

Is Drift a Serious Alternative to Natural Selection as an Explanation of Complex Adaptive Traits?

ELLIOTT SOBER

‘There are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don’t know we don’t know.’
—Donald Rumsfeld, 2003, President George W. Bush’s Secretary of Defense, on the subject of the U.S. government’s failure to discover weapons of mass destruction in Iraq

Gould and Lewontin’s (1978) essay, ‘The Spandrels of San Marco’ is famous for faulting adaptationists for not considering alternatives to natural selection as possible evolutionary explanations. One of the alternatives that Gould and Lewontin say has been unfairly neglected is random genetic drift. Yet, in spite of the article’s wide influence, this particular suggestion has not produced an avalanche of papers in which drift is considered as an explanation of complex adaptive features such as the vertebrate eye. The reason is not far to seek. Biologists whose adaptationism remained undented by the Spandrels paper continued to dismiss drift as an egregious non-starter. And even biologists sympathetic with Gould and Lewontin’s take-home message—that adaptationists need to pull up their socks and test hypotheses about natural selection more rigorously—have had trouble taking drift seriously. Both critics and defenders of adaptationism have tended to set drift to one side because they are convinced that complex adaptive traits have only a tiny probability of evolving under that process. The probability is not zero, but the consensus seems to be that the probability is sufficiently small that drift can safely be ignored.

There is a statistical philosophy behind this line of reasoning that I think is mistaken. Once the mistake is identified, drift becomes an interesting alternative to natural selection, but not because I or anyone else thinks it is a plausible explanation of complex adaptive traits. Rather, the hypothesis is interesting because it provides a *foil*;

considering drift as an alternative to natural selection forces one to identify the information that is needed if one wishes to say that an adaptive feature is better explained by one hypothesis or the other. I'll examine two relatively simple phenotypic models of selection and drift with the goal of identifying these informational requirements. It is interesting that the requirements are not trivial.

1. There is no Probabilistic *Modus Tollens*—Against Fisherian Significance Tests

Modus tollens is familiar to philosophers and scientists; it is the centrepiece of Karl Popper's views on falsifiability:

(MT) If H, then O
 not-O
 not-H

Modus tollens says that if the hypothesis H entails the observation statement O, and O turns out to be false, then H should be rejected. Since *modus tollens* is a deductively valid rule of inference (hence my use of a single line to separate premises from conclusion), perhaps the following probabilistic extension of the rule constitutes a sensible principle of nondeductive reasoning:

(Prob-MT) Pr(O|H) is very high.
 Not-O.
 not-H

According to *probabilistic modus tollens*, if the hypothesis H says that O will *very probably* be true, and O turns out to be false, then H should be rejected. Equivalently, the suggestion is that if H says that some observational outcome (not-O) has a very low probability, and that outcome nonetheless occurs, then we should regard H as false. I draw a double line between premises and conclusion in (Prob-MT) to indicate that the argument form is not supposed to be deductively valid.

Probabilistic *modus tollens* is Fisher's (1925) test of significance. Fisher describes his test as leading to a disjunctive conclusion—either the hypothesis H is false, or something very improbable has

Is Drift a Serious Alternative to Natural Selection?

occurred. Even if this disjunction followed from the premises¹, that would not mean that the first disjunct also follows, either deductively or nondeductively. I agree with Hacking (1965), Edwards (1973), and Royall (1997) that probabilistic *modus tollens* is an incorrect principle. As these authors make clear, lots of perfectly reasonable hypotheses say that the observations are very improbable; in particular, this is something we should expect when the observations are numerous and are conditionally independent of each other, given a *probabilistic* hypothesis. Consider, for example, the hypothesis that a coin is fair. If the coin is tossed a million times, the exact sequence of heads and tails that results will have a probability of $(1/2)^{1,000,000}$. However, that hardly shows that the hypothesis should be rejected. Indeed, every sequence of heads and tails has the same small probability of occurring; probabilistic *modus tollens* therefore claims that the hypothesis should be rejected *a priori*—no matter what the outcome of the experiment turns out to be—but surely that makes no sense.

It may seem that the kernel of truth in (Prob-MT) can be rescued by modifying the argument's conclusion. If it is too much to conclude that H is false, perhaps we may conclude that the observations constitute evidence against H:

(Evidential Prob-MT)	Pr(O H) is very high. not-O
	=====
	not-O is evidence against H.

This principle is also unsatisfactory, as Royall (1997, p. 67) nicely illustrates via the following story: Suppose I send my valet to bring me one of my urns. I want to test the hypothesis (H) that the urn he returns with contains 0.2% white balls. I draw a ball from the urn and find that it is white. Is this evidence against the hypothesis? It may not be. Suppose I have only two urns—one of them contains 0.2% white balls, while the other contains 0.0001% white balls. In this instance, drawing a white ball is evidence *in favour* of H, not evidence *against* it.^{2,3}

¹ Even if H and not-O are true, and Pr(not-O|H) is very low, it does not follow that not-O is 'intrinsically' improbable. It is perfectly possible for there to be another true hypothesis that confers a high probability on not-O.

² Forensic identity tests using DNA data provide further illustrations of Royall's point. For example, Crow (2000, pp. 65–67) computed the probability of a DNA match at 13 loci, based on known allele frequencies. If

Royall's story brings out the fact that judgments about evidential meaning are essentially *comparative*. To decide whether an observation is evidence against H , we usually need to know what alternatives there are to H . In typical cases, *to test a hypothesis requires testing it against alternatives*.⁴ In the story about the valet, observing a white ball is very improbable according to H , but in fact that outcome is evidence in favour of H , not evidence against it. The reason is that O is even more improbable according to the alternative hypothesis. Probabilistic *modus tollens*, in both its vanilla and evidential versions, needs to be replaced by the *Law of Likelihood*:

Observation O favours hypothesis H_1 over hypothesis H_2 if and only if $\Pr(O|H_1) > \Pr(O|H_2)$.

The term 'favouring' is meant to indicate *differential support*; the evidence points away from the hypothesis that says it is less probable and towards the hypothesis that says it is more probable.

This is not the place to undertake a defence of the Law of Likelihood (on which see Hacking 1965, Edwards 1972, Royall 1997) or to consider its limitations (Forster and Sober 2004), but a comment about its intended scope is in order. The Law of Likelihood should be restricted to cases in which the probabilities of hypotheses are not under consideration (perhaps because they are not known or are not even 'well-defined') and one is limited to information about the probability of the observations given different hypotheses. To see why this restriction is needed, consider an example presented in Leeds (2000). One observes that an ace has just been drawn from a standard deck of cards (O) and

⁴ I say that this is 'usually' true because the thesis that testing a hypothesis H is *always* contrastive is false. For example, if a set of true observational claims entails H , there is no need to consider alternatives to H ; one can conclude without further ado that H is true. And, as we know from *modus tollens*, if H entails O and O turns out to be false, one can conclude that H is false without needing to contemplate alternatives.

the individuals are sibs, the probability is 7.7×10^{-32} . The observations are very improbable under the sib hypothesis, but that hardly shows that they are evidence against it. In fact, the data favour the sib hypothesis over the hypothesis that the two individuals are unrelated. If they are unrelated, the probability is 6.5×10^{-38} .

³ A third formulation of probabilistic *modus tollens* is no better than the other two. Can one conclude that H is *probably* false, given that H says that O is highly probable, and O fails to be true? The answer is no; inspection of Bayes' theorem shows that $\Pr(\text{not-}O|H)$ can be low without $\Pr(H|\text{not-}O)$ being low.

Is Drift a Serious Alternative to Natural Selection?

the hypotheses under evaluation are H_1 = ‘the card is the ace of hearts’ and H_2 = ‘the card is the ace of spades or the ace of clubs.’ It follows that $\Pr(H_1|O) = 1/4$ and $\Pr(H_2|O) = 1/2$, while $\Pr(O|H_1) = \Pr(O|H_2) = 1.0$. It would be odd to maintain that O does not favour H_2 over H_1 ; in this case, the favouring relation is mediated by the probabilities of hypotheses, not their likelihoods.⁵ Since scientists are often disinclined to discuss the probabilities of hypotheses (when these probabilities can’t be conceptualized as objective quantities), this restriction of the Law of Likelihood accords well with much scientific practice.

