CrossMark

Drift beyond Wright–Fisher

Hayley Clatterbuck

Abstract Several recent arguments by philosophers of biology have challenged the traditional view that evolutionary factors, such as drift and selection, are genuine causes of evolutionary outcomes. In the case of drift, advocates of the statistical theory argue that drift is merely the sampling error inherent in the other stochastic processes of evolution and thus denotes a mathematical, rather than causal, feature of populations. This debate has largely centered around one particular model of drift, the Wright–Fisher model, and this has contributed to the plausibility of the statisticalists' arguments. However, an examination of alternative, predictively inequivalent models shows that drift is a genuine cause that can be manipulated to change population outcomes. This case study illustrates the influence of methodological assumptions on ontological judgments, particularly the pernicious effect of focusing on a particular model at the expense of others and confusing its assumptions and idealizations for true claims about the phenomena being modeled.

Keywords Genetic drift · Wright–Fisher · Semantic view of theories · Population genetics · Evolution

1 Introduction

Philosophers of science have paid much attention to the influence that judgments about ontology have had on methodology, and rightly so, since the nature of the causes and entities under scientific examination typically constrain the type of models and methods that should be used in making predictions. On the other hand, the influence that

H. Clatterbuck (✉)
University of Wisconsin-Madison, 5185 Helen C White Hall, 600 N Park St.,
Madison, WI 53706, USA
e-mail: clatterbuck@wisc.edu

scientific methodology has had on judgments about ontology has not been as widely appreciated. I will argue that methodological commitments can distort ontological judgments. As a case study, I will focus on methodological commitments in population genetics that have played an important, and sometimes pernicious, role in influencing ontological judgments in evolutionary biology.

In this paper, I will argue that philosophers of biology and population geneticists have paid undue attention to one particular methodologically useful but often unrealistic model of genetic drift, the Wright–Fisher model, and that this has led to mistaken judgments about the ontological nature of drift. Due to the centrality of that model in both the historical development and current practice of population genetics, some philosophers have assumed that we can understand all we need to understand about genetic drift from examinations of that model. However, I will argue that broadening our analysis beyond Wright–Fisher uncovers novel features of drift and sheds new light on its ontological status.

This case study points to a broader error lurking in efforts to read ontology directly off of mathematical models. By its nature, a model of a process does not accurately depict all of its real-world features; it contains abstractions and idealizations which permit its users to reason about a particular phenomenon of interest. Any time we focus on a particular model to the exclusion of others, we risk mistaking an abstraction or idealization of that model for a real feature of the phenomena it is modeling.

In what follows, I will show how this mistake has led to erroneous ontological judgments in the case of drift. I will start in Sect. 2 by giving a characterization of genetic drift that is neutral among its various ontological interpretations. In Sect. 3, I will briefly explain two opposing theories about the causal nature of drift and the main argumentative dialectic at the heart of the controversy between them. Then, in Sect. 4, I will argue that both sides of the debate have relied on a single model of drift, a model which has been shown to contain idealizations and ontological assumptions that are inappropriate for many populations. In Sect. 5, I will argue that when we consider alternative models of drift, one ontological view of drift—the view that drift is a genuine cause of evolutionary outcomes—ought to be favored. In Sect. 6, I will suggest some general lessons that can be applied to other cases of reasoning about mathematical models.

2 What is drift?

Before we can discuss alternative ontological theories of genetic drift, it will be necessary to get a handle on the type of phenomenon under examination. Drift is often defined in terms of what it is not; drift denotes those changes in trait frequencies (or the process that produces those changes) that can not be accounted for by differences in fitness, mutation, migration, or the other causes that move populations in predictable directions. This definition, commonly accepted (at least implicitly) by philosophers and biologists, gives rise to two key features of drift. The first is that drift is "random" in that it does not predict a directional change in trait frequencies. The second, which is supposed to follow from the first, is that drift is more pronounced in small populations.

An overused but still useful example will suffice to illustrate a drift process. Consider a population of mice living in a field that is occasionally and randomly hit by lightning strikes. The population contains two types of mice, brown-coated and white-coated, that are otherwise physically identical. Coat color is irrelevant to susceptibility to lightning strikes. Types breed true and the population size is constant. At time k, suppose that there are 6 brown-coated and 4 white-coated mice, so the trait frequencies are 0.6 brown/ 0.4 white. Lightning then strikes the field at random, killing one white-coated mouse. At time k+1, the population contains 6 brown-coated mice and 3 white coated, so the trait frequencies have changed to 0.67 brown/0.33 white. The mice then reproduce until the population size returns to 10, at which point the expectation is that there will be 6.7 brown and 3.3 white mice.[1]

This change in trait frequencies was random because each mouse was equally likely to be struck and killed by lightning, regardless of its coat color. By chance, the lightning happened to have a disproportionate effect on white mice, but it could have disproportionally killed brown mice or white and brown mice in equal proportions. Since lightning strikes are rare and indiscriminate with respect to coat color, the expected trait frequency at time k+1 is identical to the trait frequency at k.

The probability that frequencies will deviate significantly from the expected value depends on the size of the population undergoing drift. To see this, consider a much larger population containing 60 brown-coated and 40 white-coated mice. If a lightning strike had the same effect on this population, killing one white mouse, there would be 60 brown and 39 white mice, so the resulting trait frequency would be 0.61/0.39, a much smaller deviation from expectation than in the first population.

These two informal features of drift—that it is random and more pronounced in small populations—have been formalized in population genetics models of drift. These are time-discrete or continuous Markov models constituted by a transition matrix which describes the probability that a population in allelic state i in generation k will transition to state j in the k+1 generation; $P_{ij}(k+1 = j \mid k = i)$. When these probabilities are calculated for all possible j states, the result is a probability distribution over the possible allelic states of the population in the next generation.[2] In the models standardly used by population geneticists, the distributions have the following formal characteristics (Der et al. 2011, p. 82):

(Mean) The mean of the probability distribution does not change from k to k+1[3];

$$E(X_{k+1}|X_k) = X_k$$

---

[1] This example is an instance of a standard Wright–Fisher model interpretation of drift. I do not mean to imply that this is the only possible type of drift process, and I will consider others in Sect. 4.

[2] In a pure drift model, each individual in the population has an identical expected offspring distribution. In mixed models, drift can be combined with selection, mutation, and other factors.

[3] X denotes the number of individuals of the X trait. X/N gives the frequency of the trait.

(Variance) The variance of the distribution in k+1 is a function of trait frequencies in k and the population size, N[4];

$$\mathrm{Var}\left(X_{k+1}|X_k\right) = X_k \left(1 - \frac{X_k}{N}\right)\left(\frac{N\sigma_N^2}{N-1}\right)$$

Though the informal and formal properties of drift characterize the dynamics of a population undergoing drift, they do not tell us much about ontological status of drift. How we ought to interpret these models is the topic of the next section.

