Really Statistical Explanations and Genetic Drift

Marc Lange*†

Really statistical (RS) explanation is a hitherto neglected form of noncausal scientific explanation. Explanations in population biology that appeal to drift are RS explanations. An RS explanation supplies a kind of understanding that a causal explanation of the same result cannot supply. Roughly speaking, an RS explanation shows the result to be mere statistical fallout.

1. Introduction. I argue that really statistical (RS) explanation is a hitherto neglected form of scientific explanation and that, in particular, explanations in population biology that appeal to drift (a.k.a. "genetic drift," "random drift") are RS explanations. RS explanations are noncausal, in contrast to many other familiar scientific explanations (including explanations in population biology appealing to selection for some trait). That some fact has an RS explanation does not preclude its having a causal explanation. But an RS explanation supplies a kind of understanding that a causal explanation of the same result cannot supply. Roughly speaking, an RS explanation shows the result to be mere statistical fallout—that is, just a statistical fact of life.

In section 2, I will introduce RS explanations and contrast them with causal explanations. I will supply several examples of important kinds of RS explanations. In section 3, I will argue that explanations appealing to drift are RS explanations. I will contrast my conception of drift explanations with

Received December 2012; revised February 2013.

*To contact the author, please write to: Department of Philosophy, University of North Carolina, CB 3125, Chapel Hill, NC 27599-3125; e-mail: mlange@email.unc.edu.

†Thanks to Chris Haufe, Alex Rosenberg, and participants at Philosophy of Biology at Madison (2012), Andrea Woody's seminar on scientific explanation at the University of Washington (2011), and the British Society for the Philosophy of Science Special Seminar on Choice and Explanation (Oxford, 2012).

Philosophy of Science, 80 (April 2013) pp. 169–188. 0031-8248/2013/8002-0006$10.00
Copyright 2013 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/670323 Published online by Cambridge University Press

several rival conceptions in the recent literature. On my view, drift is not distinguished from selection by constituting a distinct kind of causal process. But explanations by drift are sharply distinguished from selectionist explanations since drift explanations are noncausal. Selection and drift involve different kinds of explanations, not different kinds of causal processes.

2. RS Explanations. Why in my class did the students with the lowest scores on the first exam tend not to be the students with the lowest scores on the second exam? The explanation could have been that the first exam's worst performers dropped the course before the second exam. The explanation could instead have been that the first exam's worst performers were frightened into working harder and consequently did somewhat better on the second exam. But often neither of these happened. Rather, the explanation is "regression toward the mean."

How does this explanation work? Suppose there is a statistical relation rather than a perfect correlation between two variables—for example, the outcomes of two tests. Insofar as one student scored lower than another on the first test, the former student likely (but not certainly) scored lower on the second test. In any such case, extreme scores in one variable tend to be associated with less extreme scores in the other variable (so there tends to be "regression toward the mean" from the extremes). Accordingly, just as students who did worst on the first exam tend not to do worst on the second exam, so for the same reason students who did best on the first exam tend not to do best on the second exam. This explanation does not propose that the first exam's best performers were complacent in preparing for the second exam. Rather, this explanation portrays the result being explained as just (in the words of one textbook) "a statistical fact of life" (Gravetter and Wallnau 2009, 536).

Of course, any phenomenon explained by regression toward the mean can also be explained by the particular causes that were responsible for moving various individuals toward the mean. Perhaps one student who did worst on the first exam happened to receive several questions on the second exam concerning a topic she understood; another student may have gotten luckier with her guesses on the second exam; perhaps a third was hampered by sickness during the first exam but not the second. This explanation treats the students individually, whereas regression toward the mean proceeds at the level of the population as a whole. Furthermore, regression toward the mean unites the best first-exam performers with the worst, supplying a common explanation of both groups' fates on the second exam—unlike an explanation that describes separately the causes at work on each student. Similarly, regression toward the mean unites the case of the second exam with other cases that can receive the same kind of explanation, such as the less ex-

https://doi.org/10.1086/670323 Published online by Cambridge University Press

ceptional heights of children who have exceptionally tall or short parents—Galton's (1886) original example of regression toward the mean. These cases plainly cannot be given the same sort of explanation in terms of the separate causes acting on the individuals involved.

The explanations describing the separate causes influencing various individuals may themselves be statistical explanations. For instance, suppose a fair coin is tossed 100,000 times; consider the various runs of 20 consecutive tosses beginning with toss numbers 1, 11, 21, ... Neighboring runs share 10 tosses, so a run with more than 10 heads tends to be followed by another run with more than 10 heads. Nevertheless, a run with an exceptionally high number of heads tends to be followed by a run with fewer heads. This result could be explained by regression toward the mean. Alternatively, it could be explained by the coin's 50% chance of landing heads on any given toss (independently of the outcomes of other tosses): we could compute the chance that a run with an exceptionally high number of heads will be followed by a run with fewer heads and find the chance to be high (e.g., the chance that a run of 20 heads will be followed by a run with fewer heads is just the chance of at least one tails appearing in 10 tosses, which is $1 - (.5)^{10} = 1,023/1,024$).

These two explanations are both "statistical." However, they are fundamentally dissimilar. The explanation that does not appeal to regression toward the mean is causal, whereas the regression explanation is not—in a sense that is commonly used to distinguish causal from noncausal explanations. Let's look at this distinction carefully.

The nonregression explanation appeals to the coin's fairness—that is, to the propensity of the coin (together with the coin-tossing mechanism and background conditions) to produce a "heads" outcome. This propensity is a (nonsurefire) disposition. Philosophers differ on whether dispositions themselves qualify as causes of their manifestations.[1] But whether or not a disposition is a cause of its manifestation, an explanation of some result that works by characterizing the result as manifesting some disposition is causal in the broad sense that it derives its explanatory power from describing relevant features of the result's causal history or, more broadly, of the world's network of causal relations.[2] For this reason, Lewis (1986, 220) says that a causal explanation is given when the question "Why was the CIA agent there when

1. The orthodox view (Prior, Pargetter, and Jackson 1982) is that a disposition does not help to cause its manifestation. Sober (1984, 77–78 and 84) and Shapiro and Sober (2007, 253) see this view as the wrong conclusion to draw from Molieresque "dormitive virtue" considerations.