I hope it is clear how Royall’s story and the Law of Likelihood apply to the task of evaluating drift as a possible explanation of a complex adaptive trait. The fact that $\Pr(\text{Data}|\text{Drift})$ is very low does not show that the drift hypothesis should be rejected. It does not even show that the data are evidence against the drift hypothesis. What we need to know is how probable the data are under alternative hypotheses. In particular, we need to consider $\Pr(\text{Data}|\text{Selection})$. If the data favour Selection over Drift, this isn’t because $\Pr(\text{Data}|\text{Drift})$ is *low*, but because $\Pr(\text{Data}|\text{Drift})$ is *lower* than $\Pr(\text{Data}|\text{Selection})$. Our next question, therefore, is how these two likelihoods should be conceptualized.

2. The Two Hypotheses

Let’s begin by temporarily setting to one side examples of complex adaptive features such as the vertebrate eye and consider instead an ostensibly simpler quantitative character—the fact, let us assume, that polar bears have fur that is, on average, 10 cm long. Which hypothesis—selection or drift—confers the higher probability on the trait value we observe polar bears to have?⁶

I will assume that evolution in the lineage leading to present day polar bears takes place in a finite population. This means that there is an element of drift in the evolutionary process, regardless of what else is going on. The question is whether selection also played a role.

⁵ I am grateful to Branden Fitelson for discussion on this point.

⁶ Since we are talking about a continuous variable, the proper concept is *probability density*, not *probability*, and even so, the probability (density) of the bears’ having an average fur length of *exactly* 10 cm. is zero, on each hypothesis. Thus we need to talk about some tiny region surrounding a value of 10 cm. as the observation that each theory probabilifies. The subsequent discussion should be understood in this way.

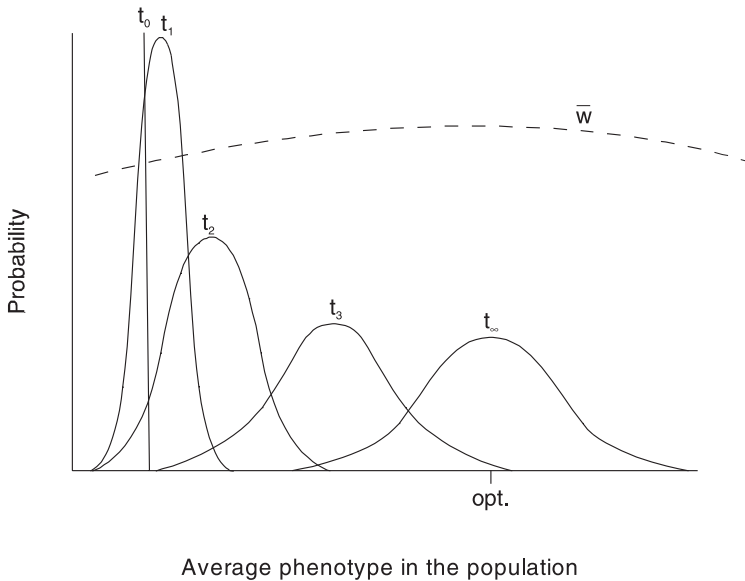
Thus, our two hypotheses are *pure drift* (PD) and *selection plus drift* (SPD). Were the alternative traits identical in fitness or were there fitness differences among them (and hence natural selection)? I will understand the idea of drift in a way that is somewhat nonstandard. The usual formulation is in terms of random *genetic* drift; however, the problem I want to address concerns fur length, which is a *phenotype*. To decide how random genetic drift would influence the evolution of this phenotype, we'd have to know the developmental rules that describe how genes influence phenotypes. I am going to bypass these genetic details by using a purely phenotypic notion of drift. Under the PD hypothesis, a population's probability of increasing its average fur length by a small amount is the same as its probability of reducing fur length by that amount.⁷ Average fur length evolves by random walk. I'll also bypass the genetic details in formulating the SPD hypothesis; I'll assume that the SPD hypothesis identifies some phenotype (O) as the optimal phenotype and says that an organism's fitness decreases monotonically as it deviates from that optimum. Thus, if 12 centimetres is the optimal fur length, then 11 is fitter than 10, 13 is fitter than 14, etc. Given this singly-peaked fitness function, the SPD hypothesis says that a population's probability of moving a little closer to O exceeds its probability of moving a little farther away. The SPD hypothesis says that O is a *probabilistic attractor* in the lineage's evolution. For evolution to occur, either by pure drift or by selection plus drift, there must be variation. I'll assume that mutation always provides a cloud of variation around the population's average trait value.

The dynamics of selection plus drift (SPD) are illustrated in Figure 1, which comes from Lande (1976), whose phenotypic model is the one I am using here. At the beginning of the process, at t_0 , the average phenotype in the population has a sharp value. The state of the population at various later times is represented by different probability distributions. Notice that as the process unfolds, the mean value of the distribution moves in the direction of the optimum specified by the hypothesis. The distribution also grows wider, reflecting the fact that the population's average phenotype becomes more uncertain as more time elapses. After infinite time (at t_∞), the population is centred on the putative optimum. The speed at which the population moves towards this final distribution depends on the trait's heritability and on the strength of selection,

⁷ Except, of course, when the population has its minimum or maximum value. There is no way to have fur that is less than 0 centimetres long; I'll also assume that there is an upper bound on how long the fur can be (e.g., 100 centimetres).

Is Drift a Serious Alternative to Natural Selection?

Figure 1



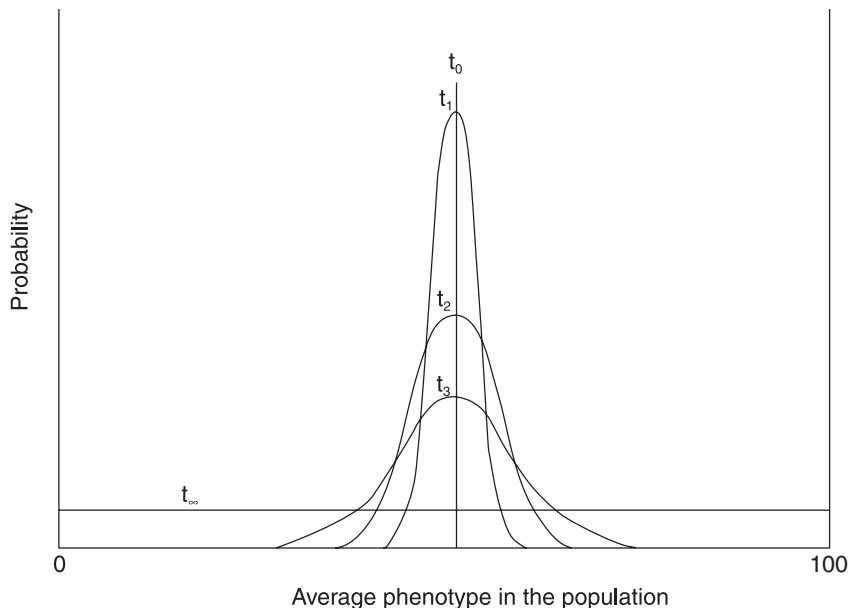
which is represented in Figure 1 by the peakedness of the \bar{w} curve; this curve describes how average fitness depends on average fur length. The width of the different distributions depends on the effective population size N ; the larger N is, the narrower the bell curve. In summary, the SPD hypothesis models the process of selection plus drift as the *shifting and squashing of a bell curve*.