3 Two competing ontological theories of drift

The distribution of trait frequencies in the population of mice changed from one generation to the next, and the resultant distribution deviated from the expected value. Should we say that drift *caused* the resulting trait frequencies to deviate from expectation? That drift *is* the deviation from expectation? Or that drift is the sampling error inherent in the stochastic process of selection by lightning strike?

What may seem like a small quibble in the case of our field mice has wider ramifications. Over the last few decades, a debate has emerged concerning the ontological status of the putative causes of evolution, particularly drift and natural selection. At stake is whether the theory of evolution is a theory of causes. If it turned out that one of our best scientific theories was not a theory of causes at all, this would be an important and surprising result.

On one side of the debate are the causal theorists, defenders of the traditional view that drift and natural selection (along with mutation, migration, etc.) are genuine causes of changes in trait frequencies in real populations. According to the causal theory, drift corresponds to some causal feature (or supervenes on some set of causal features) in the underlying causal network governing the dynamics of an evolving population (Filler 2009), though there is disagreement about what these features are [5] and whether drift is a separate causal process from selection, mutation, and so on.[6] On this view, the causal features of evolving populations bear a strictly horse-cart relationship to population genetics models; the formal properties of models are grounded in the causal properties of populations and are appropriate insofar as they describe how the causes of evolution interact to determine the dynamics of traits in a population.

---

[4] The terms in the right-hand side of the equation below denote, from left to right, the number of individuals with trait X, the number of individuals of the alternative trait, and a term accounting for variance in the offspring distribution of an individual with the X trait. The right-most term is used to calculate the variance expected population size, which I will discuss in more detail Section 4. See Der et al. (2011) for a more thorough explanation.

[5] For example, Millstein (2002) and Hodge (1987) argue that drift is indiscriminate sampling (that is, sampling without respect to intrinsic physical differences among individuals). According to Gildenhuys (2009), drift refers to "causal influences over a population" that are "non-interactive, non-pervasive, and indiscriminate causes" (p. 522). Other prominent defenses of the causal theory, such as Shapiro and Sober (2007), Stephens (2004), and Reisman and Forber (2005) do not take an explicit stand on this issue.

[6] Okasha (2006), Abrams (2007), and Clatterbuck et al. (2013) are representative of the single-process view, while Millstein (2002) and Hodge (1987) are representative of the separate-process view.

This view has been challenged by statistical theorists who argue that we ought to look at the role that drift and selection play in the mathematical models of population genetics to determine their ontological status, and that doing so reveals that they are not genuine causes but rather statistical summations over the genuine causes—the births and deaths of individual organisms—of evolution (Matthen and Ariew 2002, 2009; Walsh 2000; Walsh et al. 2002). On the statistical view, the theory of evolution is analogous to statistical mechanics, which describes the probability of different molecular concentrations without tracking the motions of individual molecules.

Causal theorists have criticized both of the statistical theorists' primary commitments, arguing that we can not read ontology off of mathematical models (Millstein et al. 2009), and that even if we could, it wouldn't follow that drift and selection are not causes. An influential argument in support of this latter claim relies on the connection between drift and population size. If we intervene to reduce the size of a population, we increase the probability that outcomes other than that predicted by selection, mutation, migration, etc. will occur; therefore, drift is a cause of those outcomes (Shapiro and Sober 2007; Stephens 2004). This argument is explicitly given within (Woodward 2003) causal interventionist framework by Reisman and Forber (2005):

(1) If an appropriately controlled manipulation of variable A results in a systematic change in variable B, then A is a cause of B.
(2) An appropriately controlled manipulation of drift (i.e. manipulating N) results in systematic changes in population-level dynamics (i.e. the probability distribution of trait frequency outcomes).
C: Therefore, drift is a cause of population-level dynamics.

In support of this argument, Reisman and Forber offer evidence that interventions that reduce population size lead to increased variance in population outcomes, both in natural "experiments" and in controlled laboratory experiments. When natural populations undergo bottlenecks, neutral traits at intermediate frequencies often go to fixation. In a controlled experiment of this phenomenon, Dobzhansky and Pavlovsky allowed replicate populations of fruit flies, consisting of two types in equal proportions, to evolve under stabilizing selection. Holding selection constant across trials, the experimenters manipulated the size of the populations by culling some populations to $N \approx 4{,}000$ and others to $N = 10$. The large populations reached roughly similar equilibrium frequencies while the small populations showed greater heterogeneity in equilibrium frequencies. Reisman and Forber argue that this manipulation of N constituted an appropriate intervention; because the strength of drift is inversely proportional to population size, N determines the magnitude of drift (Reisman and Forber 2005, p. 1115).

Statistical theorists have leveled several arguments against the second premise of the above argument. First, they have argued that it is impossible to perform an appropriate manipulation of drift since doing so would require a manipulation of individual-level causes, i.e. the lives, deaths, and reproduction of individual organisms (Walsh 2000; Walsh et al. 2002). In response, some causal theorists have argued that population-level causes, such as selection or drift, are not causal competitors with those individual level causes, since population-level causes supervene on individual level causes (Shapiro and Sober 2007). This objection raised by the statisticalists is not a problem for evo-

lutionary causes in particular but higher-level supervening causes more generally; insofar as one accepts higher-order causes, this objection should not be troubling.

However, according to a second argument from statistical theorists, even if we adopt the view that drift supervenes on features in the underlying network of a population, this does not yet show that drift is a cause. There is a natural response to the Dobzansky and Pavlovsky experiment which motivates this statisticalist response. Why was it necessary to actually perform this experiment? The experiment merely revealed a statistical truism; the results of any stochastic process will show more variance as the number of trials decreases. When you flip a fair coin 10 times, you are more likely to observe results other than 0.5 heads/0.5 tails than if you flipped it 4,000 times.

The objection, voiced by Matthen and Ariew (2009) is that manipulations of N reveal a purely mathematical relationship between N and population outcomes, not a genuinely causal one:

> What [causal theorists] miss is that the connection between population size/ variation-in-advantageous traits and drift/ selection is purely mathematical. This connection is the same as that which holds between sampling size and proportional variance from the mean in random sampling… Sample size and variance under random sampling are connected by mathematical law, and thus are not sufficiently distinct from one another to account as terms in a cause-effect relationship (p. 212).