2. Jackson and Pettit (1992, 10) make a similar claim. An explanation appealing to an omission may likewise be causal even if omissions are not causes (since they are not

https://doi.org/10.1086/670323 Published online by Cambridge University Press

His Excellency dropped dead?" is answered by "Just coincidence, believe it or not." Despite identifying no causes, this explanation is causal because it explains by virtue of describing relevant features of the result's causal history. In this case, the relevant feature is that the causes of the CIA agent's presence in the country at that time have nothing relevant in common (such as a plot) with the causes of His Excellency's death. (They have in common the big bang and the presence of oxygen in the atmosphere above the country, but neither of these is relevant in the given context.) Notice that an explanation need not describe the result's actual causal history in order to qualify as causal, as long as it explains by virtue of supplying relevant information regarding the world's network of causal relations. For instance, Jackson and Pettit (1992) correctly say that the water's boiling in the flask explains the flask's cracking by virtue of the fact that had the given collision between a water molecule and the flask not occurred (or not caused a crack to form), another collision would almost certainly have done so. The preempted "cause" did not cause the cracking, and so information about it does not describe the crack's actual causal history. But information about it explains by virtue of describing the world's causal network (telling us what that network would have been like, had it been different in a certain way). That is, information about the preempted "cause" (which we learn by being told that the water is boiling) describes the causes at work in (actual or hypothetical) cases like the given case.

Consider the above explanation of the coin-toss result that does not appeal to regression toward the mean. It is causal in that it works by describing the relevant features of the result's causal history: the chance of any toss's landing heads and the way those chances combine to yield the chance that an extreme run will be followed by a less extreme one. Since this explanation is causal, it could be deepened by being supplemented with descriptions of the causal factors responsible for giving the setup a 50% chance of yielding heads, such as the coin's mass distribution and the coin-tossing mechanism.

In contrast, the explanation by regression to the mean is not causal. It depicts the result as fallout from the statistical character of the case—not from the 50% chance of a toss's landing heads, not even from the chances of a 10-toss run's having various numbers of heads, but rather from the mere fact that there is a statistical association between the outcomes of overlapping 20-toss runs. This explanation is not deepened by being supplemented with descriptions of the coin or the tossing mechanism. These facts have no place in the explanation since the explanation does not derive its

events); such an explanation may work by describing the outcome's causal history. This is Brandon's (2006, 321) reason for characterizing an explanation appealing to omissions as "causal" even though it cites no causes.

https://doi.org/10.1086/670323 Published online by Cambridge University Press

power to explain from its describing relevant features of the result's causal history. The point of the explanation is instead to exhibit the result as arising from the fact that successive runs have a statistical relation—regardless of that precise relation or its (perhaps probabilistic) causes or, indeed, whether it has any causes at all.[3]

Accordingly, I call explanations of this kind "really statistical" (RS) explanations. An RS explanation does not proceed from the particular chances of various results—or even from the fact that some result's chance is high (or low). It exploits merely the fact that some process is chancy, and so an RS explanation shows the result to be just "a statistical fact of life."[4]

There are many kinds of noncausal explanations; not all are RS.[5] For instance, that Mother has 23 strawberries and three children explains why Mother failed to distribute her strawberries evenly (without cutting any) among her children. What makes this explanation noncausal is not that it fails to cite causes; the numbers of children and strawberries happen to be causes of Mother's failure. What makes it noncausal is that it does not work by describing the outcome's causes (or the world's causal network). In this explanation, Mother's having three children helps to explain Mother's failure not by being one of its causes, but by helping to make her success mathematically impossible (Lange, forthcoming).

This noncausal explanation is not RS. But just as this explanation is noncausal despite supplying information about causes, so an RS explanation may happen to supply information about causes. In the coin-toss case, what makes the regression explanation noncausal is neither its failure to exploit the coin's 50% chance of landing heads nor its invoking only something less specific (that there is a statistical association between the outcomes of overlapping 20-toss runs); after all, the explanation of the flask's cracking is causal despite appealing only to the water's boiling, not to a specific molecular collision. What makes the regression explanation noncausal is that

3. In some contexts, the question "Why did the students who did worst on the first exam tend not to do worst on the second exam?" is understood to be demanding a particular kind of explanation: one appealing to causes common to many of the students. Then the correct reply is to reject the question as having a false presupposition (that there is such a cause). But the fact that regression toward the mean fails to answer the why question in such a context does not show that it is never an answer; the why question may lack this presupposition.

4. Statistics textbooks commonly present regression toward the mean as answering why questions, such as "Explain why the rookie of the year in major-league baseball usually does not perform as well in his second season" (Gravetter and Wallnau 2009, 536). Lipton (2009, 622) is the only philosopher I know who has mentioned explanations that appeal to regression toward the mean.

5. Possible varieties of noncausal scientific explanations include structural explanations (Hughes 1989, 256–57), dimensional explanations (Lange 2009a), mathematical explanations (Colyvan 1998, 321–22), and explanations by symmetries (Lange 2009b).

https://doi.org/10.1086/670323 Published online by Cambridge University Press

the explainer (that there is a statistical association between the outcomes of overlapping 20-toss runs) acquires its power so to explain other than by virtue of supplying information about causes. The explanation succeeds whether or not any causal relations generate the statistical association.

Not all RS explanations exploit regression toward the mean. For example, Rutherford and Geiger set out to explain why it is that "in counting the α particles emitted from radioactive substances . . . the average number of particles [emitted] from a steady source is nearly constant, when a large number is counted, [but] the number appearing in a given short interval is subject to wide fluctuations. . . . For example, during a considerable interval it may happen that no α particle appears; then follows a group of α particles in rapid succession; then an occasional α particle, and so on" (1910, 698).[6] Rutherford and Geiger concluded that the departures from the mean in various intervals—and the fact that these departures tend to be smaller in longer intervals—are explained by "the laws of probability and that the α particles are emitted at random" (704) rather than, "for example, that the emission of an α particle [precipitates] the disintegration of neighboring atoms" (698). Once again, that this explanation is RS rather than causal is reflected in the fact that it would not have been deepened by being supplemented with an account of the causal factors responsible for the radioactive sample's rate of α-particle emission. The explanation does not appeal to whether each atom in the sample has the same chance of undergoing radioactive decay. (They do not; the sample was a mixture of many polonium isotopes.) The explanation does not appeal to whether a given sample's activity depends on its temperature (it does not). The explanation does not even appeal to whether different polonium samples have the same chance of α-particle emission. It appeals simply to a given sample's having a constant chance of emission during the experiment (however, that fact is to be explained) and then uses the "laws of probability" to explain the departures from the mean number of decays (and those departures' tendency to be greater as the intervals became briefer).

