theory*, which is explanatorily and predictively successful, contains, in an essential way, these probabilistic propensities. So if the theory of natural selection has at its foundation these probabilistic propensities in what sense could it be deterministic?

In a paragraph from which Millstein quotes we answer that question. "Biologists often describe natural selection as a *deterministic* phenomenon. By this they mean its effect is directional (i.e., has a predictable direction) as opposed to drift. They do not mean that it is deterministic in the philosophers sense." (Brandon and Carson 1996, p. 324). That last sentence is followed by a reference to Sober (1984, pp. 110–115), which I will discuss shortly, but first let me elaborate on the last sentence. As philosophers use the term, a *theory* is deterministic if from a complete *state description* of a system at time t one can derive a complete state description of that system at some later time t'. (One could complicate this characterization enormously, but this suffices for present

[2]This sort of autonomy does not commit us the truth, or even the sensibility, of the following counterfactual: If quantum mechanics were deterministic then the evolutionary process would still be indeterministic. I for one have no idea how to evaluate the truth of that statement. Rather we meant autonomy in the sense of Hacking (1990).

[3]The negative points made here do not depend on that conclusion at all. The positive ones do depend on a part of that conclusion, i.e., that the probabilistic propensities assigned to individuals are real. The autonomy of those probability values is irrelevant to the points made here.

purposes.) A *process* is deterministic then if there is such a theory that truly describes it. No biologist I know thinks of natural selection as deterministic in this sense. When they do use the term 'deterministic' to describe natural selection they do so to contrast natural selection with random drift. Both processes can result in changes in gene and genotype frequencies across generational time, but our best theories of these processes differ in that theories of selection can predict the direction, as well as rate, of cross-generational change, while our theories of drift can predict only the rate, but not the direction, of such change. That, and only that, is what is meant by "deterministic" natural selection in contrast to drift.

I really do not think we could have been much clearer in the aforementioned paragraph, but evidently our reference to Sober mislead Millstein. We referred to Sober simply to give the reader a more extended discussion of the point we were making quickly. Upon rereading Sober, I'm not sure how he could have been clearer either. Millstein quotes the following sentence from Sober (a sentence we do not quote – we simply refer to pp. 110–115): "*When it acts alone, the future frequencies of traits in a population are logically implied by their starting frequencies and the fitness values of the various genotypes.*" (Sober 1984, p. 110, italics in the original; quoted by Millstein, p. 48). However, reading only two paragraphs further one finds that Sober thinks this to be false. He says, referring to the above claim, "We can now correct the characterization offered above of why natural selection is a deterministic evolutionary force. Fitness values plus starting frequencies do *not* permit the deduction of changes in gene frequencies, even on the supposition of infinite population size. They do, however, permit a probability inference of almost unbeatable strength ..." (italics in the original, Sober 1984, p. 111). Of course, when we move to more realistic models with finite population size, this probabilistic inference loses strength (how much depends on the population size and the selection differentials). Not only are such models more realistic, they are much more interesting when one is concerned with the relationship between drift and selection, as Carson and I certainly were.

The theory of natural selection is not to be identified with any one of its many models. But given that each such model is probabilistic, and given that I have for a long time defended the view that the very foundation of the theory of evolution by natural selection is the explicitly probabilistic Principle of Natural Selection,[4] it seems highly dubious that I would think that the theory is deterministic. That doubt is well placed. I don't.

---

[4] The Principle of Natural Selection, as I have explicated it, is an instance of the Principle of Direct Inference from probability theory. I state it as follows: If *a* is better adapted than *b* in environment *E*, then (probably) *a* will have more offspring than *b* in *E*. The notion of relative adaptedness contained in this principle is defined in terms of objective probabilities of different levels of reproductive success. I argue that these probabilities should be given a propensity interpretation. The final definition of adaptedness (or expected fitness) is then the expected (in the mathematical sense) reproductive success discounted by a function of the variance, the exact form of the function depending on the nature of the variance. See Brandon 1990, pp. 9–24.

Process and outcome

The key to Millstein's approach to differentiating selection from drift is the distinction between *process* and *outcome*. In general such a distinction is useful. In this case, can we, as Millstein claims, distinguish between the processes of drift and selection without regard to the outcomes? As I will show, we cannot, at least not in a way that allows these two concepts to play their normal role in evolutionary theory.

Drift, as Millstein recognizes, can take many forms. She focuses primarily on parental sampling; that is, on the process by which organisms from one generation are "sampled" to be the parents of the next generation. For ease of exposition I will follow her in this. Focusing on this particular form of drift will result in no loss of generality; and in fact the major points I wish to make can be illustrated initially without delving into biology at all.