This objection consists of the two related claims that (a) drift does not supervene on any causal features of an evolving population but instead (b) supervenes on a mathematical feature. In support of the first claim, statistical theorists point out that fitness, mutation, migration, and the other putative causes of evolution are inherently stochastic. For illustration, in our population of field mice, each mouse is physically identical (except for coat color) and therefore has equal fitness, but this does not mean that each mouse will actually have the same number of offspring. An organism's fitness is a probabilistic expectation, derived by calculating of the number of offspring it will have if various conditions—including lightning strikes—obtain, weighted against the probability of those conditions obtaining (Mills and Beatty 1979).

If fitness encompasses all of the causes that may influence survival and reproduction and if the causes that influence fitness are the subvenience base of selection (and so on for the causes that influence mutation, migration, and so on), then it looks like there are no causes left in the underlying causal network of an evolving population for drift to pick out. Matthen and Ariew conclude that drift is non-causal, for "once a reference class has been partitioned in terms of all the factors that make a difference, the residual variation within the cells of the partition—the unassigned variation—is uncaused. It is due to chance if you like" (Matthen and Ariew 2002, p. 64).

What then does "drift" refer to? For any stochastic process with a limited number of "trials" (here, organisms in the population), there will be statistical error. For statistical theorists, drift just is this statistical error (Walsh et al. 2002). While sometimes "statistical error" refers to causes that have not been accounted for, by hypothesis, all of the causes impinging on the mouse population have been accounted for in the determination of individual and trait fitnesses. This motivates the objection's second claim.

Drift supervenes on a *mathematical* feature of an evolving population. The number of trials of stochastic evolutionary processes uniquely determines the magnitude of drift, so drift denotes a mathematical relationship between population size and variance in evolutionary outcomes.

Motivated by similar considerations, Lange (2013a, b), argues that to explain some evolutionary outcome in terms of drift is to give a "distinctively mathematical" or "really statistical", rather than primarily causal, explanation. The explanans of a drift explanation will cite mathematical features of the individual-level subvenience base and thus show why the explanandum follows by mathematical necessity.[7] For example, the explanation for why the small populations of flies had a larger variance in outcomes than the large populations need only cite the fact that the small populations experienced fewer trials of a stochastic process and the statistical fact that the variance in outcomes of a stochastic process is inversely proportional to the number of trials.

Similarly to Matthen and Ariew's argument that drift doesn't correspond to anything causal, Lange argues that while a causal explanation "derives its explanatory power from describing relevant features of the result's causal history, or more broadly, of the world's network of causal relationships", drift explanations do not bear this hallmark (Lange 2013a, p. 183). To explain why the small populations of flies deviated from expectation, it is not necessary—and arguably, it does not "deepen" the explanation at all—to provide details of the survival and reproduction of fruit flies, their offspring distributions, or the structure of their populations. Another way of getting at the same point is to say that mathematical explanations are substrate neutral. The explanation we gave for why the small populations of flies had greater variance in outcomes would work equally well at explaining the outcomes of coin-flips or any other stochastic process. It is no coincidence that almost all philosophical discussions of drift are done in terms of coin-flipping or drawing balls from an urn, Lange or the statistical theorist might argue, since there is nothing distinctively biological about drift explanations. The mathematical properties of stochastic processes are all that are needed.

While some causal theorists are not troubled by the fact that population size and evolutionary outcomes bear this strong mathematical relationship[8], I think that it does raise a strong *prima facie* challenge to what I take to be the strongest argument in the causal theory's arsenal. In order to combat this objection, causal theorists somewhat paradoxically must weaken the connection between population size, N, and population outcomes in order to redeem it as genuinely causal. They must show that this mathematical feature of populations does not uniquely determine the magnitude of drift, and further, that drift does supervene on causal features in the subvenience base of an evolving population. Fortunately for the causal theory, new theoretical work has shown that the relationship between N and drift is not as tight as once assumed.

Both sides of the debate have made a few (usually implicit) assumptions. The first is that the magnitude of drift is to be measured by the deviation of trait frequencies from

---

[7] This is stronger than the nomological necessity constitutive of causal explanations.

[8] For example, Sober (1984, p. 117) argues that the sampling size of a stochastic process is a cause of that process's outcomes. The probability distribution over outcomes of a series of coin-flips is affected by both the bias of the coin and the number of times you flipped it. See also Sober (2011) for a defense of the claim that there are *a priori* causal claims in evolutionary biology.

the frequencies predicted by selection (and mutation, migration, etc.).[9] The second is that there is a unitary function relating the population size, N, and said magnitude. To wit, consider these representative characterizations of drift, from defenders of the causal and statistical view, respectively:

> In a population of a given size, drift as a process of indiscriminate sampling *always* has the same force. It is part of the definition of drift that it is stronger when the population is smaller (Stephens 2004, p. 557), italics in original).
> Drift is manifested as a difference from the outcome predicted by the fitnesses in the population. The law of large numbers tells us that the likelihood of significant divergence from these predictions is an inverse function of the size of the population (Walsh et al. 2002, p. 459).

If these assumptions hold, then the only way to manipulate drift is to manipulate N, a mathematical feature, and the only effect of manipulating N is a change the probability of deviations from expected trait frequencies. Participants on both sides have restricted their focus to a particular model of genetic drift, the Wright–Fisher model, under which both of these assumptions hold.[10] However, I will argue that broadening our conception of drift to include alternative models unveils novel ways of intervening on drift that will strengthen the causal argument against its statisticalist detractors.

4 Alternative models of drift

The model of drift that has been assumed to be *the* model of drift (or one of a number of predictively equivalent models of drift) is the Wright–Fisher model, which was first articulated by Wright (1931) and given in its diffusion limit by Kimura (1962). Philosophers of biology perhaps cannot be blamed for this single-minded focus. The Wright–Fisher model was the first prominent account of drift within population genetics, and due to its simplicity, it continues to be a useful tool for predicting frequency changes due to drift. Indeed, the model is still the most influential account of drift among working biologists. According to Der et al. (2011), "Population geneticists typically characterize genetic drift by a single number—the variance effective population size—which simply scales all genetic quantities. As a result, there is widespread belief among many biologists that there is only a single reasonable model of drift" (p. 81).

The Wright–Fisher model describes drift processes in natural populations as binomial sampling with replacement and a constant population size N. In Wright's original formulation, drift is conceived of as a process in which each individual in the k generation produces a very large number of gametes which are true to type, and then a new population at the k+1 generation is created through N statistically independent sampling events from that nearly infinite pool of gametes. The likelihood of an X-type allele being drawn from this gamete pool is equal to the frequency of the X-type in the parent population. The transition matrix of this process is given by:

---

[9] According to some views, drift just is the actual deviation of trait frequencies from expectation (Brandon 2005). On most causal views, N is an indication of drift's power to cause such deviations.