Of course, to explain why a given sample has a constant chance of emission, we must appeal to various causal factors concerning the sample's physical constitution (as well as various natural laws). But the fact that an RS explanation appeals to a fact that itself has only a causal explanation does not make the RS explanation covertly causal. A scientific explanation is not responsible for explaining the facts in its own explanans (Hempel 2001). Admittedly, the relevant "law of probability" applies to a radioactive sample only because of the sample's physical constitution, which accounts for its constant chance of emission. But the RS explanation does not appeal to that physical constitution since the RS explanation does not explain why

6. Thanks to Chris Haufe for calling this article to my attention.

https://doi.org/10.1086/670323 Published online by Cambridge University Press

the law of probability applies. Rather, the RS explanation simply appeals to the law's applicability (in this case, the constant chance of radioactive emission) and applies the law.

In short, Rutherford and Geiger explained the departures from the mean as mere artifacts of the statistics. A causal explanation that begins with the half-lives of various polonium isotopes and their proportions in the sample, and then deduces the likelihoods of various numbers of emissions over various intervals, would also explain the observed result.[7] But all of this information about the result's causal history is beside the point when the point is to show that the result reflects the fact that the emission process is statistical, whatever the specific half-lives, proportions, and other causal details may be. This RS explanation unifies the radioactive-decay case with other cases involving fundamentally dissimilar causal mechanisms. The reason why Rutherford and Geiger recorded different numbers of emissions during different 10-minute periods is the same kind as the reason why different runs of 20 consecutive tosses of a fair coin resulted in different numbers of heads—and in both cases, there is also the same kind of reason why longer runs tend to give results closer to the mean rate. Notice that an RS explanation does not have to show that the phenomenon to be explained is certain or even likely given merely that it arises from a statistical process. An RS explanation can explain why a particular run of the experiment yielded a result extremely far from the mean value. The RS explanation is, in effect, that "it's just statistics."[8]

However, an RS explanation says a good deal more about the explanandum than merely that "it's just statistics." An RS explanation identifies the explanandum as an instance of some particular kind of behavior that is characteristic of statistical systems in virtue of their being statistical. For instance, the RS explanation of the fact that the students with the lowest scores on the first exam tend not to be the students with the lowest scores on the second exam is that the explanandum is an instance of regression toward the mean. The RS explanation of the fact that 10 of a given

7. Noncausal explanations do not preclude causal explanations, as Salmon (1989, 181–84) emphasizes. However, some philosophers who have described explanations that I would characterize as RS may disagree. For instance, regarding an explanation of why samples' departures from expectation are more likely as samples become smaller, Walsh (2007, 284, cf. 288–89) says that "the effect is explained by the statistical properties of the setup and not by the causal properties of the apples" (in this case, the weights of the various apples in the populations).

8. Some such slogan is quite common. For instance, consider Galton's discovery that over those cases in which the two parents' mean height is between 70 and 71 inches (average human height was 68.2 inches), the mean children's height is 69.5 inches. The RS explanation in terms of regression toward the mean shows that "this is a statistical, not a genetic phenomenon" (Bland and Altman 1994, 1499).

https://doi.org/10.1086/670323 Published online by Cambridge University Press

philosophy department's 15 graduate applicants accepted their offer of admission—a much higher rate than in previous years—might be that it is a statistical fluke: when a statistical process is run many times, an occasional run may have a result quite far from the mean value. An RS explanation does not have to deem the explanandum to be likely given the explanans, but it must identify some particular signature of statistical processes that the explanandum exemplifies.

Some signatures have well-known names, such as "regression toward the mean," but others do not. There is no moniker for the fluctuations (observed by Rutherford and Geiger) in the frequencies of various outcomes in runs of repetitions of a given experiment, or for the considerable departures of some of these frequencies from their mean values, or for the tendency of departures from the mean value to diminish as the runs lengthen. Nevertheless, all of these familiar behaviors are characteristic of runs of independent, identically distributed trials; the runs exhibit these features solely by virtue of their statistical character (regardless of its causal basis, if any). These behaviors can thus receive RS explanations.[9]

RS explanations are especially illuminating when the phenomenon being explained does not initially appear to be merely a matter of statistics. For example, suppose that for 20 successive days, two students are given a different problem to solve daily, where the first student to solve the day's problem will win one point. The student in the lead at the end of the 20-round contest will win a prize. Suppose we run this contest for many pairs of students, and in more than half of the contests, there is one student who is either never in the lead or in the lead just once. It would be tempting to conclude that this is because students who fall behind early tend to lose heart and slacken off or because in every contest, one student is much more talented than the other. But the correct explanation may in fact be that the outcome of the contest is the product of a "random walk." Regardless of the difference between the two students' abilities (i.e., between their chances of winning a point in any round)—even if there is no difference between their abilities—it is more likely than not that one student will be behind in the contest only once or never, when a given student's chance of winning a

9. Of course, the RS reason why a run of 200 coin tosses has only 20 heads (where the mean outcome is 100 heads) is the same as the RS reason why a run of 200 coin tosses has only 5 heads: such wild fluctuations occasionally happen with runs of independent, identically distributed trials. That both outcomes receive the same RS explanation should not disturb us. It has long been widely (although not universally) accepted that different outcomes can receive the very same causal statistical explanation. For instance, if Jones contracts paresis, then the considerations responsible for Jones's likelihood of contracting paresis (such as his having latent untreated syphilis) will explain why he does, and if Jones does not contract paresis, then the same considerations will explain why he does not (Salmon 1989, 62). I see no reason why different possible outcomes cannot likewise receive the same RS explanation.

https://doi.org/10.1086/670323 Published online by Cambridge University Press

given round is the same in every round.[10] Surprisingly, this "one-sidedness" is a characteristic feature of random walks.[11] To explain the result as the product of a random walk is like explaining some phenomenon as a case of regression toward the mean. Both explain the result as a characteristic kind of statistical fallout.