According to Millstein, drift as a process can be characterized as an *indiscriminate* sampling process. The contrast, of course, is with a *discriminate* sampling process. And the difference between discriminate and indiscriminate sampling processes is itself easy to understand. A sampling process is indiscriminate if and only if each entity in the pool to be sampled has an equal probability of being chosen. For example, if we are pulling balls from an urn, that process is an indiscriminate sampling process if and only if each ball in the urn has an equal probability of being pulled. In the biological case, imagine a population of 1000 individuals, only 100 of which will get to mate and reproduce. If each of the 1000 individuals has an equal probability of becoming a parent then that process is indiscriminate sampling. Less abstractly, if there are no physical differences that make the difference between those 100 who are successful versus the remaining 900 who are not, then this parental sampling is indiscriminate.[5]

In contrast, suppose that there are physical differences that make a difference with respect to the probability of becoming a parent. Then the sampling is discriminate and, according to Millstein, is a case of natural selection. But notice that all that is required to differentiate these *processes*, discriminate versus indiscriminate sampling, is the difference between *processes* where each possible result is equiprobable versus those where they are not. Unfortunately that distinction does not map well onto the ways biologists differentiate drift from selection.

Again consider sampling balls from an urn. Suppose that there are 100 balls in the urn, 50 red and 50 black. Further suppose that each ball has an equal probability of being chosen on any one draw ($p = 0.01$). (The equal

---

[5]This way of putting the point suggests a propensity interpretation of probability, where the physical properties of the chance set-up determine the long-run frequencies of the outcomes. This is the interpretation I favor. But the points made here are compatible with a limit-relative-frequency interpretation.

number of red and black together with the equal probability of each ball being chosen results in the p(red) = p(black) = 0.5.) A sampling here consists in pulling a single ball from the urn, recording its color, and then replacing the ball. Let us now imagine a trial of four such samplings. The most probable outcome, the one that will occur most often in a long sequence of such trials, i.e., the *mode*, is 2 red and 2 black. That outcome is also the *expected* or *mean* outcome, i.e., the one that corresponds to the overall frequency of the two types in the urn.[6] But there are four other possible outcomes: all red; all black; 3 red, 1 black; and 1 red, 3 black. Here are the probabilities of each of these possible outcomes:

  Example 1

 p(all red) = 0.0625

 p (3 red, 1 black) = 0.25

 p (2 red, 2 black) = 0.375 p (1 red, 3 black) = 0.25

 p (all black) = 0.0625

Notice that while the expected outcome is indeed the modal outcome, it occurs in slightly fewer than 4 out of 10 trials, or conversely, outcomes that deviate from the expectation occur in slightly more than 6 out of 10 trials.

Now let us change the setup so as to change the probabilities of the results (red or black). There are two ways to do this. First we could change the relative frequency of the two colors in the urn, e.g., if we replace five of the black balls with five red, then we would get: p(red) = 0.55 and p(black) = 0.45. But in this case the sampling would still be indiscriminate, i.e., every ball would still have the same probability of being drawn. To make the sampling discriminate we need to change the probabilities of individual balls being drawn so that they are no longer equal. So let us repaint the balls using slightly sticky red paint and slightly less sticky black paint, but retaining the equal numbers of red and black. Let us suppose that this change in the physical characteristics of the two types of balls results in the following probabilities: each red ball now has p = 0.011 of being drawn and each black ball has p = 0.009 (where before both types had p = 0.01). Now the process of sampling, i.e., of drawing balls from the urn, is discriminate. The probability of red is now 50×0.011 = 0.55 and the probability of black is 50×0.009 = 0.45. Let us again sample (with replacement) four balls. Again there are five possible outcomes and we can again calculate the probability of each:

---

[6]The expected outcome equals the overall frequency of the two types in the urn only under the assumption of equiprobability. When we discard that assumption, as we will, the expected outcome equals the sum of the products of the probabilities of the two types times their frequencies. Although we are here talking about the simple case of drawing balls from an urn, that expected value is just two complications away from my definition of expected fitness. See footnote 2 above. (The two complications are: (1) In biology reproduction is not all or nothing, there are different levels of reproductive success; and (2) The expectation needs to be discounted by the variance.)

Example 2
p (all red) = 0.091506
p (3 red, 1 black) = 0.299475
p (2 red, 2 black) = 0.367538
p (1 red, 3 black) = 0.200475
p (all black) = 0.041006

Notice that the outcome, 2 red, 2 black, is still the most likely outcome, but its probability has been lowered as has the difference between it and the next most likely outcome. Also notice the *expected* outcome, 55% red and 45% black, is of course not possible with a sample size of four.[7]

One more example: here we make the red paint stickier still so that the resulting probability for each red ball getting picked is 0.015 and the probability for each black ball is 0.005. Again there are 50 red and 50 black balls. Now the probability of red is 0.75 and of black 0.25. Sampling four balls we get the following distribution of probabilities for the five possible results:

Example 3
p (all red) = 0.316406
p (3 red, 1 black) = 0.421875
p (2 red, 2 black) = 0.210938
p (1 red, 3 black) = 0.046875
p (all black) = 0.003906

In this case, as in the first example, the expected result (75% red, 25% black) is also the modal result.