[10] Notable exceptions are Gildenhuys (2009), Millstein et al. (2009), and Plutynski (2007).

$$\text{Pr}_{ij} = \text{Pr}\left(X_{k+1} = j | X_k = i\right) = \binom{N}{j}\left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$$

This ontological picture lends itself to the traditional depiction of drift as akin to sampling balls from an urn. Consider an urn at the k generation containing 6 green and 4 red balls ($N = 10$). A new urn is populated at the k+1 generation by randomly drawing a ball, making a copy to place in the new urn, and then replacing the drawn ball. This process is repeated 10 times to create a new population of $N = 10$. The Wright–Fisher equation gives the probabilities that the new urn will contain particular numbers of green and red balls.

The most likely outcome is that the frequency of green balls in the k+1 urn will be 0.6, the same as in the k urn, but other frequencies are possible. For example, the probability that the new urn will contain at least 7 green balls is quite large (=0.382). Now consider an urn at the k generation containing 60 green and 40 red balls ($N = 100$), from which a new urn is created at the k+1 generation by the same process of binomial sampling with replacement. The probability that the new urn will contain 70 or more green balls is much smaller (=0.023).

This analogy illustrates a few key features of the Wright–Fisher model. First, initial trait frequencies and the population size entirely determine the transition matrix. Deviations from expectation are more likely as the two types approach equal frequencies in the k generation and as the number of trials, N, decreases. The probability that the urn containing 10 balls would "jump" from 0.6 green to at least 0.7 green was much larger than the probability that the urn containing 100 balls would. Second, in general, big jumps in trait frequencies are highly improbable for sizeable populations. The probability that the small urn will jump from 0.6 green in the k generation to all green in the k+1 generation is 0.006. For the large urn, the probability is even more miniscule (=0.6^{100}).

A third feature is not illustrated in the above example but can easily be seen. The Wright–Fisher model predicts that a novel mutant in the k generation is very likely to be lost, even when it is selectively advantageous. To model selection in our urn cases, suppose that some balls are more likely to be drawn than others by a factor determined by the selection co-efficient, s. Now imagine an urn of 99 green balls and a single new yellow mutant ($N = 100$). Even the yellow type is twice as fit as the green type ($s = 1/2$, a very large value of s for natural populations)[11], the probability of losing the yellow type in the k+1 generation is 0.13. In general, when s is small and Ns is greater than 1 (which is the case in most populations), the probability that a newly introduced favorable allele will go to fixation is approximately $2s$ (Wright 1931, p. 133).[12]

---

[11] Here, we can model this by supposing that the ball drawing is no longer indiscriminate, and a token yellow ball is twice as likely to be drawn than a token green ball. See Brandon (2005, pp. 157–158) for a similar example.

[12] From Wright, the probability of fixation (π) of a beneficial allele introduced at frequency $1/2N = 2s/(1-e^{-4Ns})$. Notice that as Ns increases, the value of the denominator goes to 1, so $2s$ will be a good approximation of the probability of fixation for most values of s. For a more detailed derivation and discussion, see Kimura (1962, pp. 715–716). I am grateful to a reviewer for helpful comments on this issue.

Before we compare the Wright–Fisher model to others, a quick aside will be helpful. As we have seen, philosophers of biology often associate the strength of drift with population size, N, which is a simple count of the organisms in a population.[13] However, biologists more typically measure the strength of drift with the effective population size, N_e (see Charlesworth 2009 for a thorough explanation of the concept). The Wright–Fisher model contains various assumptions that are not true of most actual populations; for instance, it assumes that each individual in the population contributes to the gamete pool, that there is no associative mating within types, and so on. The effective population size N_e of a real population P describes the size N of a theoretical population which obeys the assumptions of the Wright–Fisher model whose dynamics accurately model those of P. For example, a population that contains 10,000 individuals, in which only some of which reproduce at a given time and in which organisms preferentially mate with their own type, might behave *as if* it was a population of 2,000 individuals under the Wright–Fisher model.

While the causal theory benefits from an appreciation of the theoretical nature of the effective population size, I will not spend much time arguing the point here since it is illustrated much more dramatically by the alternative models I will consider next. However, two points here will foreshadow my argument in the next section. First, the effective population size for a population P is determined by facts about the underlying network of biological causes of P, in particular, facts about the mating and reproductive tendencies of organisms in P. Second, the effective population size is the true measure of the strength of Wright–Fisherian drift; it predicts greater deviation of trait frequencies from expectation when the actual population is smaller but also when factors like assortative mating make the effective population size smaller. These facts show that the magnitude of Wright–Fisherian drift supervenes on causal features of an evolving population, not merely the mathematical feature N, and that these are features that are not subsumed in selection, mutation, or the other higher-level causes of evolution. Lastly, they suggest that the mathematical relationship which figures in the Wright–Fisher model is often not a literally true (or even nearly true) description of real populations it describes, so it is an error to draw ontological judgments about the population directly off of the model.

As I noted in Sect. 2, there are two informal features that are taken to be constitutive of drift—that it is random and more pronounced in small populations—and the binomial sampling process described by the Wright–Fisher model obeys these features.[14] However, since Wright, more general drift models have been constructed that abandon some of the restrictive assumptions of the Wright–Fisher model. Consider the

---

[13] I am not claiming that philosophers have wholly ignored the distinction between population size and effective population size; for instance, Stephens (2004) uses the effective population size as an indicator of the strength of drift. However, few in this debate have paid close attention to the significant causal factors which determine effective population size. Gildenhuys (2009) is a welcome exception.

[14] As Der et al. (2011) show, the Wright–Fisher model, as well as the other models I consider, also obey the formal characterizations of drift described by (Mean) and (Variance). The question of whether there are drift models that exhibit the informal properties of drift while not obeying these formal properties is interesting but beyond the scope of this paper. For my purposes, it suffices to show that there are models which meet the formal requirements which population geneticists use to define drift processes but differ interestingly in their resulting population dynamics.

Cannings model which specifies a transition matrix via θ, an independent identically distributed random variable specifying the offspring probability distribution of types (which are equal under a pure drift model):

$$\text{Pr}_{ij} = \text{Pr}\left(X_{k+1} = j | X_k = i\right) = \text{Pr}\left(\sum_{i=1}^{i} \theta_i = j\right)$$

The binomial variable is one candidate for θ, but there are others (Cannings 1974).

Recent theoretical work by Der et al. has shown that there are versions of these alternative models that obey the two characteristic features of drift (they call these "generalized Wright–Fisher models" or GWFs) yet differ dramatically in the outcomes that they predict. The reason for this is that (Mean) and (Variance) describe the first two statistical moments of a GWF process, but processes satisfying these are free to differ in their higher moments. The variance effective population size that is usually discussed in applications of the standard Wright–Fisher model is determined by the variance in offspring distributions of types in a population. However, offspring distributions with the same variance can differ in skew, heavy-tailedness, and so on.