The RS explanation of the improvement in the scores of the worst-performing students does not say merely that this improvement was just "a statistical fact of life." It goes on to identify the result as a case of regression toward the mean in particular. Thus, the RS explanation does not say merely that the result has no causal explanation appealing to some cause common to many of the students. Rather, it supplies a genuine explanation by identifying the particular signature of statistical processes that the explanandum exemplifies. That is why RS explanations are characterized in science (see n. 4) as answering why questions, not merely as rejecting some.

An RS explanation may proceed without appealing to any laws of nature—unlike statistical explanations that require laws to associate a given setup with various chances of yielding various possible outcomes. An RS explanation does not appeal to the particular propensities at work but merely to the fact that there are chances. Consequently, instead of subsuming the result to be explained under a statistical law of nature, an RS explanation exploits a theorem of the probability calculus (what Rutherford and Geiger called "the laws of probability").[12]

An RS explanation does not proceed by the brute-force approach of putting everything in and computing everything out (as in the nonregression explanation of the coin-toss result)—that is, by calculating the chance of the result being explained from more basic chances (e.g., the chance of a toss's landing "heads"), after perhaps inferring those chances from the setup conditions (e.g., the coin's mass distribution) and the natural laws by which those conditions are associated with some propensities. Rather, an explanation is RS if and only if it works by identifying the result being explained as an instance of some characteristically statistical phenomenon such as regression

10. In the most evenly matched case (i.e., when each student has a 50% chance of winning any given round), the chance that one player will be in the lead only once or never is 0.5379 (Feller 1957, 79).

11. The RS explanation does not have to include the mathematical derivation showing this "one-sidedness" to be a characteristic feature of random walks. The RS explanation appeals merely to the fact that it is such a feature. Similarly, I argue (Lange, forthcoming) that "distinctively mathematical" explanations in science (such as the explanation I gave of Mother's failure to distribute the strawberries evenly) include mathematical facts but not their proofs.

12. Of course, an RS explanation does not appeal solely to mathematical necessities. It is a contingent fact that a given system is statistical—e.g., that the students' scores on the two exams are correlated statistically.

https://doi.org/10.1086/670323 Published online by Cambridge University Press

toward the mean, departure from the expectation value, or the one-sidedness of random walks. Their working in this way distinguishes RS explanations from causal explanations and is responsible for the other features setting RS explanations apart (such as their exploiting theorems of the probability calculus in place of natural laws). Whereas the brute-force approach can be applied to explain any result of statistical processes, an RS explanation exists only when the phenomenon to be explained arises from the mere fact that statistical processes are at work.

3. Drift. Population biology calls on drift to explain many kinds of phenomena. It explains why the frequencies of various traits (or genotypes or genes) in a given population experiencing no mutations, migration, or selection pressure nevertheless change over time. It explains why a population's frequencies depart from their expectation values (which reflect the selection pressure on the population) and why these departures tend to be greater for smaller populations. It explains why different populations of the same size, with the same initial frequencies, and experiencing the same selection pressures (and no mutations or migrations) nevertheless have different frequencies at the end of a given period—and why these differences tend to be greater for smaller populations. Drift explains all of these results as "just statistics"—that is, as "sampling error" (whether drift enters the picture when certain organisms but not others become parents, or certain gametes but not others from a given parent contribute to a zygote, or in some other way). Explanations by drift are RS explanations.[13]

To see how this characterization of drift explanations can help us to understand them better, I briefly contrast my view with some of the other ways that drift explanations have recently been characterized. It might initially appear that "drift" refers to all and only departures of a population's frequencies from their expectation values. However, it is commonly noted (e.g., Beatty 1992, 36; Plutynski 2007) that if drift consists of such departures, then drift cannot explain these departures (as population biologists call on it to do). Accordingly, some philosophers identify drift as the propensities of a population to have frequencies departing to various degrees from their expectation values: "In a finite series of trials, there is a spread of possible outcomes, each with its own probability. Drift is simply the uncertainty associated with this spread of possibilities" (Matthen 2009, 484).

13. Matthen (2009) has recently proposed that statistically abstractive (SA) explanation is a hitherto neglected form of explanation and that explanations by drift are SA explanations. An SA explanation is not an RS explanation. Rather, it is a statistical explanation of the usual sort, its only distinction being that it derives the chance of the fact being explained by placing the case in a reference class that is not objectively homogeneous. In doing so, it ignores certain statistically relevant factors on the grounds that they are uninteresting to a particular theory in play (such as population genetics).

https://doi.org/10.1086/670323 Published online by Cambridge University Press

On this view, the ground (i.e., the disposition's "categorical base" or "causal basis") for the propensities associated with drift is no different from the ground for the propensities associated with natural selection. Indeed, drift and selection are not distinct causal processes, and an explanation appealing to drift is no different from an explanation appealing to selection—except for the kind of outcome being explained: "Drift explains what natural selection cannot in the sense that drift accounts for those differences between the actual outcome of a series of births, deaths, and reproductions and the outcome predicted by differences in trait fitness" (Walsh, Lewens, and Ariew 2002, 465). "Natural selection and drift are not distinct processes working on a finite population, but mathematically connected aspects of the same accumulation" (Matthen 2009, 484). On this view, neither drift nor selection for some trait is a causal process. Drift consists just of the chances for various frequencies to depart to various degrees from their expectation values through births and deaths resulting from various separate processes affecting individual organisms.