What we have done in these three examples is start with the one distribution of probabilities that corresponds to indiscriminate sampling, namely the equiprobable distribution. The next two examples move away from that distribution towards increasing differences between the probabilities for red and for black. Probability theory allows for an infinite number of distributions of the probabilities for each of the 100 balls being drawn. The only constraint is that these 100 probability values sum to 1. But if we continue moving in the direction of our examples, there is a distribution of maximal probability difference. We will discuss it shortly, but first let us ask the question: has the change in our experimental setup, a change from indiscriminate sampling to discriminate sampling, resulted in a *qualitative* change with respect to drift?

We cannot satisfactorily answer that question until we jettison Millstein's approach to defining drift in a way that makes no reference to outcomes. I will argue explicitly for this move presently, but for now let us, tentatively, adopt an outcome-oriented conception of drift, namely, that *drift is any deviation from the expected result due to sampling error*.[8] I label this approach "outcome-oriented" to differentiate it from Millstein's, but notice that it does refer to a

---

[7]Brandon and Carson (1996) discuss examples such as this where drift is forced to occur. I will not stress that point here.

[8]This is the characterization of drift that Brandon and Carson (1996) adopt. It is standard in the biological literature, see e.g., Roughgarden 1979, chapter 5.

process, viz. sampling, and so here drift is defined in terms of both process and outcome. If that is what drift is, then the difference between Example 1 and Examples 2 and 3 is quantitative, not qualitative. Consider Examples 1 and 3. In Example 1 the probability of the expected result is 0.375. That means that in a long sequence of such trials, approximately 37.5% would yield exactly the expected result while 62.5% would yield a result that deviated, more or less, from the expectation. In Example 3 the expected result will occur more frequently in a long series of trials, approximately 42% of the time. Both setups will regularly lead to results that deviate from the expectation; the difference between them is quantitative, not qualitative.

Let us be more explicit about the analogy between drawing balls from an urn and biology. Getting drawn from the urn corresponds to parental sampling, i.e., becoming one of the organisms that reproduces. The equiprobable distribution represents the case where all organisms in a population have equal fitness. Differences in probabilities represent selection differentials. So we have just illustrated the well-known result from biological studies of drift: *everything else being equal, the greater the selection differentials the smaller the expected effect of drift*. Conversely, the smaller the selection differentials the greater the effect of drift (everything else being equal). But from an evolutionary point of view nothing changes much when we move from very small selection differentials to absolute neutrality. Indeed biologists use the term *effectively neutral* for alleles or traits where the selection differentials are so small relative to population size that drift is expected to dominate. Thus the *qualitative* distinction Millstein marks between indiscriminate sampling and discriminate sampling does not map onto a qualitative distinction in evolutionary processes.

With respect to drift the equiprobable distribution does not stand out among the infinite number of possible distributions. However another sort of distribution, what I will call distributions of *maximal probability difference* (MPD for short), does. This set of distributions is qualitatively distinct with respect to drift. In our simple urn case, where there are 100 balls in the urn, the MPD occurs when one of the balls has a probability of 1 of being drawn, and all other balls have a probability of 0. Thus here there are 100 such distributions. Given the nature of our setup, only one ball could have a probability $= 1$ of being drawn. One can imagine different setups. Suppose we were to simultaneously grab four balls out of the urn. Then a distribution where four particular balls have a probability of 1 of being drawn while the other 96 have a probability of 0 is an MPD distribution. In general we get the maximal probability difference when all the probabilities equal either 0 or 1, and at least some equal 0 and some 1.[9] In the biological case a maximal fitness value, which

---

[9]For any setup of the sort we are considering there will be exactly one equiprobable probability distribution. In our simplest urn drawing model there are exactly 100 distributions of maximal probability difference, different setups will have different combinatorial possibilities, but as long as the number of objects is finite there will only be a finite number of MPD distributions. The import of this will be discussed later.

is an expected number of offspring, say 15.75, is normalized to 1. So the biological analogue of MPD is where all the fitnesses in a population equal either 1 or 0, again with at least some of each value. Call this a population with *maximal fitness difference*.

Distributions of MPD are qualitatively distinct with respect to drift. Drift *cannot* occur with such a distribution. Balls are sampled from an urn, or organisms are sampled from a population, but with all probabilities equaling either 0 or 1 there can be no sampling error, no deviation from the expected result. The expected result occurs with probability 1. Notice that it follows from what was said above about drift, that with the MPD distribution, the expected effect of drift is minimized. That is true, but under-informative in that it masks the qualitative difference between such distributions and the infinite number of other possible probability distributions. With an MPD distribution, drift is not just highly unlikely, it is impossible.[10]

What about natural selection? If it is to be distinguished from drift, then we cannot identify it with the process of discriminate sampling, since, as we have seen, that process does not differ qualitatively from indiscriminate sampling with respect to drift (except in the extreme case of MPD). To get anywhere in our exploration of drift we were forced to an outcome based characterization of drift. A similar move will be needed for natural selection.