Understood in this way, the SPD hypothesis constitutes a relatively simple conceptualization of natural selection in a finite population. The SPD hypothesis assumes that the fitness function is singly peaked and that fitnesses are *frequency independent*—e.g., whether it is better for a bear to have fur that is 9 centimetres long or 8 does not depend on how common or rare these traits are in the population. I also have conceptualized the SPD hypothesis as specifying an optimum that remains unchanged during the lineage's evolution; the optimum is not a moving target. Indeed, the hypothesis assumes that there is a fur length that is optimal for all bears, regardless of how they differ in other respects. I have constructed the SPD hypothesis with these features, not because I think they are realistic, but because they make the problem I wish to address more tractable. My goal, recall, is to identify the information one must have on hand if one wishes to say whether

SPD or PD has the higher likelihood. Informational requirements do not decline when models are made more complex; in fact, they increase.

Figure 2 depicts the process of pure drift (PD); it involves just *the squashing of a bell curve*. Although uncertainty about the trait's future state increases with time, the mean value of the distribution remains unchanged. In the limit of infinite time (at t_∞), the probability distribution is flat, indicating that all average phenotypes are equiprobable. The rate at which the bell curve gets squashed depends on N , the effective population size; the smaller N is, the faster the squashing occurs.⁸

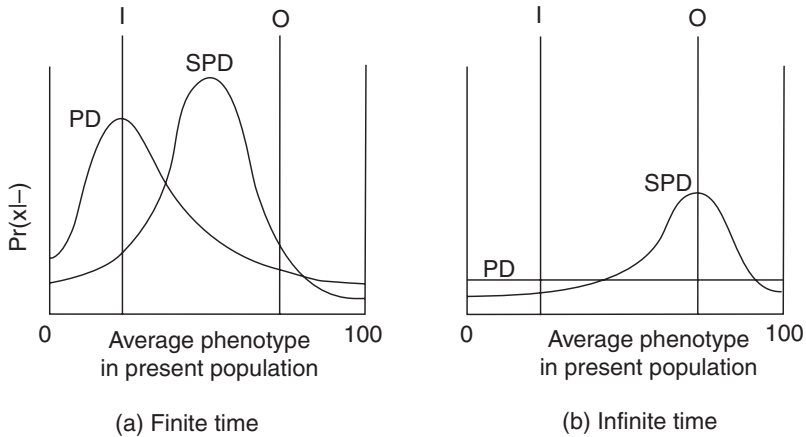
Figure 2



⁸ The case of infinite time makes it easy to see why an explicitly genetic model can generate predictions that substantially differ from the purely phenotypic models considered here. For example, under the process of pure random *genetic drift*, each locus is homozygotic at equilibrium. In a one-locus two-allele model in which the population begins with each allele at 50%, there is a 0.5 probability that the population will be *AA* and a 0.5 probability that it will be *aa*. In a two-locus two-allele model, again with each allele at equal frequency at the start, each of the four configurations *AABB*, *AAbb*, *aaBB*, and *aabb* has a 0.25 probability. Imagine that genotype determines phenotype (or that each genotype has associated with it a

Is Drift a Serious Alternative to Natural Selection?

Figure 3



For both the PD and the SPD curves, it is important to understand that the curves do not describe how an individual's fur length must change, but rather describe possible changes in the population's average fur length. It is perhaps easiest to visualize what is involved by thinking of each curve as describing what will happen if 1000 replicate populations are each subjected to the SPD or the PD process, where each begins with the same initial average fur length. After a given amount of time elapses, we expect these different populations to have different average fur lengths; these different averages should form a distribution that approximates the theoretical distributions depicted in the figures. What these curves say about a single population is that there are many different average fur lengths it might have, and that these different possibilities have the different probabilities represented by the relevant curve.

We now are in a position to analyse when SPD will be more likely than PD. Figure 3a depicts the relevant distributions when there has been finite time since the lineage started evolving from its initial state (I). Notice that the PD distribution stays centred at I, whereas the SPD curve has moved in the direction of the putative optimum (O). Notice further that the PD curve has become more flattened than the SPD curve has; selection impedes spreading out.

different average phenotypic value) and it becomes obvious that a genetic model can predict nonuniform phenotypic distributions at equilibrium. The case of selection-plus-drift is the same in this regard; there are genetic models that will alter the picture of how the average phenotype evolves. See Turelli (1988) for further discussion.

Figure 3b depicts the two distributions when there has been infinite time. The SPD curve is centred at the optimum while the PD curve is flat. Whether finite or infinite time has elapsed, the fundamental fact about the likelihoods is the same: *the SPD hypothesis is more likely than the PD hypothesis precisely when the population's actual value is 'close' to the optimum*. Of course, what 'close' means depends on how much time has elapsed between the lineage's initial state and the present, on the intensity of selection, on the trait's heritability, and on N , the effective population size. For example, if infinite time has elapsed (Figure 3b), the SPD curve will be more tightly centred on the optimum, the larger N is. If 10 is the observed value of our polar bears, but 11 is the optimum, SPD will be more likely than PD if the population is small, but the reverse will be true if the population is sufficiently large.

Given that there are several biological parameters that affect the curves associated with the SPD and the PD hypotheses, it is worth asking which parameter values make it easier to discriminate between the two hypotheses and which values make this more difficult. One crucial factor is the amount of time that has elapsed between the ancestor and the present day species whose trait value we are trying to explain. With only a little time, the predictions of the two hypotheses are mainly determined by I , the population's initial condition, and so the two hypotheses will be pretty much indistinguishable. Only with the passage of more time do the processes postulated by the two hypotheses significantly influence what they predict; with infinite time, the predictions are determined entirely by the postulated processes and the initial condition has been completely 'forgotten.' This means that more time is 'better,' in terms of getting the two hypotheses to make different predictions. The same point holds for the intensity of selection and the heritability; for a fixed amount of time since the initial state I , the higher the values of these parameters the better, in terms of getting the SPD curve to shift significantly away from the PD curve. It is more difficult to gauge the net epistemological significance of N , the effective population size; as noted before, small N makes the PD curve flatten faster, whereas large N makes the SPD curve narrower.

Although the criterion of 'closeness to the putative optimum' suggests that there are just two possibilities that need to be considered in deciding whether SPD is more likely than PD, it is more fruitful to distinguish the four possibilities that are summarized in the accompanying table. In each, an arrow points from the population's initial state (I) to its present state (P); O is the optimum postulated by the SPD hypothesis. The first case (a) is the

Is Drift a Serious Alternative to Natural Selection?

most obvious; if the optimum (O) turns out to be identical with the population's actual trait value (in our example, fur that is 10 cm. long), we're done—SPD has the higher likelihood. However, if the present trait value is different from the optimum value, even a little, we need more information. If we can discover what the lineage's initial state (I) was, and if this implies that (b) the population evolved *away* from the putative optimum, we're done—PD has the higher likelihood. But if our estimates of the values of I and O entail that there has been (c) overshooting or (d) undershooting, we need more information if we wish to say which hypothesis is more likely.

		Which hypothesis is more likely?
(a) present state coincides with the putative optimum	--- ----- ----- I → P=O	selection-plus-drift
(b) population evolves away from the putative optimum	--- -----*----- --- P ← I O	pure drift
(c) population overshoots the putative optimum	--- ----- ----- --- I → O → P	?
(d) population undershoots the putative optimum	--- ----- ----- --- I → P O	?