Many philosophers (e.g., Millstein 2006; Shapiro and Sober 2007) have criticized this view as failing to recognize selection for some trait as a unified causal process acting on a population as a whole—in other words, for failing to understand that "selection *for* is the causal concept *par excellence*" (Sober 1984, 100).[14] (As we will see, these critics generally regard drift and natural selection as distinct causal processes.) More relevant to our concern with drift is that the view expressed by Matthen, Walsh, et al. fails to account for those cases in which drift explains a phenomenon that does not consist of a departure from expectation values. For example, consider a population in which some rare, widespread catastrophe (e.g., an epidemic) wipes out many individuals and "selects" without regard to their heritable traits but leaves the population with the same frequencies as before—

14. To defend this view properly would take me too far afield. I do wish to note, however, that selection for some trait is a causal process (as Shapiro and Sober [2007, 256] call it in a passage I quote shortly), not a cause. In selection, various features of the population undergoing selection and of its environment are causes of the outcome frequencies. Selection (like diffusion, erosion, heat flow, and electrical conduction) is a process involving population-level causes and effects, the result being that certain traits are "selected" over others so that a new set of trait frequencies is the effect that is explained (just as temperature differences are causes in the process of heat flow, with a new spatial distribution of heat as the effect to be explained, and potential differences are causes in electrical current flow, pressure differences in diffusion, and so forth). To say that "selection for properties causes differences in survival and reproductive success" (Sober 1984, 100) is to specify the causal process at work, not to say that selection is a cause in place of or alongside these features of the population and environment. This can easily lead to confusion. To cite one example, McLaughlin (2007, 279–80) interprets Sober's remark above as mistakenly treating selection as a cause and then goes on to insist that selection "just is (nonaccidental, differential) reproductive success."

https://doi.org/10.1086/670323 Published online by Cambridge University Press

namely, with frequencies coinciding with their expectation values. Population biologists would attribute the outcome (including the population crash and the resulting frequencies) to drift. This explanation by drift is an RS explanation: the occurrence of such a catastrophe (like a period of exceptionally frequent α-particle emission from Rutherford and Geiger's radioactive sample) is just a statistical fluke that is virtually bound to happen from time to time in the long run. Of course, the outcome also has a causal explanation—namely, that an epidemic struck the population—and there are even causal explanations of why certain individuals but not others were victims (although these causal explanations are not selectionist since differences in heritable traits made no difference to which creatures survived the catastrophe).

Let me give another kind of example in which drift explains a phenomenon that does not consist entirely of a population's departure from expectations. Consider two populations of the same (finite) size and with the same initial frequencies that are experiencing the same selection pressures (and no mutations or migrations). Suppose they have different frequencies at the end of a given period: one has frequencies coinciding with their expectation values and the other has frequencies departing from their expectation values. Population biologists would attribute this difference to drift; that is, they would hold drift responsible for explaining why the two populations' frequencies diverged despite their having the same initial conditions and their experiencing the same selection pressures. The same explanation would apply whether both populations' frequencies departed from their expectation values or only one population's did; drift explains their divergence.

The drift explanation is that such variation is characteristic of statistical processes (as Rutherford and Geiger also emphasized). It is an RS explanation. It is not the case that one population in the pair was acted on by selection and the other by drift. Both are subject to the same selection pressures. Indeed, the outcome could be explained by drift or by those selection pressures (since the selection pressures, like a coin's propensity for landing heads, do not determine an outcome). But an RS explanation is noncausal, whereas an explanation in terms of what is being selected for is causal. A selectionist explanation describes the ground of the propensities for various frequencies and so explains by virtue of supplying information regarding the sorts of causes at work in cases with the initial frequencies of and selection pressures on these two populations. By contrast, the RS explanation in terms of drift identifies the result being explained as the same sort of typically chancy behavior as Rutherford and Geiger noted their radioactive source exhibiting: considerable fluctuations in short periods. Thus, in interpreting drift explanations as not appealing to some sort of causal process, I agree

https://doi.org/10.1086/670323 Published online by Cambridge University Press

with Matthen, Walsh, et al.—but without agreeing with them that selection for some trait is likewise not a causal process. On my view, a selectionist explanation of the two populations' divergence is like a causal explanation of why runs of 20 coin tosses with an exceptional frequency of heads tend to be followed by less exceptional runs—namely, an explanation that begins by describing the causal basis of the 50% chance of any toss's landing heads. In contrast, an explanation by drift is like an explanation of the coin-toss outcome as regression toward the mean.

Here is another example in which (contrary to Matthen, Walsh, et al.) drift is invoked to explain a result that does not consist just of a population's departure from expectation values. Consider a population under constant selection pressure with frequencies that in the short run depart from their expectation values but in the long run largely coincide with their expectation values. Again, drift explains this behavior—but not because drift explains only the short-run result (i.e., only the departures from expectation). Rather, an RS explanation appealing to drift explains the difference between the short- and long-run results as characteristic of statistical processes. (Rutherford and Geiger observed this feature, too.) This drift explanation cannot be accommodated by the view that "a series of births, survivals, deaths, and reproductions manifests drift just if the outcome—measured as changes in trait frequencies—diverges from that predicted by differences in fitness" (Walsh et al. 2002, 459). Note the "just if."

Of course, this outcome can also be explained causally: in terms of the selection pressure. As we have seen, RS explanations and causal explanations work differently and respond to different interests that we may have in asking "Why?" As I understand it, a population biologist who wants to know whether the difference between the population's short-run and long-run frequencies is a matter of drift or a matter of selection is asking whether the selection pressure early in the period under study was different from the selection pressure during the rest of the period. She is asking about causes. But she is not presenting drift as a possible causal explanation—a rival to selection for some trait. Rather, in asking whether the difference between the short-run and the long-run results is explained by drift or selection, she is asking whether this difference has an RS explanation by drift or whether some difference in selection pressure accounts for the difference in frequencies. These are rivals, whereas a causal explanation for the population's short- and long-run frequencies that appeals to various unchanging selection pressures is compatible with an RS explanation by drift for the difference between these frequencies.[15]