I have characterized natural selection as follows: *natural selection is differential reproduction that is due to differential adaptedness (or fitness) to a common selective environment* (see, e.g., Brandon 1990, chapters 1 and 2). This identifies an outcome – differential reproduction – but it has an explicitly causal component to it as well. Natural selection is not just any case of differential reproduction, but is those cases that are *due* to differential adaptedness (or fitness). The whole point of the propensity interpretation of fitness, or adaptedness (Brandon 1978; Mills and Beatty 1979), at least as I have developed it, is to provide the conceptual machinery adequate for an explanatory theory of natural selection. Among other things this requires being able to distinguish those cases of differential reproduction due to differential adaptedness from cases of differential reproduction that are drift. This, in essence, is the so-called "tautology problem".

With that conception of natural selection in mind let us return to the continuum of probability distributions discussed above. As we saw, the equiprobable distribution allows for drift (it does not make drift necessary). But selection *cannot* occur under such a regime, because selection requires differential fitness or adaptedness – in our urn model that translates into probability differences. So although the equiprobable distribution is not qualitatively distinct from the infinity of other possible distributions with respect to drift; it is qualitatively distinct with respect to selection. Selection requires fitness

---

[10]This is in contrast to the case of infinite population size, where drift is *highly* unlikely, but not impossible. See discussion in Section 1 above.

differences, so this is the single probability distribution that precludes selection. Put in other words, with indiscriminate sampling, selection cannot occur.

On the other hand, the set of MPD distributions forces selection to occur, while, as we saw above, precluding drift. Selection must occur with an MPD distribution because, to use our simple model, balls with a probability of 1 of being drawn will be drawn, and those with a probability of 0 will not be, otherwise those probability values are not truly 1 and 0. Just to make the analogy explicit: organisms with a fitness of 0 will not reproduce. Organisms with a fitness of 1 will. This is quite clear if we make the simplifying assumption that reproduction is all or nothing, i.e., an organism either reproduces exactly n offspring or 0 offspring. Then the urn model fits exactly. But when we allow for the possibility of producing different numbers of offspring, as real organism are wont to do, then two organisms could achieve the same expected number of offspring in different ways. For example, type A could always have 2 offspring, while type B could sometimes have 1 and sometimes 3, but with a mean of 2. Type C has 0 fitness. So, if you have read Gillespie (1977), you will know that selection can occur here between A and B (with A being favored) even though this seems to be the analogue of an MPD distribution.[11] But, in fact it is not. When appropriately discounted, the fitness of B is lower than A, and so this is not an MPD-type distribution.

Notice the asymmetry between drift and natural selection vis-à-vis the two poles of our probability distribution continuum. At the equiprobable end *drift is possible and selection is impossible*. At the MPD end *selection is necessary (not just possible) and drift is impossible* (Figure 1). It is worth noting that the modalities just attributed to drift and selection do not depend on population size. The probability of drift and selection does so depend, but not the possibility or impossibility.

Putting that slight asymmetry aside, could we use these two poles to make a process-oriented distinction between selection and drift? That is, could we equate the process of drift with indiscriminate sampling and selection with *maximally* discriminate sampling? No. From a mathematical point of view, the equiprobable distribution and the finite set of MPD distributions represent an infinitesimally small fraction of the possible probability distributions. Were we to categorize sampling with an equiprobable distribution as drift and sampling with an MPD distribution as selection, we would be leaving all but an infinitesimally small fraction of cases uncategorized (Figure 1). Biologically things are probably worse still for this sort of strategy. Has any real biological population in the history of life on Earth ever realized one of these two extremes? I will not pretend to know the answer to that question, but I would not be surprised if it were no. Certainly it is safe to say that no population ever studied has met these conditions. Which means that the biological action is in the distributions that fall between the equiprobable and

---

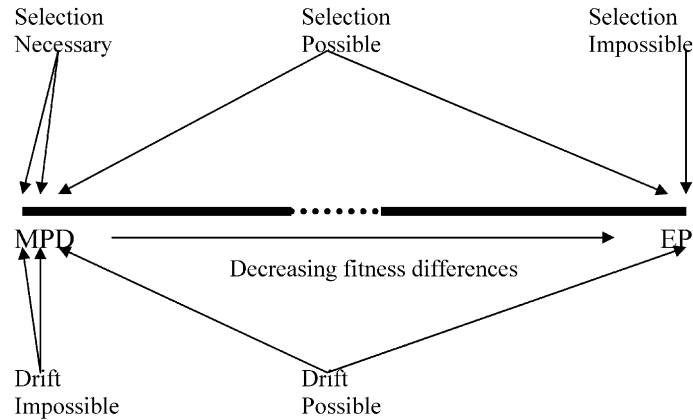[11]For further discussion of this see Brandon 1990, pp. 18–22.

Figure 1. The heavy horizontal line, with dotted centre section, represents the infinite number of possible fitness distributions from maximal probability differences (MPD – all fitness = 0 or = 1, with some of both) on the left to the equiprobable distribution (EP – all fitness the same) on the right. The arrows emanating from the different descriptions of the modalities of selection and drift indicate the areas of the distribution falling under these descriptions.

the MPD. It is here that we need to be able to distinguish drift from selection. And we can.