For illustration, I will focus on the Eldon–Wakeley model, which has been proposed as a model of drift in populations of Pacific oysters (Eldon and Wakeley 2006, Der et al. 2011, 2012). In this model, "individuals produce one offspring each generation, until a random time at which a single individual replaces a fraction λ of randomly chosen individuals from the entire population" (Der et al. 2011, p. 83). This counts as a pure drift process, satisfying (Mean) and (Variance). Each individual (and therefore each type) has equal fitness, an identical, but highly skewed, offspring distribution (each individual has a high probability of leaving one offspring, a small chance of being randomly chosen to leave many offspring, and a small chance of being replaced if another individual is chosen to replace its offspring in the next generation), and therefore, the expected distribution in a given generational transition is identical to that in the parent generation. The variance will depend on the value of λ (which in turn is determined by N).

Now, consider an extreme version of this process in which an individual is randomly chosen at k to replace the entire population with its offspring at k+1 (so $\lambda = 1$), where the average time until this random bottleneck effect is N. In the ball-drawing analogy, this would be like having an urn filled with 100 balls, 60 green and 40 red, from which each ball typically contributes exactly one representative to a new urn, containing 60 green and 40 red balls. However, once every 100 generations or so, you take the first ball you draw, copy it 100 times, and use those copies to fill the new urn. This new urn has a 0.6 chance of having all green balls and a 0.4 chance of having all red balls.

This type of drift process will behave much differently than the binomial sampling process that we considered above. First, it will permit much bigger jumps. For $N = 100$, the probability of going from 60 green and 40 red balls to 100 green balls in one generation is virtually 0 (0.6^{100}) on the Wright–Fisher model, but it is 0.006 on

the extreme Eldon–Wakeley model.[15] For $N = 10$, the probability of such a jump is 0.006 on Wright–Fisher but is ten times more probable (0.06) on the Eldon–Wakeley model.

Second, long periods of stasis between bottleneck events allow new mutants to persist and therefore for selectively favored mutants to increase in frequency between drift events. Since the drift event randomly selects an individual to populate the next generation, the probability of a type being chosen is equal to its frequency at the time of the drift event. Therefore, long periods of stasis increase the probability that a fitter allele will eventually go to fixation. The probability of fixation of a newly introduced favored allele will depend on the relative values of s and the expected time until bottleneck, N, and will go to 1 in the limit of very rare bottlenecks ($N \rightarrow \infty$). This is radically different than the Wright–Fisher model, which predicts that the probability of fixation of a new favored mutant is $2s$, no matter how large N gets. (Der et al. 2012, p. 1332) conclude that "selection operates very differently in the Eldon–Wakeley model than it does in the standard Wright–Fisher [model]… in particular, the form of genetic drift that arises in populations with reproductive skew tends to amplify the effects of selection, relative to the standard form of Wright–Fisherian drift—even when both models are normalized to have the same variance-effective population size".

The standard Wright–Fisher and Eldon–Wakeley models represent two extremes on a continuum of GWF models which predict strikingly different population dynamics and outcomes. Consideration of these models shows that the assumptions on which the debate between the causal and statistical theories have often been predicated—that the only relevant outcomes of drift are deviations of trait frequencies from expectation[16] and that population size is the unitary measure of the probability with which drift will produce those outcomes—are false. To see this, compare Stephen's statement (quoted above) that drift "*always* has the same force" with a characterization of drift from Der et al.:

> The Wright–Fisher model acts as a vigorous suppressor of selection. Once again, this supports the idea that the form of genetic drift encoded by the Wright–Fisher model (and those models with the same diffusion limit) is extremely strong. In the Wright–Fisher model, drift counteracts the deterministic force of selection more powerfully than in any other generalized population model (Der et al. 2011, p. 87).

The strength of drift can be measured with respect to various outcomes, including suppression of selection, time to fixation, and the probability of fixation of an advantageous mutant. Further, the size of a population does not determine the value of these outcomes; as we have seen, a population of $N = 1,000$ behaves differently if it is undergoing an extreme Eldon–Wakeley process than if it is undergoing Wright–Fisher binomial drift. In the next section, I will argue that these differences in the magnitude

---

[15] The probability that green will go to fixation is the probability the population will go through a bottleneck event in that generation ($1/N = 0.01$) times the probability that it will be a green ball that is chosen to populate the next generation (its starting frequency, 0.6).

[16] This is not a universal assumption of causal theorists. For instance, see Stephens (2004) and Filler (2009) for discussions of drift and changes in heterozygosity.

of drift in differently constituted populations suggest new ways of manipulating drift, ways that buttress the causal theory against its statisticalist detractors.

5 Redeeming the causal theory

The contrast between the Eldon–Wakeley and Wright–Fisher models sketched above highlights the fact that different drift models contain different ontological assumptions about the drift process at work in a population, and whether these assumptions correspond to the actual drift process that a population is undergoing determines which model will deliver accurate predictions about population outcomes. Here, I agree with the criticism of statistical theorists from Millstein et al. (2009). They argue:

> [Models] are always ideal structures. Interpreting a model, then, involves proposing that certain features of the model, some or all of variables and functions, correspond to certain features of the world… The justification of a model's interpretation depends crucially, we claim, on claims about the physical processes that affect the allele frequencies and their dynamics (Millstein et al. 2009, p. 5).

On this semantic interpretation of theories, it is not possible to read the ontology of an evolutionary process directly off of a mathematical model.

In support of the semantic view, Millstein et al. point out that there are different drift models (they mention the Cannings and Moran models[17] that carry different assumptions about the causal network of a population undergoing drift.[18] They state, "It is not clear whether there is any particular reason to choose one drift model over another in trying to *understand* the concept of drift, though there may be other reasons to choose a particular model (e.g. tractability)" (Millstein et al., p. 6). Though the presence of alternative models of drift is an important point in favor of their semantic view, which I share, their argument is strengthened considerably by emphasizing the predictive inequivalence of those models in addition to differences in tractability or other virtues.

If various models were predictively equivalent, then it would lend support to the statisticalist argument that the underlying biological details do not make a difference for evolutionary outcomes that are attributable to drift. If drift was invariant across changes in the underlying causal network—in other words, if it was substrate-neutral—then this would be evidence that what drift models are latching onto is a merely mathematical regularity, and that to explain an outcome in terms of drift is to give a really statistical explanation.

---

[17] The Eldon–Wakeley model is an extension of the Moran framework. For a more detailed discussion, see Der et al. (2012, p. 1332).