15. Compare "it was drift, not selection" to "this is a statistical, not a genetic phenomenon" quoted in n. 8.

https://doi.org/10.1086/670323 Published online by Cambridge University Press

Suppose the answer to the population biologist's "Drift or selection?" question regarding the difference between the population's short- and long-run frequencies is "Drift, since the degree of selective advantage afforded by various traits was unchanged throughout the period." This answer plainly does not require that no trait have been selected for during the period. This drift explanation fails to accord with the idea championed by Millstein (2002; cf. Hodge 1987, 252–53) that drift always consists of a sampling process that is indiscriminate (in that heritable physical differences are causally irrelevant to differences in reproductive success). Of course, the distinction between discriminate and indiscriminate sampling is genuine, and some phenomena explained by drift result from indiscriminate sampling (as in the epidemic example). However, as many philosophers have noted (e.g., Brandon 2005; Matthen 2009, 467), a product of discriminate sampling can also be explained by drift. For example, as I mentioned, drift explains why two populations of the same size, with the same initial frequencies, and experiencing the same selection pressures (and no mutations or migrations) have different frequencies at the end of some period. Another example of drift with discriminate sampling is Beatty's famous case (1984, 195) in which dark moths in a given period are eaten by predators disproportionately because, although trees with dark bark are more common than light trees and consequently dark moths are fitter than light, "the dark moths chanced to land on light trees more frequently than on dark trees," whereas the light moths landed on a representative sample of trees. The sampling here is discriminate since predators are more likely to eat moths that contrast with their backgrounds.[16] On my view, the outcome in this period has an RS explanation in terms of drift because it constitutes just the sort of fluctuation that is characteristic of statistical processes. It also can be given a causal explanation in terms of the selection pressure—or, more deeply, in terms of the setup's physical features that are responsible for the selection pressure (e.g., that 60% of the trees have dark bark).

Of course, some paradigm instances of drift are cases of indiscriminate sampling, as Millstein emphasizes. For example, according to Kimura's "neutral theory of molecular evolution," "the overwhelming majority of evolutionary changes at the molecular level [result from] random fixation (due to random sampling drift in finite populations) of selectively neutral (i.e., selectively equivalent) mutants" (1991, 367). On my view, once again, when

16. By adding another population of moths in which both light and dark moths happened to land on representative samples of trees, we could turn Beatty's example into my earlier example involving two populations with the same initial frequencies under the same selection pressures but with frequencies that have diverged by the end of a given interval.

https://doi.org/10.1086/670323 Published online by Cambridge University Press

drift explains the fixation of some selectively neutral mutant, the explanation is RS because it identifies the result as just the sort of fluctuation that is characteristic of statistical processes. (The explanation thus characterizes the fluctuation as akin to a period in which Rutherford and Geiger's radioactive sample undergoes an extreme rate of α-particle emission. Recall that the explanans in an RS explanation need not make the explanandum likely.) The result would still involve this sort of fluctuation even if the mutant were very slightly deleterious, for example. However, on Millstein's view, the result would then not be the product of drift at all since sampling would be discriminate rather than indiscriminate. I agree with Brandon (2005) that discriminate sampling does not differ qualitatively from indiscriminate sampling. Accordingly, drift explanations should be available in either case.

One motivation for Millstein's identification of drift with indiscriminate sampling (and selection with discriminate sampling) is that otherwise, it is difficult to see how drift and selection can be distinct kinds of population-level causal processes. Accordingly, Brandon (2005, 167) concludes that "there is a single process, sampling," and so "the outcome, deviation or lack thereof from expectation, is a necessary component of our concepts of drift and selection," respectively. But I have already argued that drift need not be associated with departure from expectation values. Shapiro and Sober (2007) suggest that drift and selection are distinct population-level causal processes (just as Millstein says) but that (contrary to Millstein) drift does not require indiscriminate sampling. Rather, drift occurs whenever a population has a finite effective size (255). What, then, sets drift apart from selection? Shapiro and Sober say: "We view selection and drift as distinct processes whose magnitudes are represented by distinct population parameters (fitnesses on the one hand, effective population size on the other). Changes in each of these parameters will be associated with changes in the probabilities of different outcomes. If you intervene on fitness values while holding fixed population size, this will be associated with a change in the probability of different trait frequencies in the next generation. And the same is true if you intervene on population size and hold fixed the fitnesses" (256).

In short, according to Shapiro and Sober, even if drift cannot be eliminated from a population undergoing selection, drift remains distinct from selection. Drift, unlike selection, is stronger insofar as population size is smaller. Abrams agrees: "drift and selection are distinct causal factors because they can be independently manipulated by changing, respectively, population size and fitness differences" (2009, 667). Several other philosophers (e.g., Reisman and Forber 2005) have joined Shapiro and Sober in arguing that since population size can be manipulated to yield predictable differences in the chances of various frequencies, drift is a causal process. In addition to manipulability, some philosophers (e.g., Millstein 2006, 632) have also appealed

https://doi.org/10.1086/670323 Published online by Cambridge University Press

to counterfactual dependence and probability raising as signs of causal relations. On all of these grounds, philosophers have argued that in drift, "the size of the population can be said to play a causal role" (631) and so that drift is a causal process distinct from selection.[17]

A fundamental problem with these arguments is that the explanatory power of effective population size in drift explanations is mediated not by any laws of nature, but merely by the "laws of probability"—that is, by math.[18] No causal law is responsible for making greater departures from expectation values more likely as population size diminishes. Given the chances of various possible outcomes on any particular independent trial, the relation between population size and the likelihoods of various population-wide outcomes is not contingent.[19] For all of its manipulationist, counterfactual-dependence, and probability-raising credentials, population size does not act as a population-level cause in drift explanations since these credentials are mathematically necessary, not beholden to any mere law of nature.[20] This feature of drift explanations fits well with their interpretation as RS explanations since RS explanations employ theorems of the probability calculus in place of statistical laws of nature.

17. Sober (personal communication) currently believes that drift and selection are not distinct causal processes; rather, there is only one population-level causal process at work. However, drift and selection involve different causes co-occurring (population size and fitness differences, respectively) that jointly influence the trait frequencies, just as in a series of coin tosses, there is only one causal process by which heads and tails are accumulating, but the coin's bias and the number of tosses are distinct causes, separately manipulable.

18. There are other problems with these arguments, too. In some cases, selection is stronger insofar as population size is smaller (see Sober 2001).