Moths, reference classes and selective environments

Every day moths get eaten by birds. Now I'm as fond of moths as the next person, but generally this fact about moths does not upset me. However, in an otherwise insightful article published almost 20 years ago Beatty (1984) unleashed into the philosophical literature a population of moths that confused him and that seem to have caused considerable confusion since. It is time to clarify this confusion and to let these poor moths rest in peace.

Though fictional, Beatty's moths are modeled on Kettlewell's well-known studies (Kettlewell 1955, 1956). They come in two forms – light and dark – and inhabit a forest that contains light and dark trees in a ratio of 40:60. It is clear that Beatty assumes, though he never explicitly states this, that neither form of the moth behaviorally discriminates between light and dark trees. In other words, the moths land at random on trees and so have a 40% chance of landing on a light tree and a 60% chance of landing on a dark tree. Birds then prey on the moths based on their conspicuousness relative to their background. Beatty then assumes that the fitness distributions of the two types overlap in that environment (195).

Now Beatty asks us to suppose that, *by chance*, more of the dark moths happen to land on light trees than would be expected, but not vice versa, i.e., the light moths behaved more or less as expected. As a result the frequency of

the light moths increases in the next generation, contrary to expectation. What are we to make of this? Beatty says:

> Is the change in frequency of genes and genotypes in question a matter of natural selection, or a matter of random drift? That is, is the change in question the result of sampling discriminately or indiscriminately with regard to fitness differences? It is not easy to maintain that the sampling was entirely indiscriminate with regard to differences in survival and reproductive ability. At least it is difficult to maintain that the death by predation of *conspicuously* dark moths in this environment is indiscriminate sampling, whereas the death of *conspicuously* light moths in the same environment is selection. On the other hand, it is also difficult to maintain that selection alone is the basis of the change. At least, it is difficult to maintain that the fittest were selected. (pp. 195–196)

Beatty concludes with the following:

> In other words, it seems that we must say of some evolutionary changes that they are to some extent, or in some sense, a matter of natural selection *and* to some extent, or in some sense, a matter of random drift. And the reason (one of the reasons) we must say this is that it is conceptually difficult to distinguish natural selection from random drift, especially where *improbable results of natural selection* are concerned. (p. 196)

This statement is so qualified that one might think it would be hard to disagree with. But as Millstein correctly notes, were we to accept it we would be forced to say that the large amount of work that has gone into the selectionist–neutralist debate is based on a conceptual mistake. Millstein's solution, as we have seen, is to say that outcomes don't matter and that so long as the sampling was discriminate then this is selection. And so she unambiguously classifies Beatty's case as selection. We have also seen that her approach is a non-starter.

Notice that Beatty's problem would not have arisen had the woods been 100% dark or 100% light. The problem arose because dark moths by chance landed on light trees more often than they should have, and this problem could not have arisen had the environment been homogeneous with respect to background color. This matters because fitness is obviously relative to an environment and Beatty is unsure as to which environment he should relativize these unlucky moth's fitness. Should it be the environment that is characterized by the statistical distribution of the colors of the trees in the woods (60% dark, 40% light), or the environment that the moth happened on just prior to its demise (light)? This is a genuine biological problem and one of the utmost importance to the theory of evolution by natural selection. But before sketching its solution, let me point to its analogue in probability theory.

The dominant objectivist interpretation of probability during the 20th century was the limit-relative-frequency interpretation. According to it, the probability of attribute A just is the limit of the relative frequency of A in an infinite series of trials. But which infinite series of trials? The problem becomes particularly acute when we try to assign a probability to a single case. For instance, suppose we are about to flip a particular US quarter and want to know the probability of its landing heads on that particular flip. Unfortunately that particular flip belongs to an indefinite number of potentially infinite sequences of events, e.g., flips of any sort of coin, flips of any sort of US coin, flips of US quarters, flips that occur on Tuesdays, flips that start with head side up, and so on. To which sequence do we assign it? That is the *reference class problem*. If we could unambiguously assign it to some particular sequence, then we could just assign the limit of the relative frequency of heads in that sequence as the probability in that particular case. But if there is no objective way of assigning the event to one particular sequence, then it seems that we cannot give an objective probability to the single case.

One might think that the propensity interpretation of probability fares better with respect to the problem of assigning a probability to a single case since according to it, a probability just is some set of physical features of an object (e.g., coin) in a particular *chance setup* (e.g., tossing devise). But Weslely Salmon has argued (e.g., Salmon 1970, pp. 38–40) that the reference class problem arises just as acutely when we try to specify the nature of the chance setup.[12] Does any toss of any coin count as part of our chance set up? Or just tosses of US quarters? Or, etc.? A Pyrhric victory could be achieved by saying that it is just this single toss that counts as our chance setup. But then we could never compile reliable statistics to validate any claim about the probability.