[18] One of their primary arguments is that the statisticalist view cannot adequately capture actual scientific practice since the development of alternative models of drift was motivated by the desire for models that made more realistic biological assumptions than the Wright–Fisher model, which Millstein et al. claim "makes no sense if drift were only a statistical outcome" (p. 6). I agree with their contention that biologists have often conceived of drift as a causal process, yet I do not think this argument is a particularly compelling refutation for statistical theorists who want to substantially revise the way that biologists (as well as philosophers) think about drift.

Millstein et al. are correct that the fact that a relationship is represented mathematically in a model does not entail that the relationship being modeled is itself merely mathematical.[19] However, this lack of an entailment does not show that the statistical-ist theory is false; we need some reason to think that the relationship being modeled is genuinely causal. The predictive inequivalency of different drift models, such as the Wright–Fisher and Eldon–Wakeley models, plays this very role by showing that, contrary to the statisticalists' assumptions, details of the causal network underlying a population do change the outcomes of drift. Drift does supervene on causal properties of an evolving population, and by manipulating these properties (thereby manipulating drift), we can cause changes in evolutionary outcomes.

The different ontological assumptions of alternative drift models suggest ways that we can manipulate drift in real populations. For instance, we can alter a population that is undergoing a Wright–Fisher type of drift process so that it switches to a drift process more like that described in other models, such as the Eldon–Wakeley model; in effect, we can randomly sample the population differently. What then are the different onto-logical commitments of the standard Wright–Fisher and alternative models that gen-erate such predictive differences and possible manipulations? A full treatment of the issue is beyond the scope of this paper, but a few examples will suffice for my ends here.

The Wright–Fisher is a particularly good model of some paradigm cases of drift, including gametic sampling. Suppose an individual is a heterozygote, *Aa*, with respect to some trait and reproduces 10 times. Each reproductive event can be seen as sampling a single gamete from this pool to contribute a single offspring to the next generation. The probability of sampling an *A* gamete from our organism is equal to the starting frequency of *A*'s in the gamete pool, which is 0.5 (and so on for *a* gametes).[20] The expectation is that 5 of the organism's offspring will receive an *A* gamete and 5 will receive an *a* gamete, but given a limited number of trials, the actual outcomes may deviate from expectation.[21]

We can expand this analysis to consider how a population of diploid organisms will evolve, where $N = 10$ (so the total number of alleles is 20) and the initial allele frequencies are 0.5 *A* and 0.5 *a*. According to the Wright–Fisher model, on which:

(a) each gamete has an equal chance of contributing to the next generation (there is no selection),
(b) gametic sampling events are independent and identically distributed, and
(c) in an individual sampling event, each gamete gives rise to a single offspring[22],

the probability that one or the other allele will increase in frequency from 0.5 to at least 0.8 is

---

[19] I will return to this point in Sect. 6.

[20] I will ignore meiotic drive or other segregation distorters.

[21] Filler (2009) argues that drift's tendency to reduce heterozygosity shows that drift is an evolutionary force.

[22] This does not mean that every gamete will contribute a single descendent to the daughter population. Sampling with replacement allows for the possibility that a gamete could be sampled more than once. However, this condition does ensure that the expected offspring distribution of a gamete is binomial.

$$\text{Pr}_{ij} = 2 \sum_{j=16}^{j=20} \binom{20}{j} \left(\frac{10}{20}\right)^{j} \left(1 - \frac{10}{20}\right)^{20-j} = 0.006$$

However, there are possible manipulations to this population that will yield changes in its evolutionary dynamics, even while holding the population size ($N = 10$, number of alleles $= 20$) constant and maintaining neutrality of the A and a alleles. Suppose we keep assumptions (a) and (b) but change (c), so that now, in each generation, there is a $1/2N$ chance that a randomly selected allele will populate the entirety of the next generation (similar to a replacement event in an extreme Eldon–Wakeley process).

Using this example, we can formulate a new version of the manipulationist argument for the causal theory. The Wright–Fisher and Eldon–Wakeley models each contain a variable denoting a probability distribution over the number of offspring that may be produced by each individual in a generation. As assumption (c) shows, the Wright–Fisher model sets constraints on this variable. However, in some real populations, organisms may produce offspring numbers that fall outside of the range specified by the Wright–Fisher model. According to Der et al. (2012, pp. 1331–1332):

> The Wright–Fisher model assumes that individuals each produce a Poisson-distributed number of offspring each generation, subject to the constraint of a constant population size (Karlin and McGregor 1964). This formulation excludes the possibility of a highly skewed distribution of offspring numbers, which has been observed empirically in some species. Marine species in particular, as well as some plants and fungi, sometimes produce a very large number of offspring when faced with high mortality early in life (Hedgecock 1994).

Suppose that we have a population of $N = 10$ diploid Pacific oysters that initially have offspring distributions that fall within the specified range of the Wright–Fisher model. Then, we intervene on the population in such a way as to increase the probability that individuals have a far greater number of offspring (relative to the population size) than allowed by the Wright–Fisher model[23], while keeping s constant and offspring distributions equal for every individual. Sampling from this offspring distribution increases the probability that a single individual's offspring will constitute a large fraction of the population in the next generation. If we increase the skew as described, the probability of a neutral allele increasing in frequency from 0.5 to at least 0.8 will be greater than 0.006 (the probability of that outcome obtaining under binomial sampling).[24]

---

[23] The quote from Der et al. (2012) suggests ways of manipulating the probability distribution over offspring number so as to increase skew; when mortality rates are high, individuals may increase the number of gametes they produce in a given breeding episode, so we could intervene on offspring distributions by changing the environment in a way that increases mortality rates.

[24] The proof of this is simple. In $(1 - 1/2N)$ of the generations, the probability of a jump from 0.5 A to > 0.8 A or 0.5 a to >0.8 a is given by the binomial equation, but in $1/2N$ of the generations, the probability of such a jump is 1, which by necessity makes the probability of a jump greater than it would be in a binomial process. The exact probability of such a jump will depend on the value of λ, N, and the frequency of A and a when a replacement event occurs. The probability of fixation of a selectively favored allele will also increase.

In summary, the Wright–Fisher model states that the offspring distribution variable can only take on certain states specified by the binomial distribution. We can shift the population from a Wright–Fisher process to an Eldon–Wakeley drift process by manipulating the population such that the offspring distribution variable takes states that are incompatible with the Wright–Fisher model. Manipulations of this variable change the probability distribution over allele frequencies in the next generation. Since each individual still has an identical offspring distribution, this is a pure drift process. Following Reisman and Forber (2005), we can construct the following manipulationist argument:

(1) If an appropriately controlled manipulation of variable A results in a systematic change in variable B, then A is a cause of B.
(2) An appropriately controlled manipulation of drift (i.e. a manipulation of the number of offspring produced by individuals undergoing sampling) results in systematic changes in population-level dynamics (i.e. the probability of transitioning from 0.5 frequency of the *A* allele to a >0.8 frequency of the *A* allele).
 C Therefore, drift is a cause of population-level dynamics.