19. The relevant theorem of the probability calculus is that in n independent trials where the chance of a given kind of outcome in any given trial is p, the chance of i such outcomes is $[n!/i!(n-i)!]p^i(1-p)^{n-i}$.

20. However, Sober (2011) believes that although the fitnesses of various traits in a population stand in metaphysically necessary connections to the traits' expected frequencies, they are causes of those frequencies; no contingent law of nature is needed to mediate between a cause and its effect. (This may not represent a departure from Sober's prior view of fitnesses as "causally inert" [1984] since Sober there was arguing that fitnesses are not causes of the survival and reproduction of individual organisms; he did not address whether they are causes of trait frequencies in a population.) Sober (2011) gives manipulationist reasons for thinking fitness to be a cause of trait frequency, and he says that a contingent law is not needed to mediate a causal relation because (appealing to Davidson) it is an analytic truth that E's cause would cause E (i.e., if E occurs and has a cause, then E's cause caused E). However (I reply), this truth's analyticity does not show that a causal relation requires no contingent law to connect cause and effect since "contingent" here should be contrasted with "metaphysically necessary" (a modal matter), not with "analytic" (a conceptual matter). Although "E's cause" may be analytically connected to "E," it does not follow that the event that is E's cause is connected to E by metaphysical necessity. In contrast, that trait A is fitter than trait B is connected with metaphysical

https://doi.org/10.1086/670323 Published online by Cambridge University Press

Sober (1984, 115) says that coin tossing "affords the same distinction" as selection versus drift, revealing why they are distinct causal processes. Just as selection cannot occur without drift, so a coin flipped with some chance of landing heads must be flipped some definite number of times. Nevertheless, the distinction remains: "Two coins may have the same chance of landing heads but may be tossed a different number of times. Two other coins may be tossed the same number of times but may differ in their chances of landing heads. The first two coins have something in common, and so do the second" (115). Of course, this last remark is correct. But I do not think that the coin example shows drift and selection to be distinct causal processes any more than a series of coin tosses consists of several causal processes running simultaneously.

The coin-toss example also illustrates my claim that the number of coin tosses is not a cause of the resulting frequency of heads because no contingent law of nature gives it its alleged "causal role." Suppose we aim to explain why the coin's percentage departures from its expected frequency of heads have generally been greater in runs of 50 tosses than in runs of 50,000 tosses. We could explain this fact by the brute-force approach of using the coin's chance of landing heads on an individual toss to compute the chances of various great departures from the frequency's expectation value in runs of 50 and 50,000 tosses. We could deepen this causal explanation by replacing the propensity with its ground in the setup, including the coin's mass distribution and the tossing mechanism. Then the explanation would require laws of nature to associate this setup with a propensity for (for example) a head. These are causal explanations. But the role in these explanations played by the number of coin tosses in the run (which enters after a law has associated the setup with a toss's chance of yielding a head) is beholden not to any mere law of nature but entirely to a theorem of the probability calculus. The run's length is thus not a cause.[21]

---

necessity to a relation among A's and B's expected frequencies. That connection holds in virtue of what fitnesses essentially are, whereas the event that is E's cause is not essentially so. Furthermore, I do not see how fitness's manipulationist credentials do anything to undermine Sober's previous (1984) argument that fitness is a mere actuarial property reflecting various potential causal factors that may not actually have come into play and so is unsuited to being a cause. That argument, in my view, precludes fitnesses from being causes even of trait frequencies. (Of course, I agree with Sober [1984] that fitness can still figure in causal explanations since it can describe the world's causal relations without figuring in any itself.) For more, see Lange and Rosenberg (2011).

21. Walsh (2007) has also recently argued that the number of tosses is not a cause. Here is his argument, as I understand it. Consider 100 fair coin flips resulting in roughly 50% heads, although in the 10 runs of 10 successive flips each, there were significant departures from 50% heads. Why is that? A drift explanation appeals to 10 flips being much smaller than 100 flips. But if the number of flips is a cause, then (Walsh says) we would have to say that drift as a cause is strong in each subpopulation but weak in the

https://doi.org/10.1086/670323 Published online by Cambridge University Press

Suppose an explanation instead sets aside the coin's particular chance of landing heads on an individual toss as well as the fact that we are dealing with runs of 50 and 50,000 tosses—and even that we are dealing with coin tossing in particular. The explanation then proceeds not by describing relevant features of the outcome's causal history, but rather by characterizing the fact being explained as an instance of a general, characteristic feature of statistical processes: outcomes are more likely to adhere more closely to expectation values in the longer than in the shorter run. This explanation is noncausal; it computes no chances regarding the case at hand; it is mediated not by any ordinary laws of nature, but rather by a theorem of probability. This is the kind of explanation that drift supplies.[22]

On my view, one instance of drift may involve indiscriminate selection (one kind of causal process), whereas another instance involves discriminate selection (a causal process of another kind). Drift is not distinguished from selection by constituting a distinct kind of causal process. But explanations by drift are sharply distinguished from selectionist explanations since drift explanations are noncausal. That, in every case of drift, some population-level causal process is at work does not entail that there is some distinct kind of population-level causal process that constitutes drift or that a drift explanation is a causal explanation. What is distinctive about drift, I have suggested, is not the outcome or the causal process but the kind of noncausal explanation it supplies.

---

population as a whole—which is a "contradiction" (296) since the 100-flip population is composed of the 10 subpopulations. However, I see no contradiction in taking the number of tosses as a cause: that there were only 10 tosses could be a cause of the outcome for a given subpopulation but plainly is not supposed to be a cause of what happens in the overall population. Walsh generates a contradiction only by turning the number of tosses into the strength of a "force": drift is weak for the overall population but strong in each subpopulation, which is a contradiction ("drift is objectively both weak and strong in the population" [296]) considering that the overall population comprises the subpopulations. But without treating drift as a force (as philosophers sometimes do metaphorically) acting on each subpopulation (and perforce the overall population), Walsh generates no contradiction; at least, I see no contradiction arising simply from $n = 10$ being a cause for a subpopulation but $n = 100$ being a cause for the population as a whole. (Matthen [2009, 409] gives an argument similar to Walsh's: "How can strong nondirectional (and therefore noncancelling) forces operating on parts of a population give rise to a weak force operating on the whole?")