Hans Reichenbach, perhaps the leading exponent of the frequency interpretation, thought the reference class problem lacked a fully adequate solution. The best he could do was to recommend that we adopt as the reference class "the narrowest class for which reliable statistics can be compiled" (Reichenbach 1949, p. 374). But this is a pragmatic recommendation that seeks to maximize two variables – narrowness and reliability – that are at odds with each other. The narrower our class the less reliable our statistics, and the more reliable we seek to make our statistics the less narrow our reference class. Thus in the end Reichenbach refused to use the word 'probability' to apply to single events, instead in such cases he used the word 'weight', and argued that 'probability' applied literally only to sequences (Salmon 1970, p. 41).

Salmon, a student of Reichenbach, succeeded where Reichenbach failed. To solve the reference class problem he developed the concept of *homogeneity*. A class is homogeneous with respect to some attribute (or outcome) A if and only

---

[12]I don't completely agree with Salmon. I think that the problem is less serious for the propensity interpretation. However, Salmon is surely right that the propensity interpretation still faces the reference class problem.

if there is no *place selection* of the class that is statistically relevant to *A*. The concept of *place selection* comes from von Mises, and is any partition of the reference class *not* made in terms of the attribute (or outcome) in question (Salmon 1970, pp. 42–43). In other words, we can partition a sequence of tosses of a coin any way we like (e.g., tosses of this quarter versus some other coin; tosses on Tuesday versus on other days; tosses that start with the head up versus those that start tail up, etc.) just so long as we don't do so in terms of the outcome itself. Thus tosses that yield heads versus tosses that yield tails is not a place selection and is not allowable. (Obviously that partition is statistically relevant, but it is cheating.) Salmon then adopts the following *reference class rule*: "choose the broadest homogeneous reference class to which the single event belongs" (Salmon 1970, p. 43).

Before moving on let me note one important point. The notion of homogeneity as defined above is an ontological concept not an epistemic one. One can, and Salmon does, define *epistemic homogeneity* – a class is epistemically homogeneous with respect to attribute *A* if and only if there is no *known* place selection of it that is statistically relevant to *A* (Salmon 1970, p. 44). But we are here interested in objective homogeneity.

I have treated the biological problem that Beatty raises elsewhere (see Brandon 1990, chapter 2 and Antonovics et al. 1988), and so will be brief here. When dealing with natural selection the relevant notion of the environment is what I've termed the *selective environment*. Intuitively, the selective environment is the arena within which selection occurs. It is measured in terms of the relative actualized reproductive success of two or more competing types. What it is defined as, i.e., what is being measured, is the relative fitnesses (expected reproductive success) of two or more competing types across space, or time, or some other suitable dimension. For the sake of concreteness, consider a spatial scale. A region of space is selectively homogeneous with respect to types A_1, A_2, ..., A_n if and only if the relative fitnesses of types A_1, A_2, ..., A_n are constant within that region. In other words, a region of space is selectively homogeneous if and only if it cannot be partitioned by means of a place selection in a way that is statistically relevant to the relative fitness of the competing types.

A region of space is *heterogeneous* if it is not homogeneous. In other words, a spatial gradient, say a transect up a mountainside, is selectively heterogeneous if and only if the relative fitness of the competing types changes along the transect. Empirically this would be indicated by a genotype-environment [$G \times E$] interaction. We have lots of evidence from ecological genetics that selective environments are indeed heterogeneous. But, of course, the scale of such heterogeneity depends on all sorts of ecological factors; it may be a matter of millimeters (e.g., in the annual plant *Erigeron annuus*[13]) or of hundreds of kilometers.

---

[13]See Straddon and Bennington (1998).

What about Beatty's moths? The case is fictional and Beatty does not supply enough information to answer that question. But we can fill in details in ways that would allow us to address the question. First, we can describe the case where the whole wooded area in question is selectively homogeneous. That may sound counterintuitive. Surely, one might think, a light colored tree is not part of the same selective environment as a dark tree. The "relative fitness" of light and dark moths differs dramatically on these two trees. But this objection mistakenly treats fitness as an instantaneous property of an organism rather than as a property of the organism's life history.[14] Here is where behavior becomes crucial. Plants stand still and wait to be counted; moths do not. So it is quite conceivable that in some particular populations of *Erigeron* one spot in a field is in one selective environment, while 5 mm away is a different selective environment. That is not conceivable for moths. Whatever the spatial scale of selective heterogeneity for moths it is going to be considerably greater than that of a plant like *Erigeron*. Moths fly around and land on many different trees. Their probability of being devoured by a bird depends on the match, or lack thereof, of their color and the statistical average color of the background that they create by their behavior. Thus if the two tree types are distributed randomly about the woods and both types of moths show no behavioral preference for one type of tree over the other, then the woods in question are selectively homogeneous. This is fully consistent with Beatty's story. Then this is a case where both selection and drift occurred. (We cannot quantify the effect of each since the story is not quantitative.)