The examples I have already discussed suggest additional manipulations of drift that will yield changes in evolutionary outcomes. If we intervene on our population of field mice by physically clustering individuals of a type together, thereby imposing correlations among individuals of a type, we can increase the probability that the population will exhibit large jumps in trait frequency. For example, suppose that we intervene to make brown mice cluster closely with other brown mice and white mice cluster with white mice, and that so clustered, a lightning strike will kill multiple mice at once. Even if a lightning strike is equally likely to strike a mouse of either coat color, this manipulation will change the probability that the population frequencies will change dramatically in the event of a lightning strike.

Lastly, recall that the probability of fixation of a newly introduced advantageous allele is $\approx 2s$ for nearly all populations on the Wright–Fisher model while on the Eldon–Wakeley model, it approaches 1 as the time period between drift events increases. We can weaken the power that drift has to suppress selectively-favored mutations in a population by increasing the time between population bottlenecks and/or changing the offspring distributions of individuals in the population to resemble those described by the Eldon–Wakeley model.

In each of the three examples given above, manipulations to causal properties of evolving populations—properties such as the frequency of lightning strikes and bottlenecks, the probability that individuals will produce large numbers of offspring, and demographic properties like population density—yield changes in evolutionary outcomes. This shows that drift does not merely supervene on mathematical properties. Further, some of these causal properties are not in the subvenience base of any of the other canonical evolutionary causes.

I expect that a statistical theorist would level the same objection against these more nuanced manipulationist arguments that she did against the argument from manipulations of population size. She may argue that alternative models of drift merely show that the mathematical relationship between drift and evolutionary outcomes is more complicated than once thought, but it does not show that the relationship is genuinely

causal. After all, she might argue, the outcomes predicted by the Eldon–Wakeley model follow by mathematical necessity from facts about the offspring distributions of individuals in the population (which are determined partially by λ) and the expected time until bottleneck (N). Perhaps I am correct that philosophers of biology have erred in focusing exclusively on population size, but once we have acknowledge that error, nothing else follows.

I think that this response is mistaken. It is trivially true that once a mathematical model has been constructed to describe and predict population dynamics, the relationships between variables in that model will be mathematically related. However, causal information is crucial in determining which mathematical model will be a true or accurate model of a population, and its variables will represent causal features of the population.

Alternative models of drift also show why Lange is mistaken in calling drift explanations "really statistical explanations". Statistical explanations, he argues, are not "deepened" by being supplemented with facts about the causal features underlying a chance process, and "these facts have no place in the explanation since the explanation does not derive its power to explain from its describing relevant features of the result's causal history" (Lange 2013a, pp. 172–173). On the contrary, the causal features of a population evolving under drift, beyond just the number of trials of the stochastic process, are crucial in determining the outcomes we should expect and also for explaining outcomes that do occur.

For example, suppose you want to explain why an allele starting at an intermediate frequency in a small population went to fixation in the next generation, and that the correct explanation is that it was due to drift. Lange argues that this explanation is statistical, for we need only cite the fact that the population was small—in other words, that there were a small number of trials of a stochastic process—to explain why the allele had a high probability of going to fixation. However, since some drift processes will make this outcome more likely than others, even when we hold population size constant, we can deepen the explanation by adverting to causal features of the population, i.e. "The individual offspring distributions in the population of Pacific oysters were highly skewed because individuals were producing large numbers of gametes in response to environmental stresses that made early mortality likely, so allele fixation in that time period in a population of that size undergoing drift was more probable than it would have been if the individuals had binomial offspring distributions". The number of trials still matters in drift explanations, but so does the type of drift process at work.

While I do not intend to elaborate and defend a particular causal theory of drift here, I will briefly state how such a theory might work in light of the alternative models of drift I have considered. On the causal theory, there are various causes acting on individuals within a population that lead to different rates of survival, death, and reproduction. When traits are regularly and projectibly correlated with differential survival and reproduction, "selection" denotes this correlation and measures its strength and direction. "Drift" ranges over those causal factors that are not regularly and projectibly correlated with differential survival and reproduction among types.[25]

---

[25] This is another way of saying that drift is "indiscriminate". See Millstein (2002) for a more thorough explication of the concept.

As alternative models of drift demonstrate, these latter causal factors may vary in differently constituted populations and interact differently with selection and the other causes of evolution. Recall that according to the statistical theorists, drift is what remains once we have taken account of all of the "factors that make a difference" to evolution (Matthen and Ariew 2002, p. 64). On the contrary, drift corresponds to factors that do make a significant difference in the evolution of actual populations, and therefore, it would be an error to consign drift to the dustbin of non-causal "mere chance".

6 Implications

I have been concerned with a special ontological question within the domain of evolutionary biology, but that dialectic holds lessons for ontological investigations in science more generally. In particular, I want to call attention to two errors that have led to confusion in the debate over the causal nature of drift. The first error is confusing a mathematical description of a causal process for a true description of a mathematical process. The second error is in focusing too narrowly on one methodologically useful model at the expense of others and therefore mistaking the particular idealizations and assumptions of that model for true features of the causal process that it describes.

The first error has received more attention in philosophical discussions of scientific models. Natural phenomena often exhibit regularities that yield themselves to mathematical descriptions, and science has increasingly relied on mathematical models to describe and predict them. According to causal realists, there are strong reasons to be leery about reading ontology directly off of mathematical models.[26] The most general reason is that it seems to put the descriptive cart before the ontological horse. It is the nature of the phenomena of interest (along with our epistemic and perhaps aesthetic interests) that determines whether a model is appropriate or accurate and not vice versa. Further, because models contain idealizations and abstractions, information is lost in the process of modeling; a methodologically useful model may leave out details that would be important for ontological judgments.

A related problem, raised by Millstein et al. (2009) is that mathematical descriptions underdetermine ontology. They argue that "it is a mistake to derive definitions from mathematics alone… since *many*, very different definitions can be derived from the same equation", and these include both physical and purely mathematical interpretations (p. 4). Two illustrations of this underdetermination will be helpful.

Millstein et al. offer the example of the Hardy-Weinberg law which states that an infinite, randomly mating, diploid population containing alternative alleles A and a at frequencies p and q, will maintain equilibrium genomic frequencies $p^2 + 2pq + q^2 = 1$ (where p^2 denotes the AA genotype, $2pq$ the Aa genotype, and q^2 the aa genotype). In

---

[26] The question of whether we should be realists or instrumentalists about models is somewhat orthogonal to this debate. Antirealists will argue that it is a mistake to go beyond the models to make any judgments whatsoever about ontology. I will not respond to this argument since none of the parties to the debate in question are antirealists. Instead, here I defend the more modest claim that if we are engaged in the practice of making ontological judgments in science, then we should not read them directly off of mathematical models.

its biological interpretation, this equation represents the outcome of causal features in the population, namely random mating in the absence of selection or drift. However, there are other phenomena that this equation can be used to describe, such as the area of a square with sides of length $p + q$. This example shows that the mere fact that a process can be represented mathematically does not entail that it is a purely mathematical (non-causal) relationship (Millstein et al. 2009, p. 4).