22. That an explanation by drift cites no causes does not entail that it is not a causal explanation. Brandon (2006, 329) says that drift explanations are causal precisely because they describe what happens by default—i.e., in the absence of causes making it happen otherwise (see n. 2). I have argued that drift explanations are noncausal because they are RS, not because they cite no causes.

https://doi.org/10.1086/670323 Published online by Cambridge University Press

REFERENCES

Abrams, Marshall. 2009. "How Do Natural Selection and Random Drift Interact?" *Philosophy of Science* 74:666–79.
Beatty, John. 1984. "Drift and Natural Selection." *Philosophy of Science* 51:183–211.
———. 1992. "Random Drift." In *Keywords in Evolutionary Biology*, ed. Evelyn Fox Keller and Elisabeth Lloyd, 273–81. Cambridge, MA: Harvard University Press.
Bland, J. M., and D. G. Altman. 1994. "Statistic Notes: Regression." *British Medical Journal* 308:1499.
Brandon, Robert. 2005. "The Difference between Drift and Selection: A Reply to Millstein." *Biology and Philosophy* 20:153–70.
———. 2006. "The Principle of Drift: Biology's First Law." *Journal of Philosophy* 103:319–35.
Colyvan, Mark. 1998. "Can the Eleatic Principle Be Justified?" *Canadian Journal of Philosophy* 28:313–36.
Feller, William. 1957. *An Introduction to Probability Theory and Its Applications*. 2nd ed. Vol. 1. New York: Wiley.
Galton, Francis. 1886. "Regression towards Mediocrity in Hereditary Stature." *Journal of the Anthropological Institute of Great Britain and Ireland* 15:246–63.
Gravetter, Frederick, and Larry Wallnau. 2009. *Statistics for the Behavioral Sciences*. 8th ed. Belmont, CA: Wadsworth.
Hempel, Carl G. 2001. "Science Unlimited?" In *The Philosophy of Carl G. Hempel*, ed. James Fetzer, 329–44. New York: Oxford University Press.
Hodge, M. J. S. 1987. "Natural Selection as a Causal, Empirical, and Probabilistic Theory." In *The Probabilistic Revolution*, vol. 2, ed. Lorenz Krüger, Lorraine Daston, and Michael Heidelberger, 233–70. Cambridge, MA: MIT Press.
Hughes, R. I. G. 1989. *The Structure and Interpretation of Quantum Mechanics*. Cambridge, MA: Harvard University Press.
Jackson, Frank, and Philip Pettit. 1992. "In Defense of Explanatory Ecumenism." *Economics and Philosophy* 8:1–21.
Kimura, Mootoo. 1991. "The Neutral Theory of Molecular Evolution: A Review of Recent Evidence." *Japanese Journal of Genetics* 66:367–86.
Lange, Marc. 2009a. "Dimensional Explanations." *Noûs* 43:742–75.
———. 2009b. "A Tale of Two Vectors." *Dialectica* 63:397–431.
———. Forthcoming. "What Makes a Scientific Explanation Distinctively Mathematical?" *British Journal for the Philosophy of Science*.
Lange, Marc, and Alex Rosenberg. 2011. "Can There Be *A Priori* Causal Models of Natural Selection." *Australasian Journal of Philosophy* 89:591–99.
Lewis, David. 1986. "Causal Explanation." In *Philosophical Papers*, vol. 2, 214–40. New York: Oxford University Press.
Lipton, Peter. 2009. "Causation and Explanation." In *The Oxford Encyclopedia of Causation*, ed. Helen Beebee, Christopher Hitchcock, and Peter Menzies, 619–31. Oxford: Oxford University Press.
Matthen, Mohan. 2009. "Drift and 'Statistically Abstractive Explanation.'" *Philosophy of Science* 76:464–87.
McLaughlin, Brian. 2007. "On Selection Of, For, With, and Against." In *Thinking about Causes*, ed. Peter Machamer and Gereon Wolters, 265–83. Pittsburgh: University of Pittsburgh Press.
Millstein, Roberta. 2002. "Are Random Drift and Natural Selection Conceptually Distinct?" *Biology and Philosophy* 17:33–53.
———. 2006. "Natural Selection as a Population-Level Causal Process." *British Journal for the Philosophy of Science* 57:627–53.
Plutynski, Anya. 2007. "Drift: A Historical and Conceptual Overview." *Biological Theory* 2: 156–67.
Prior, Elizabeth, Robert Pargetter, and Frank Jackson. 1982. "Three Theses about Dispositions." *American Philosophical Quarterly* 19:251–57.
Reisman, Kenneth, and Patrick Forber. 2005. "Manipulation and the Causes of Evolution." *Philosophy of Science* 72:1113–23.

https://doi.org/10.1086/670323 Published online by Cambridge University Press

Rutherford, Ernest, and Hans Geiger. 1910. "The Probability Variations in the Distribution of α Particles." *Philosophical Magazine* 6 (20): 698–707.

Salmon, Wesley. 1989. "Four Decades of Scientific Explanation." In *Scientific Explanation*, ed. Philip Kitcher and Wesley Salmon, 3–219. Minnesota Studies in the Philosophy of Science 13. Minneapolis: University of Minnesota Press.

Shapiro, Larry, and Elliott Sober. 2007. "Epiphenomenalism: The Dos and the Don'ts." In *Thinking about Causes*, ed. Peter Machamer and Gereon Wolters, 235–64. Pittsburgh: University of Pittsburgh Press.

Sober, Elliott. 1984. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Cambridge, MA: MIT Press.

———. 2001. "Two Faces of Fitness." In *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*, vol. 2, ed. Rama S. Singh, Costas B. Krimbas, Diane B. Paul, and John Beatty, 309–21. Cambridge: Cambridge University Press.

———. 2011. "*A Priori* Causal Models of Natural Selection." *Australasian Journal of Philosophy* 89:571–89.

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

Walsh, Denis M., Tim Lewens, and André Ariew. 2002. "The Trials of Life: Natural Selection and Random Drift." *Philosophy of Science* 69:452–73.

https://doi.org/10.1086/670323 Published online by Cambridge University Press