3. Estimating Biological Parameters

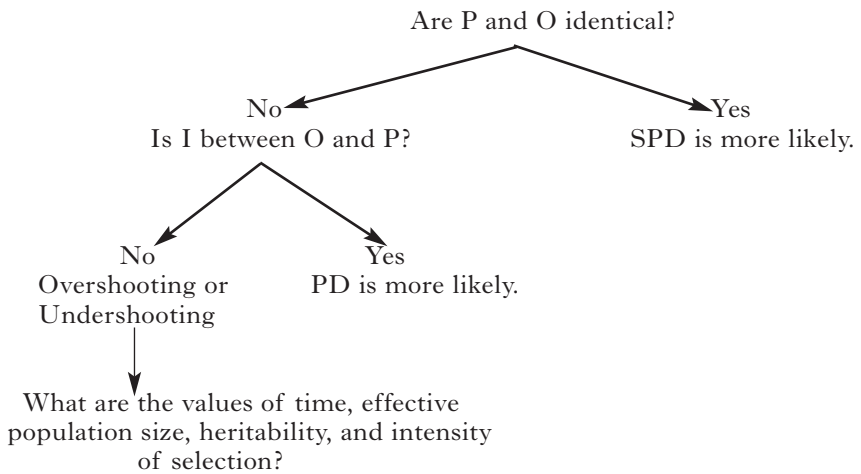
If answering the question of whether SPD is more likely than PD depends on further details, how should those details be obtained? One possibility is that we simply *invent* assumptions that allow each hypothesis to have the highest likelihood possible, and then compare these two 'best cases.' With respect to the SPD hypothesis, this would involve assuming that the actual fur length of 10 centimetres also happens to be the optimal value, that the population has been evolving a very short time, that its initial state was the same as its present state, and that the population is extremely large. The net effect of these assumptions is to push the likelihood assigned to SPD very close to the maximum possible value of unity. The problem is that this is a game that two can play. By assuming that the population was large, that there has been little time since the population started evolving, and that the population began with a trait value of 10, the PD hypothesis will have that same high likelihood.

Simply inventing favourable assumptions has a second defect, in addition to the fact that it leads to a stand-off between the two hypotheses. The problem is that the assumptions are merely that—they are *invented*, not *independently supported*. When we want to

know whether SPD is more likely than PD, we are not asking whether we can invent a detailed description of selection that has a higher likelihood than any hypothesis about drift that we are able to invent. Rather, what we want to know is how SPD and PD compare when each is fleshed out in ways that are independently plausible.⁹

This means that we can use the four possible relationships depicted in the table to construct the protocol for asking questions presented in Figure 4. We first must figure out what the optimal fur length would be for polar bears, if fur length were subject to natural selection. If it turns out that the optimal value (O) is identical with the observed present value (P), we need go no farther—we can conclude that SPD is more likely than PD. However, if the optimum and the actual value differ,¹⁰ we must estimate what the ancestral condition (I) was at some earlier point in the lineage. If the estimated values for O and I, and the observed value for P, are such that I is between O and P, we are finished—we can conclude that PD is more likely than SPD. However, if I is not between O and P, we must estimate additional biological parameters if we want to say which hypothesis has the higher likelihood.

Figure 4



⁹ I criticize ‘intelligent design theory’ for being unable to provide independently supported information about the putative designer’s goals and abilities in Sober (2002a, 2004a, 2004b).

¹⁰ The question of whether O and P are identical must be approached statistically. If $O = 10.01$ and the 100 sampled polar bears have a mean fur length of 10.0, the conclusion may be that O and P are not statistically distinguishable.

Is Drift a Serious Alternative to Natural Selection?

Let us begin with the first question in the protocol. If present day polar bears have an average fur length of 10 cm, how are we to discover what the optimal fur length is? If there is, as I'm assuming, variation in the present population around its mean value, we can observe the survival and reproductive success of individuals that have different trait values. We also might wish to conduct an experiment in which we attach parkas to some polar bears, shave others, and leave others with their fur lengths unchanged. We then can monitor what happens to these experimental subjects, and these observations will allow us to estimate the fitness values that attach to different fur lengths. This experiment will allow us to construct an empirically well-grounded estimate of what the optimal fur length is.

There is a second approach to the problem of identifying the optimal fur length, one that is less direct and more theoretical. Suppose there is an energetic cost associated with growing fur. We know that the heat loss an organism experiences depends on the ratio of its surface area to its volume. We also know that there is seasonal variation in temperature. Although it is bad to be too cold in winter, it also is bad to be too warm in summer. We also know something about the abundance of food. These and other considerations might allow us to construct a model that identifies what the optimal fur length is for organisms that have various other characteristics. Successful modelling of this type does not require the question-begging assumption that the bear's actual trait value is optimal or close to optimal. This methodology has been applied to other traits in other taxa (Alexander 1996; Hamilton 1967; Parker 1978); there is no reason why it should not be applicable in the present context.

As noted earlier, the simple model of selection we are considering assumes a stationary target—the optimal fur length for bears *now* is the same as the optimum that existed while the lineage was evolving. This is almost certainly a highly unrealistic assumption. If we dropped it, we'd have to worry about how an estimate of present optimal values would allow one to infer what trait value was optimal during the time the trait was evolving. As Tinbergen (1964, p. 428) observed '[w]hen one finds that a certain characteristic has survival value ... one has demonstrated beyond doubt a selection pressure which prevents the species in its present state from deviating ... However, the conclusion that this same selection pressure must have been responsible in the past for the *moulding* of the character studied is speculative, however probable it often is.'

Although the SPD hypothesis' assumption of a stationary target

allows us to bypass this question, the question of how to infer past from present cannot be evaded in connection with the second question in the protocol described in Figure 4. We need to decide whether I, the state the population occupied at some earlier time, falls between O and P. The natural way to address this question is to estimate I. Biologists attempt to solve this estimation problem by exploiting the fact that polar bears and other bears alive today share common ancestors. By using other traits that these bears possess, they infer a phylogenetic tree of the sort depicted in Figure 5 in which polar bears and their relatives are tip species. The fur lengths of polar bears and their relatives can then be written onto the tips of that tree. The observed character states of these tip species provide evidence about the character states of ancestors, which are represented by interior nodes. What inference procedure should we use to infer these ancestral trait values?

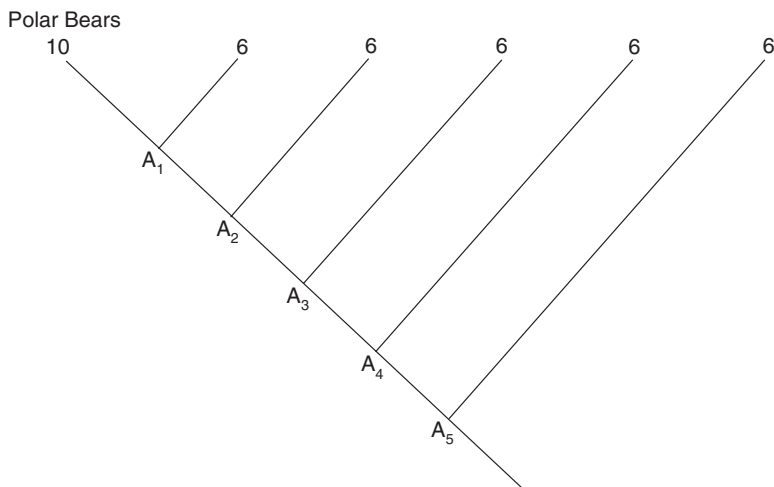


Figure 5

Before addressing that question, I want to explain why Figure 5 shows that our question about SPD versus PD is ambiguous—there are many SPD versus PD questions about polar bear fur length, not just one. It is obvious that present day polar bears have multiple ancestors, and equally obvious that different ancestors may well have had different fur lengths. If these were all known, the problem of explaining why polar bears now have fur that is 10 centimetres

Is Drift a Serious Alternative to Natural Selection?

long would decompose into a number of subproblems—why the fur length present at A_5 evolved to the length present at A_4 , why A_4 's fur length evolved to the value found at A_3 , etc. SPD may be a better answer than PD for some of these transitions, but the reverse might be true for others. Similarly, suppose we infer the fur length of just one of these ancestors and ask whether the evolution of fur length from that ancestral value to the 10 cm we observe in present day polar bears favours SPD or PD. The answer may depend on which ancestor we consider. If the inferred optimum is 12 and we infer that $A_2 = 11$, then the lineage leading from A_2 to the present has evolved away from the putative optimum, and we conclude that PD is more likely than SPD. If, however, we focus on A_3 , and infer that $A_3=8$, then the lineage leading from A_3 to the present undershot the optimum, and we are in case (c) described in the Table; with further biological information, it may turn out that SPD is more likely than PD. The question of whether SPD or PD is the more likely explanation of an observed trait value thus needs to be relativized to a choice of ancestor.