On the other hand we could certainly fill out the story in other ways so that the woods are selectively heterogeneous. Suppose the size of the woods is considerably greater than the average size of the territory a moth typically inhabits. Suppose further that the two tree types are not distributed randomly about the woods but rather are patchy, with the ratio of light to dark differing significantly among patches. Finally suppose that the size of these patches is on the order of, or greater than, the size of typical moth habitats. Then the woods are selectively heterogeneous, and to fully understand the evolutionary dynamics of this population one would need a model of what I have termed *compound selection*. Compound selection involves selection *within* selectively homogeneous environments and distribution *among* such environments (see Brandon 1990, pp. 71–77). But in any realistic scenario there will again be both selection and drift.

The details of Beatty's story are under-specified, and the appropriate conceptual tools are not utilized, so it is impossible to say anything definitive about it. Let those moths rest in peace.

---

[14]Biologists regularly measure the relative fitnesses of organisms during one part of their life cycle, and as we make that period shorter and shorter we would converge on instantaneous fitness. Such a measure is sufficient for understanding the ecological process of selection that occurs during that part of the life cycle. But that measure corresponds to evolutionary fitness only under the assumption that selection occurs *only* during that part of the life cycle.

Process and outcome redux

In the urn case there is a process of sampling balls from the urn. In the biological case there is a process of sampling from among the members of one generation to get the parents of the next. In both cases the sampling may be governed by the single equiprobable distribution, in which case it is indiscriminate, or, much more likely, it may be discriminate. Ignoring MPD distributions, in either case the result of the sampling may, or may not, be the expected one. That is, there is a single process, sampling. In any given case, whether or not that process generates the expected result can be known, but can be known only after the fact. (Of course, prior to the sampling process we can make probabilistic prediction about this.) Thus we can distinguish drift from selection, using the conceptions of drift and selection tentatively suggested above, but only after the fact. (While again we can make good probabilistic predictions prior to the fact.) This means that the outcome, deviation or lack thereof from expectation, is a necessary component of our conceptions of drift and selection. It also means that, as we saw above, characterizations of sampling processes alone will be incapable of making the selection/drift distinction.

I've just claimed that we can know, after the fact, whether or not drift was involved in the transmission of gene and genotype frequencies from one generation to the next. Moreover, I claim, we can quantify the extent to which it is involved. Granted this will be practically difficult in real biological cases, but is it possible in principle?

First consider the common analogy between the evolutionary process and information theory. In information theory we have a source (S), a communication channel and a receiver (R). A message (information) is sent from S to R via the communication channel. Noise in the channel is a measure of the independence of the received message from the sent message, in other words, a measure of the difference between the received message from that which was sent. Can we know, quantitatively, this difference? Yes. After the transmission we can compare the original message to that which was received, and we can quantify the extent to which they differ.[15] Prior to the transmission of some particular message we can quantify the "noisiness" of the channel, either based on past experience, or on some more direct knowledge of the processes involved in transmission. And given the level of noisiness we can predict, probabilisti-

---

[15]It is important to note that this quantitative measure of the effect of "noise" on the message can be wrong. Imagine a message that consists of a finite series of 1's and 0's. Now suppose the third place in the original message is a 1. In the transmission process this 1 gets changed into a 0 and then later back into a 1. If we look only at the final outcome we will mistakenly say that noise had no effect on the third place. But this sort of mistake is correctable. We can add more monitoring stations between S and R. This has an important analogue in the biological measurement of fitness. Our most accurate measures of fitness will be those that follow most completely the whole life cycles of the organisms under study. It follows that the method that I am outlining for quantifying the effects of selection and drift will be accurate exactly so far as our estimates of fitness are accurate. With bad estimates of fitness we can easily mistake selection for drift and vice versa.

cally, the extent to which our message will be transformed. But whether or not noise has an effect in a particular case can be known only after the fact. There is here a single process, transmission over a noisy channel, which sometimes does, and sometimes does not, result in a transformation of the message.

How is this analogous to the process of evolution by natural selection? The signal to be sent is the product of genotypic frequencies and their associated fitnesses (where sampling is discriminate), or in the limiting case of indiscriminate sampling, is just the genotypic frequencies themselves. "The *effect* of sampling error is to introduce 'noise' into the communication channel. Because of this noise, the signal that is received fluctuates from generation to generation." (Roughgarden 1979, p. 58, emphasis added). Just as in the information theory case, here too we can, after the fact, know qualitatively whether or not drift has occurred, i.e., know whether or not the next generation's genotypic frequencies match those predicted by the past generation's frequencies and fitnesses. And if they do not match we can quantify this lack of fit, i.e., we can quantify the effect of drift. As should be clear by now, this *post hoc* quantification of drift requires knowledge of the fitnesses of the individuals of the last generation.

There are two obvious ways to object to the above account of the difference between selection and drift. First, one could deny the existence of individual fitnesses. And one could deny the existence of such things for at least two quite different reasons. One could pursue an instrumentalist stance towards fitness and say that they are not real, but are merely useful instruments. But then one would probably still accept my approach to differentiating selection and drift – because it is useful – and just not give it the realist interpretation I give (Rosenberg 1985, 1988, 1994). Or one could deny that the probabilities that make up an individual's fitness attach to individuals at all. Instead, on this view, the probabilities that play a role in evolutionary theory attach only to ensembles of individuals (see Matthen and Ariew 2002; Walsh et al. 2002). The major point of Brandon and Carson (1996) was that a realist stance towards individual fitness was justified. It is my hope that this article has added further evidence for that claim.