Now back to the problem of inferring the character states of ancestors. A standard method that biologists use is *parsimony*—we are to prefer the assignment of states to ancestors that minimizes the total amount of evolution that must have occurred to produce the trait values we observe in tip species. This is why assigning ancestor A_1 in Figure 5 a value of 8 and the other ancestors a value of 6 is said to have greater credibility than assigning them all a value of 10. But why should we use parsimony to draw this inference? This is a large question, which I won't attempt to answer here. However, a few points may be useful. First, it turns out that if drift is the process at work in a phylogenetic tree, then the most parsimonious assignment of trait values to ancestors (where parsimony means minimizing the *squared* amount of change) is also the assignment of maximum likelihood (Maddison 1991). On the other hand, if there is a directional selection process at work, parsimony and likelihood can fail to coincide (Sober 2002b). This point can be grasped by considering a very simple example—the lineage leading from a single ancestor to present day polar bears. If bears now have a trait value of 10, the most parsimonious assignment of character state to the ancestor is, of course, 10. But suppose the lineage has been undergoing strong selection that is pushing the lineage towards an optimal fur length of (say) 13; if so, the most likely assignment of trait value to the ancestor is some number less than 10; the exact value of that best estimate depends on the amount of time separating ancestor and descendant, the

intensity of selection, and the heritability.¹¹ It follows that parsimony does not provide evidence about ancestral character states that is *independent* of the PD and SPD hypotheses we wish to test.¹²

Given that it is question-begging in the context of testing drift against selection to use parsimony to infer the initial state (I) of the lineage leading to present day polar bears, one remedy may be to consider a range of possible values for I. If the drift hypothesis says that the maximum likelihood estimate of the fur length of an ancestor is 8 and the selection hypothesis says that the maximum likelihood estimate is 6, then perhaps one should see how setting I to different values between 6 and 8 affects the likelihood comparison of PD and SPD. If the greater likelihood of one hypothesis over the other is robust, then the problem of inferring I can be set to one side. However, I see no reason to expect that the problem will usually disappear in this way.¹³

Similar issues arise in connection with the last question in the protocol. If $O \neq P$ and if I is not between O and P, then we need estimates of effective population size, heritability, intensity of selection, and so on. If we can't answer the questions in this protocol, we are in no position to say which hypothesis is more likely.

¹¹ Here's an analogy: imagine you want to swim across a river that has a strong current. The way to maximize your probability of reaching a target on the other side is *not* to start directly across from it; rather, you should start a bit upstream. How far upstream you should go depends on the width of the river, the strength of the current, and on how strong a swimmer you are.

¹² This is the central problem with the procedure for testing adaptive hypotheses proposed by Ridley (1983); see Sober (2002b) for further discussion.

¹³ The discovery of fossils helps solve this problem, but does not solve it completely. Even if fur length could be inferred from a fossil find, it is important to remember that we can't assume that the fossils we observe are *ancestors* of present day polar bears. They may simply be *relatives*. If so, the question persists—how is one to use these data to infer the character states of the most recent common ancestor that present day polar bears and this fossil share? The fact that the fossil is closer in time to this ancestor than is an organism that is alive today means that the fossil will provide stronger evidence. If present day polar bears have a fur length of 10 and the fossil has a fur length of 6, and it is known that the fossil is temporally much closer to the most recent common ancestor than present day polar bears are, then the maximum likely estimate of the ancestor's character will not be 8, but will be closer to 6 (assuming a pure drift process). Likelihood and parsimony do not agree in this case.

Is Drift a Serious Alternative to Natural Selection?

My analysis of how the PD and SPD hypotheses ought to be compared rests on the demand for 'independent evidence' concerning the values of various biological parameters. Some of these parameters are shared between the two models. For example, the likelihood of each depends on what the initial state (I) of the lineage was, on how much time elapsed between I and the present state P, and on the effective population size. The demand for independent evidence in these cases is the demand that the estimate not depend on assuming that the PD hypothesis is true or on assuming that the SPD hypothesis is true. However, there are parameters in the SPD hypothesis that do not occur in the PD hypothesis—most obviously the putative optimum O. What would it mean to have 'independent evidence' concerning the value of O? After all, if an optimal phenotype exists, the SPD hypothesis must be true and the PD hypothesis false. Here the demand for independent evidence is the demand that one's estimate of O not rest on the assumption that the present state P of the population is optimal or close to optimal. The point is to find the most reasonable estimate of O, on the assumption that the SPD hypothesis is true.

4. A Digression on Dichotomous Characters

Perhaps the epistemological difficulties just described would disappear if we redefined the problem. Instead of asking why polar bears now have fur that is 10 centimetres long, perhaps we should ask why they have 'long' fur rather than 'short.' Isn't it clear that polar bears are better off with long fur than they would be with short? If so, long is the *optimal* fur length in this dichotomous character. Doesn't this allow us to conclude without further ado that SPD has higher likelihood than PD, according to the protocol described in Figure 4? Apparently, you don't need to know the ancestral fur length or other biological details to make this argument.

It is interesting how often informal reasoning about natural selection focuses on dichotomous qualitative characters. For example, sociobiologists ask why human beings 'avoid incest,' not why they avoid it to the degree they do. The adaptive hypothesis is that selection favours outbreeding over inbreeding. This hypothesis renders the observed 'avoidance of incest' more probable than does the hypothesis that says that a pure drift process occurred. Of course, the problem gets more difficult if we estimate how much inbreeding there is in human populations and then ask whether that quantitative value is more probable under the SPD hypothesis or

the PD hypothesis. However, why can't an adaptationist admit that this quantitative problem *is* more difficult and still insist on the correctness of the simple likelihood argument just described to solve the qualitative problem?

There is a fly in the ointment. What does it mean to say that fur is 'long' rather than 'short?' No matter where the cut-off is drawn to separate 'short' fur from 'long,' it will do violence to the trichotomy implicit in our fundamental finding about Figure 3—that SPD is more likely than PD when the observed fur length is 'close' to the optimum, whereas fur lengths that are too long or too short render PD more likely than SPD. Long fur is not, contrary to appearances, unambiguous evidence favouring the hypothesis of natural selection.

The problem with imposing dichotomous descriptors ('long' versus 'short') on a quantitative character in the case of fur length seems to arise from the fact that fur length has an intermediate optimum. But why should this be a problem in connection with a feature like incest avoidance, where the less of it the better (or so I will assume for the sake of posing this problem)? In this case, there will be two regions of parameter space, not three. If human beings have a rate of incest that is 'close' to zero, then SPD is likelier than PD; otherwise, the reverse is true. The problem is to say how close is close enough (Sober 1993). How much incest is consistent with saying that SPD is more likely than PD? Of course, if human beings had a zero rate of incest, we'd be done—SPD would be more likely than PD. But if the rate is nonzero, it is unclear how to classify the observation, and so it is unclear whether SPD is more likely. We need further biological information to answer this question. Moving to a dichotomous description of the data doesn't change that fact.

5. The Significance of Adaptive Complexity—from Fur Length to the Camera Eye

It may strike the reader that the example I have been considering—fur length in polar bears—is rather simple and therefore differs in important respects from the problem of testing adaptive hypotheses about a complex structure like the vertebrate eye. In fact, I'm not so sure that bear fur length really *is* so simple. But even if it is, the problems just adumbrated apply with equal force to the task of explaining complex adaptive features. There are additional wrinkles, as we shall see, but questions concerning the relationship

Is Drift a Serious Alternative to Natural Selection?

of the initial state (I), the present state (P), and the optimal state (O) remain relevant to identifying the predictions of the SPD hypothesis.

Fur length has an obvious 'transformation series.' If a population is going to evolve from an average fur length of 3 cm to an average of 5 (let us assume by a series of small changes), then it must pass through an average that is around 4. But consider the evolution of the camera eye in the vertebrate line. If we trace this lineage back far enough, we will find an ancestor that does not have any eye at all. Again assuming that changes must be small, we can ask what the intermediate stages were through which the lineage must have passed as it evolved from no eye to a camera eye. A more general approach would be to conceive of this problem probabilistically; there may be more than one possible transformation series, with different probabilities attaching to different possible changes in character state. The reconstruction of this transformation series is a nontrivial evolutionary problem.

There are nine or ten basic eye designs found in animals, with many variations on those themes. In broad strokes, this variation can be described as follows: vertebrates, squid, and spiders have camera eyes, most insects have compound eyes (but so do many shallow water crustacea), the *Nautilus* has a pinhole eye, the clam *Pecten* and the crustacean *Gigantocypris* have mirror eyes, and flatworms, limpets, and bivalve molluscs have cup eyes. When biologists place these features at the tips of an independently inferred phylogenetic tree (and use parsimony to infer the character states of ancestors), they conclude that these and other basic designs evolved somewhere between 40 and 65 times in different lineages (Salvini-Plawen and Mayr 1977; Nilsson 1989). For each monophyletic group of tip species that share a given eye design (e.g., vertebrates with their camera eyes), we can ask whether the trait value exhibited favours SPD over PD.

The first question in the protocol described in Figure 4 is to determine which design is optimal for each species or taxonomic group. In the case of polar bear fur length, we considered a simple experiment that could provide information about this. Is there a similar experiment for the case of eye design? Present technology makes this unfeasible. Although it is easy enough to remove or diminish the efficiency of whatever light-sensitive apparatus an organism possesses, it is harder to augment those devices or to substitute one complex structure for another. On the other hand, considerable information is now available concerning the optical properties of different eye designs, though a great deal remains to

be learned. Nilsson (1989, p. 302) agrees with Land's (1984) contention that 'if the *Nautilus* had a camera-type eye of the same size, it would be 400 times more sensitive and have 100 times better resolution than its current pinhole eye.' He has similar praise for camera eyes as compared to compound eyes: 'if the human eye was scaled down 20 times to the size of a locust eye, image resolution would still be an order of magnitude better than that of the locust eye. Diffraction thus makes the compound eye with its many small lenses inherently inferior to a single-lens eye (Nilsson 1989, p. 306).'

If the camera eye is *globally* optimal, the protocol described in Figure 4 may appear to entail that SPD is more likely than PD with respect to the organisms that have camera eyes. However, this protocol was predicated on the assumption of a monotonic fitness function. As it turns out, current thinking about eye evolution rejects monotonicity. This means that we have to rethink the protocol in the context of the idea of multiple adaptive peaks. Biologists have been led to reject monotonicity by considering why the camera eye isn't more widely distributed, given that the camera eye seems to be fitter than both the pin hole and the compound eye. Spiders and squid are as lucky as we are, but bees and the *Nautilus* are not. Why not? Nilsson (1989, p. 306) suggests that compound eyes are trapped on a local adaptive peak. He agrees with Salvini-Plaven and Mayr (1977) that

... at an early stage of evolution, the simple eye would be just a single pigment cup with many receptors inside ..., whereas the compound eye would start as multiple pigment cups with only a few receptors in each ... At this low degree of sophistication, neither of the two designs stands out as better than the other. It is only later, when optimized optics have been added, that the differences will become significant. But then there is no return, and the differences remain conserved.

Expressed in terms of the idea of a transformation series, the thought is that evolving from a compound eye to a camera eye would have to retrace steps—there is no way to go directly from one complex design to the other—and the retracing would involve passing through arrangements of lower fitness.

How does this non-monotonic fitness function affect the protocol for testing SPD against PD? If the population's initial state (I) and its present state (P) are in the zone of attraction of the same adaptive peak, the analysis proceeds as before. However, if I is in the zone of attraction of local optimum O, but the lineage evolves to a point in the zone of attraction of O' (where $O \neq O'$), the

Is Drift a Serious Alternative to Natural Selection?

analysis will be more complicated. The probability that a population beginning in the zone of attraction of O will 'jump across a valley' and end in the zone of attraction of O' (where $O \neq O'$) depends on the height of O , the effective population size, the amount of time there is between the initial state and the present, the heritability, and the width of the valley (Lande 1985). This means that a population that now sits on top of the *global* adaptive peak can still be such that SPD has lower likelihood than PD, if the population's initial state and its other parameters have the right values. The protocol described in Figure 4 therefore needs to be modified. With multiple peaks, the fact that $P=O$ (where O is *any* of the optima, even the global optimum) does not suffice to settle which hypothesis has the higher likelihood. The protocol needs to begin, not with the question of whether $P=O$, but rather by asking whether $P=O$ (for some optimum O) *and* whether I is in the zone of attraction of O . The lineage's initial state (I) must enter the protocol from the outset. The shift from a monotonic fitness function to one with multiple peaks thus requires that the test procedure be more historical.

Are there are other differences between polar bear fur length and the vertebrate eye that affect how SPD and PD should be tested? What of the alleged fact that the collection of traits that we call 'the vertebrate eye' is more complex than the simple trait of having fur that is 10 cm long? Complex traits are often said to have a larger number of components than simpler traits. To explore this idea, let's consider a complex trait T^* that can be understood as the conjunction of a number of simpler constituent traits T_1, T_2, \dots, T_n . The vertebrate eye is 'complex' because it has this feature *and* that feature *and* this other feature besides. If the PD hypothesis says that these traits evolved independently of each other, then the likelihood of the PD hypothesis decomposes as follows:

$$\Pr(T^*|PD) = \Pr(T_1 \& T_2 \& \dots \& T_n |PD) = \Pr(T_1|PD) \cdot \Pr(T_2|PD) \cdot \dots \cdot \Pr(T_n|PD).$$

It isn't inevitable that a hypothesis invoking drift without selection should take this form; for example, two genes that are closely linked may each experience random drift. However, for the sake of a simple example, let's consider this formulation of the PD hypothesis.

What does the SPD hypothesis say about the evolution of the complex trait T^* ? Let's begin with the easiest case, in which selection is taken to act simultaneously and independently on each

Elliott Sober

of the simple constituents. In this case, the likelihood of the SPD hypothesis is

$$\Pr(T^* | \text{SPD}) = \Pr(T_1 \ \& \ T_2 \ \& \ \dots \ \& \ T_n \ | \text{SPD}) = \Pr(T_1 | \text{SPD}) \cdot \Pr(T_2 | \text{SPD}) \cdot \dots \cdot \Pr(T_n | \text{SPD}).$$

If each component trait T_i favours SPD over PD—that is, if $\Pr(T_i | \text{SPD}) > \Pr(T_i | \text{PD})$, for each i —then

$$\frac{\Pr(T^* | \text{SPD})}{\Pr(T^* | \text{PD})} > \frac{\Pr(T_i | \text{SPD})}{\Pr(T_i | \text{PD})}.$$

The complex trait provides stronger evidence for SPD over PD than does any of its simple constituents.¹⁴

Here's an analogy: Suppose an investment firm predicts that stock T_1 will rise in price tomorrow, and the prediction comes true. This successful prediction offers some evidence, if only a little, that the firm's prediction was based on knowledge of the stock market¹⁵ rather than being the result of a coin toss the firm performed. The evidence favouring intelligent design over chance increases as the number of correct predictions increases. The 'simple' fact that stock T_1 increased in price counts as evidence, but far more weighty is the 'complex' fact that the company correctly predicted price increases for T_1 , T_2 , ..., and T_n . It isn't that complex facts are automatically evidence for design over chance, whereas simple facts are not; rather, the point is that when several simple facts are each evidence for design over chance, the complex constructed from those simples is stronger evidence still (assuming some degree of conditional independence amongst them). What holds for intelligent design versus chance holds also for SPD versus PD.