Second, one might admit the possible existence of individual fitness, but deny the possibility of our knowledge of fitness. My comment on this stratagem applies equally to the denial of the reality of fitness as well. It is that there is a whole field of evolutionary biology, ecological genetics, which has as its basic goal the measurement of individual fitnesses. Given that biologists have successfully measured individual fitness in the field and in the lab since, at least, 1898 (see Weldon 1898) I'm not much moved by the claim that we cannot measure it. In my view, the actual is always possible.

Conclusions

I have offered definitions of selection and drift that are outcome-oriented and so can be definitively applied only after the fact. *Drift is any deviation from the*

expected levels of reproduction due to sampling error. Selection is differential reproduction that is due to (and in accord with) expected differences in reproductive success. As we have seen, a purely process-oriented approach cannot make the appropriate distinction between selection and drift because both result from the same process – sampling. While Millstein is surely right to think of selection and drift as processes, her attempt to differentiate them purely in terms of process does not work. The outcomes are necessary not only to operationalize the distinction, but also to make sense of it conceptually.

On my approach drift and selection are outcomes that can be distinguished given the appropriate probabilities that govern the sampling process. Assigning probabilities to individuals' lives, deaths and levels of reproductive success depends, at least, on solving the reference class problem with respect to biology. That has been solved with the concept of selective environmental homogeneity. In the last section I briefly mentioned, but did not pretend to fully address, other potential philosophical objections to my approach. All I can say here is that one had better hope that my approach is the correct one; otherwise the explanatory power of evolutionary theory is greatly diminished.

References

Antonovics J., Ellstrand N.C. and Brandon R.N. 1988. Genetic variation and environmental variation: expectations and experiments. In: Gottlieb L.D. and Jain S.K. (eds), Plant Evolutionary Biology. Chapman and Hall, London, pp. 275–303.

Beatty J. 1984. Chance and natural selection. Philos. Sci. 51: 183–211.

Brandon R.N. 1978. Adaptation and evolutionary theory. Stud. Hist. Philos. Sci. 9: 181–206.

Brandon R.N. 1990. Adaptation and Environment. Princeton University Press, Princeton.

Brandon R.N. and Carson S. 1996. The indeterministic character of evolutionary theory: no "no hidden variables proof" but no room for determinism either. Philos. Sci. 63: 315–337.

Eshel I. and Feldman M.W. 1982. On evolutionary genetic stability of the sex ratio. Theoret. Populat. Biol. 21: 430–439.

Gillespie J. 1977. Natural selection for variances in offspring number: a new evolutionary principle. Am. Natural. 111: 1010–1014.

Hacking I. 1990. The Taming of Chance. Cambridge University Press, Cambridge.

Kettlewell H.B.D. 1955. Selection experiments on industrial melanism in the lepidoptera. Heredity 9: 323–342.

Kettlewell H.B.D. 1956. Further selection experiments on industrial melanism in the lepidoptera. Heredity 10: 287–301.

Mills S. and Beatty J. 1979. The propensity interpretation of fitness. Philos. Sci. 46: 263–286.

Millstein R. 2002. Are random drift and natural selection conceptually distinct?. Biol. Philos. 17: 33–53.

Matthen M. and Ariew A. 2002. Two ways of thinking about fitness and natural selection. J. Philos. 99: 55–83.

Rausher M.D. and Englander R. 1987. The evolution of habitat preference. II. Evolutionary genetic stability under soft selection. Theoret. Populat. Biol. 31: 116–139.

Reichenbach H. 1949. The Theory of Probability. University of California Press, Berkeley.

Rosenberg A. 1985. The Structure of Biological Science. Cambridge University Press, Cambridge.

Rosenberg A. 1988. Is the theory of natural selection a statistical theory? Can. J. Philos. 14(Suppl.): 187–207.

Rosenberg A. 1994. Instrumental Biology or the Disunity of Science. University of Chicago Press, Chicago.

Roughgarden J. 1979. Theory of Population Genetics and Evolutionary Ecology: An Introduction. Macmillan Publishing Co, New York.

Salmon W. 1970. Statistical Explanation and Statistical Relevance. University of Pittsburgh Press, Pittsburgh.

Sober E. 1984. The Nature of Selection. MIT Press, Cambridge, MA.

Straddon D. and Bennington C.C. 1998. Fine-grained spatial and temporal variation in selection does not maintain genetic variation in Erigeron annuus. Evolution 52: 678–691.

Walsh D., Lewens T. and Areiw A. 2002. Trials of life: natural selection and random drift. Philos. Sci. 69: 452–473.

Weldon W.F.R. 1898. Opening address, section D, zoology. Nature 58: 499–506.