The above argument shows how a complex character is related to its simpler constituents *if* each of those constituents favours SPD over PD. But how does the complexity of a trait affect how fast the trait can be expected to evolve towards its optimum? Wagner (1988) and Orr (2000) show that the complexity of a trait, as measured by the number of independent dimensions that are needed to score different trait values for their relative fitness, is relevant to the rate

¹⁴ Notice that I am here comparing a complex trait with each of its simpler constituents. I am not comparing complex and simple traits that are not related in this way. I see no reason to believe the stronger thesis that *all* complex adaptive features provide stronger evidence for SPD over PD than *any* simple adaptive feature provides.

¹⁵ Here I assume for the sake of a simple example that the firm wants to make accurate predictions.

Is Drift a Serious Alternative to Natural Selection?

of evolution under SPD—for fixed mutation size, distance from the optimum, and population size, more complex traits evolve more slowly under SPD. This means that the complexity of a trait is relevant to the answers we will get when we ask the questions described in our protocol; however, notice that the questions we need to ask are the same for both simpler and more complex traits.

I just conceptualized the SPD hypothesis as saying that selection acts independently and simultaneously on each constituent of the complex trait T^* . However, selection processes are often not thought of in this way. Consider, for example, an SPD hypothesis concerning a complex trait that has two constituents, T_1 and T_2 , where the hypothesis says that the process begins with T_1 undergoing selection, and that T_2 is favoured by selection only after T_1 has become pretty common; before that time, T_2 is subject to drift. Here selection for T_1 and selection for T_2 occur sequentially, not simultaneously. This is a more complicated formulation of the SPD hypothesis than the one just described, but the point of relevance is the same. If SPD is more likely than PD with respect to the evolution of trait T_1 , and the same is true with respect to trait T_2 , then the complex trait T^* will favour SPD over PD more strongly than either of the simple constituent traits does. The complexity of T^* is in this sense relevant to the problem of testing selection against drift, but not because this property of T^* allows one to ignore the protocol presented in Figure 4.

As a third and final formulation of SPD, it is interesting to consider traits that have the property that intelligent design theorists term ‘irreducible complexity’ (Behe 1996). The idea is not new; it is exemplified in what Paley thought was special about the vertebrate eye. The eye has a number of parts and these parts must be arranged *just so* if the eye is to perform its function. The fitness function implicit in this idea is that T^* is a conjunction of n quantitative traits T_1 & T_2 & ... & T_n , where there is a fitness advantage to an organism only if it has a rather narrowly circumscribed value for each conjunct. This is easy to visualize if T^* has just two conjuncts, T_1 and T_2 . Imagine a plane whose x- and y-axes represent values for T_1 and T_2 and where the elevation above the plane represents fitness. T^* will be irreducibly complex if the fitness surface is flat, except for a narrowly circumscribed bump, which represents the fitness advantage an organism receives if it has the right values for both T_1 and T_2 .

Before discussing how irreducible complexity affects the comparison of the SPD and PD hypotheses, I want to comment on what intelligent design theorists say about such traits. They claim

that no mindless natural process can produce adaptive traits of this sort. This assertion involves a double overstatement. First, intelligent design theorists do not have the power to foresee the mindless processes that science will someday learn to describe (neither does anyone else, of course). Second, evolutionary theory doesn't say that irreducibly complex traits *cannot* evolve; at most, the theory says that they have a *low probability* of evolving (Sober 2002a). But, as noted at the start of this paper, the fact that the probability of T* is low, given SPD, is not sufficient for any conclusion about SPD to be drawn. The relevant question must be comparative.

The competing hypotheses we have been considering—SPD and PD—each regard the lineage leading to the present trait value P as beginning in some initial character state I. Our present question is whether SPD makes the attainment of P more probable than PD does, if the fitness function exhibits irreducible complexity. If P is optimal, the answer is *yes* (note that there is just one adaptive peak on the surface we are taking to characterize the idea of irreducible complexity). However, there is a difference that irreducible complexity introduces. Irreducible complexity involves a fitness function that is flat except that there is a narrowly delimited bump on it; the contrasting case is a fitness function that everywhere slopes upward towards a single peak. When a trait evolves from I to the optimal value of O, the idea of irreducible complexity entails that the trait must first move through a pure drift process before it reaches the small region in which selection can take it to the optimal value O; in contrast, if the fitness function is monotonic, evolving from I to O will be driven by selection from start to finish, This means that if the observed trait T* coincides with what the SPD hypothesis says is the optimal trait value, and if T* is also irreducibly complex, then T* favours SPD over PD *more weakly* than would be the case if T* were optimal but not irreducibly complex. The likelihoods of SPD and PD are more similar for traits that are irreducibly complex than they are for traits that are not.

I have described four ideas about how the complexity of a trait might be understood. In the first, we shifted from a monotonic fitness function to multiple adaptive peaks, and the result was that the protocol described in Figure 4 had to be modified, even though the relationship of I, O, and P remains critical. In the next three interpretations, a complex trait was understood as one that has a number of constituents; in (i) selection acts simultaneously and independently on each conjunct; in (ii), it acts sequentially on each conjunct; in (iii) selection occurs only when an organism has the

Is Drift a Serious Alternative to Natural Selection?

right trait value for each of the constituent conjuncts. These differences can affect whether SPD or PD will have the higher likelihood, but they do not affect the questions about those hypotheses that need to be addressed.

6. Concluding Comments

I have focused in this paper on the task of deciding whether selection or drift, each conceptualized phenotypically, is the better explanation of a single trait value found in a single species or taxonomic group. The reason I formulated the problem in this way is that it is simple (and so it is a good place to begin) and also because many biologists invoke natural selection to explain observations of this kind and would be loathe to think of drift as an alternative that is worth considering. For example, sociobiologists and evolutionary psychologists often focus on traits they think are universal in our species and seek to explain them in terms of natural selection.¹⁶ If the trait is adaptive, the drift hypothesis gets dismissed out of hand. And if the trait is both adaptive and complex, the suggestion that drift should be considered will strike many biologists as both pedantic and obscurantist. I have described the biological information that is needed if one wants to defend the claim that selection is more likely than drift. Appeal to intuition is not enough. I also have described some of the difficulties that must be overcome if such information is to be obtained. Although it may seem 'obvious' that adaptive features are better explained by appeal to selection than by appeal to drift, producing evidence for that 'obvious' claim is far from trivial. And the fact that the adaptive feature is 'complex' does not let one off the hook.

Although this paper addresses the problem of explaining a single trait value found in a single group of organisms, some biologists will be disinclined to consider this an appropriate datum for testing selection against drift. Rather, their preference will be to consider

¹⁶ It is a familiar point in the debate about adaptationism that the genetic system can prevent the fittest of the available phenotypes from evolving to fixation; see Sober (1993), pp. 123–7 for a simple exposition of this point in connection with the idea of heterozygote superiority and footnote 8 of the present essay for the bearing of this point on the SPD and PD hypotheses. I have chosen to explore purely phenotypic models in order to show that it is far from inevitable that selection is more likely than drift as an explanation of complex adaptive features, even when we restrict our attention to such models.

facts about the distribution of character states across a group of taxa. Instead of asking why polar bears have fur that is 10 cm long, they will want to ask why bears living in colder climates tend to have longer fur than bears living in warmer climates. The choice between these *explananda*—a point value for a single taxon versus a correlation within a group of taxa—is significant. This is because the protocol for testing SPD against PD is different in the two cases. The protocol for explaining a point value is given in Figure 4. For discussion of the protocol relevant to explaining a correlation, see Sober and Orzack (2003). It turns out that less biological information is needed for one to decide between selection and drift in this case.

Regardless of how shifting the *explanandum*—from the point value of a single taxon to a correlation within a set of taxa—changes the epistemological landscape, the argument of this paper constitutes a challenge to those who wish to address *explananda* of the first kind. Although biologists sometimes assert that the traits they study are ‘optimal,’ they often are happy to grant that their confidence in the role played by natural selection does not require strict optimality. The vertebrate eye has a blind spot, but this imperfection has not made many friends for the hypothesis of random genetic drift. After all, even if a complex adaptive feature is imperfect, it still is complex and adaptive, and that is the end of the matter, or so one might be inclined to think. However, according to the protocol described in Figure 4, imperfection (in the sense that the observed trait value differs from what one thinks is the optimal trait value) requires that other questions be asked and answered. Paley (1802) thought it was intuitively obvious that the watch found on the heath favours intelligent design over chance, and that this conclusion does not depend on whether the watch keeps perfect time.¹⁷ Latter day adaptationists have followed Paley’s lead. According to the models of selection and drift explored here, the problem is in fact more subtle.

In the Spandrels article, Gould and Lewontin complain that if one adaptive hypothesis is refuted, it is *easy*—indeed, *too easy*—to invent another. They do not say so, but the same point applies to a pluralistic hypothesis that accords an important role to selective and to nonselective processes as well; if one hypothesis of this type is

¹⁷ In defiance of Gould’s (1980) argument concerning the panda’s thumb, Paley (1802) thought that intelligent design hypotheses are not refuted by adaptive imperfection. In Sober (2004a), I agree with Paley against Gould on this point.

Is Drift a Serious Alternative to Natural Selection?

refuted, another hypothesis of the same kind can be invented (Sober 1993). This *tu quoque* does not mean that Gould and Lewontin's criticism is wrong; rather, the lesson is that every theoretical approach must come to terms with this problem. In the present context, the solution I have suggested is to require that biological parameters be estimated on the basis of independent evidence. It is easy enough to *stipulate* that the optimal fur length for polar bears is 10 cms; what is more difficult is to justify this estimate in terms of independently plausible observational and theoretical considerations. If one adaptive hypothesis fails, by all means let another be invented! But inventing parameter values is not the same thing as justifying them; and justifying the estimates used in a hypothesis is not the same thing as testing that hypothesis against alternatives.

Acknowledgments

My thanks to James Crow, Branden Fitelson, Patrick Forber, Richard Lewontin, Will Provine, and Larry Shapiro for useful discussion, and to Johan Van Bentham for his serendipitous reference to Rumsfeld.

References

- Alexander, R. M. 1996. *Optima for Animals*. Princeton: Princeton University Press.
- Behe, M. 1996. *Darwin's Black Box*. New York: Free Press.
- Crow, J. 2000. 'The Future of Forensic DNA Testing—Predictions of the Research and Development Working Group.' *National Institute of Justice*: NCJ 183697.
- Edwards, A. 1972. *Likelihood*. Cambridge: Cambridge University Press.
- Fisher, R. 1925. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- Forster, M. and Sober, E. 2004. 'Why Likelihood?' In M. Taper and S. Lee (eds.), *The Nature of Scientific Evidence*, Chicago: University of Chicago Press.
- Gould, S. (1980). *The Panda's Thumb*. New York: Norton.
- Gould, S. and Lewontin, R. (1978): 'The Spandrels of San Marco and the Panglossian Paradigm—A Critique of the Adaptationist Paradigm.' *Proceedings of the Royal Society B*, **205**, 581–98.
- Hacking, I. 1965. *The Logic of Statistical Inference*. Cambridge: Cambridge University Press.

Elliott Sober

- Hamilton, W. D. 1967.: 'Extraordinary Sex Ratios.' *Science*, **156**, 477-488.
- Kingsolver, J. and Koehl, M. 1985. 'Aerodynamics, Thermoregulation, and the Evolution of Insect Wings—Differential Scaling and Evolutionary Change.' *Evolution*, **39**, 488–504.
- Kirkpatrick, M. 1996. 'Genes and Adaptation—a Pocket Guide to the Theory.' In M. Rose and G. Lauder (eds.), *Adaptation*. New York: Academic Press, pp. 125–46.
- Land, M. 1984. 'Molluscs.' In M. Ali (ed.), *Photoreception and Vision in Invertebrates*. New York: Plenum, pp. 699–725.
- Lande, R. 1976. 'Natural Selection and Random Genetic Drift in Phenotypic Evolution.' *Evolution*, **30**, 314–34.
- Lande, R. 1985. 'Expected Time for Random Genetic Drift of a Population between Stable Phenotypic States.' *Proc. Nat. Acad. Sci. USA*, **82**, 7641–5.
- Leeds, S. 2000. 'Other Minds, Support, and Likelihoods.' unpublished manuscript.
- Maddison, W. 1991. 'Squared-Change Parsimony Reconstructions of Ancestral States for Continuous-Valued Characters on a Phylogenetic Tree.' *Systematic Zoology*, **40**, 304–14.
- Nilsson, D. 1989. 'Vision Optics and Evolution.' *Bioscience* 39: 298–307.
- Orr, A. (2000): 'Adaptation and the Cost of Complexity.' *Evolution*, **54**, 13–20.
- Orzack, S. and Sober, E. 2001. 'Adaptation, Phylogenetic Inertia, and the Method of Controlled Comparisons.' In S. Orzack and E. Sober (eds.), *Adaptationism and Optimality*, Cambridge University Press, pp. 45–63.
- Paley, W. 1802. Natural Theology, or, Evidences of the Existence and Attributes of the Deity, *Collected from the Appearances of Nature*. London: Rivington.
- Parker, G. 1978. 'Search for Mates.' In J. Krebs and N. Davies (eds.), *Behavioral Ecology—An Evolutionary Approach*. Oxford: Blackwell.
- Ridley, M. 1983. *The Explanation of Organic Diversity*. Oxford: Oxford University Press.
- Royall, R. 1997. *Statistical Evidence—a Likelihood Paradigm*. London: Chapman and Hall.
- Salvini-Plaven, L. and Mayr, E. 1977. 'On the Evolution of Photoreceptors and Eyes.' In M. Hecht, W. Sterne, and B. Wallace (eds.), *Evolutionary Biology*, vol. 10, New York: Plenum, pp. 207–63.
- Sober, E. 1993. *Philosophy of Biology*. Boulder: Westview Press.
- Sober, E. 2002a. 'Intelligent Design and Probability Reasoning.' *International Journal for the Philosophy of Religion*, **52**, 65–80.
- Sober, E. 2002b. 'Reconstructing Ancestral Character States—A Likelihood Perspective on Cladistic Parsimony.' *The Monist*, **85**, 156–76.
- Sober, E. 2004a. 'The Design Argument.' In W. Mann (ed.), *The Blackwell Guide to Philosophy of Religion*. Oxford: Blackwell Publishers.
- Sober, E. 2004b. 'Intelligent Design is Untestable. What about Natural Selection?' In P. Hajek, L. Valdes-Villanueva, and D. Westerstaahl (eds.),

Is Drift a Serious Alternative to Natural Selection?

Proceedings of the 11th Congress of the International Union for History and Philosophy of Science, Division of Logic, Methodology, and Philosophy of Science. Elsevier.

Sober, E. and Orzack, S. 2003. 'Common Ancestry and Natural Selection.' *British Journal for the Philosophy of Science*, **54**, 423–37.

Tinbergen, N. 1964. 'On Aims and Methods of Ethology.' *Zeitschrift für Tierpsychologie*, **20**, 410–33.

Turelli, M. 1988. 'Population Genetic Models for Polygenic Variation and Evolution.' In B. Weir, E. Eisen, M. Goodman, and G. Namkoong (eds.), *Proceedings of the Second International Conference on Quantitative Genetics*. Sunderland, MA: Sinauer, pp. 601–18.

Wagner, G. 1988. 'The Influence of Variation and of Developmental Constraints on the Rate of Multivariate Phenotypic Evolution.' *Journal of Evolutionary Biology*, **1**, 45–66.