Another example comes from vector addition in Newtonian physics. Suppose that Jack and Jim are engaged in a test of strength. Jack attempts to push an object north (on a frictionless plane, of course), and Jim tries to push it south. Jack, being the stronger of the two, exerts 1,000 Newtons (N) on the object while Jim can only manage to push with 800 N of force. We can model this situation using a free-body diagram:

$$F_{Jack} = 1000 \text{ N}$$

$$F_{Jim} = 800 \text{ N}$$
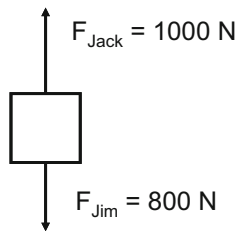
It follows from standard vector addition $(1{,}000 \text{ N} + -800 \text{ N})$ that the resultant force on the object will be 200 N north. This model can be interpreted as representing Jack and Jim's test of strength, implying that there is a cause of 200 N north acting on the object, and predicting that the object will move north. However, there are plenty of other causal and mathematical relationships that could be represented equally well by the model alone (for example, the simple mathematical fact that $1000 + -800 = 200$). Should we infer that the relationship between Jack's and Jim's pushes was purely mathematical or that to explain the resulting motion of the object with the above free-body diagram is to give a purely mathematical explanation? Neither of these claims seems plausible.[27]

Of course, this is not to deny the importance of mathematical models in ontological investigations. Models, and empirical generalizations more broadly, provide theorists with information about the structural features of a causal process under investigation which in turn offers useful restrictions on the causes that are suitable candidates for explanations of those empirical generalizations.

However, this role of mathematical models suggests the second error in reading ontology directly off of mathematical models. If one starts with an appropriate and accurate model of a phenomenon, one can proceed in looking for causes that manifest the formal structural properties of that model. However, if one starts with an inappropriate, incomplete, or inaccurate model, then this process will often lead to errors in ontological investigations. This is the problem raised by alternative models in the debate over the causal status of drift.[28]

---

[27] Though see Lange and Rosenberg (2011) for a defense of the latter claim's plausibility.

[28] Another case of a methodological device in population genetics being given an erroneous ontological interpretation is discussed in Clatterbuck et al. (2013).

The Wright–Fisher model is an appropriate and accurate model of some kinds of drift. It also has considerable methodological virtues; it is mathematically and intuitively tractable, and since it was the first major model of drift, it has been widely used and developed. However, some of these methodological advantages come at the cost of simplifying assumptions and idealizations that are inappropriate for many actual populations. As the work of Der et al. shows, these assumptions are not harmless. Not only will the Wright–Fisher model deliver inaccurate empirical predictions in some cases[29], it has also led some to erroneous beliefs about what drift is.

One remedy is for theorists to consider alternative models of the same phenomena and carefully elucidate the ontological assumptions made by each. It might not always be feasible to determine whether these ontological assumptions are true in all instances. However, an easier task is to determine whether those assumptions are invariant across models of the same phenomenon. If they are not, then we ought to tread carefully in making ontological pronouncements on the basis of just one of the many models that are available.

References

Abrams, M. (2007). How do natural selection and random drift interact? *Philosophy of Science*, *74*(5), 666–679.

Brandon, R. N. (2005). The difference between selection and drift: A reply to Millstein. *Biology and Philosophy*, *20*(1), 153–170.

Cannings, C. (1974). The latent roots of certain Markov chains arising in genetics: a new approach, I. Haploid models. *Advances in Applied Probability*, *6*, 260–290.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, *10*(3), 195–205.

Clatterbuck, H., Sober, E., & Lewontin, R. (2013). Selection never dominates drift (nor vice versa). *Biology & Philosophy*, *29*, 1–16.

Coyne, J. A., Barton, N. H., & Turelli, M. (1997). Perspective: a critique of Sewall Wright's shifting balance theory of evolution. *Evolution*, *51*, 643–671.

Der, R., Epstein, C. L., & Plotkin, J. B. (2011). Generalized population models and the nature of genetic drift. *Theoretical Population Biology*, *80*(2), 80–99.

Der, R., Epstein, C., & Plotkin, J. B. (2012). Dynamics of neutral and selected alleles when the offspring distribution is skewed. *Genetics*, *191*(4), 1331–1344.

Eldon, B., & Wakeley, J. (2006). Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*, *172*(4), 2621–2633.

Filler, J. (2009). Newtonian forces and evolutionary biology: A problem and solution for extending the force interpretation. *Philosophy of Science*, *76*(5), 774–783.

---

[29] While an exploration of the empirical ramifications of alternative drift models is beyond the scope of this paper, I suspect that the work of Der et al. will be fruitful in making more accurate predictions about drift. For instance, debates over the plausibility of Wright's Shifting Balance Theory have appealed to features of Wright–Fisherian drift—such as the probability of fixation of a novel mutant in a small population, time until fixation of neutral alleles, and the interaction of selection and drift—which may be different if the populations are undergoing other types of drift (for an overview of the debate, see Coyne et al. (1997)).

Gildenhuys, P. (2009). An explication of the causal dimension of drift. *The British Journal for the Philosophy of Science*, *60*(3), 521–555.

Hedgecock, D. (1994). Does variance in reproductive success limit effective population size of marine organisms? In A. R. Beaumont (Ed.), *Genetics and evolution of aquatic organisms* (pp. 122–134). London: Chapman and Hall.

Hodge, M. J. S. (1987). Natural selection as a causal, empirical, and probabilistic theory. In L. Kruger, G. Gigerenzer, & M. Morgan (Eds.), *The probabilistic revolution* (Vol. 2, pp. 233–270). Cambridge, MA: MIT Press.

Karlin, S., & McGregor, J. (1964). Direct product branching processes and related Markov chains. *Proceedings of the National Academy of Sciences of the United States of America*, *51*(4), 598.

Kimura, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics*, *47*, 713–719.

Lange, M. (2013a). Really statistical explanations and genetic drift. *Philosophy of Science*, *80*, 169–188.

Lange, M. (2013b). What makes a scientific explanation distinctively mathematical? *British Journal for the Philosophy of Science*, *64*, 485–511.

Lange, M., & Rosenberg, A. (2011). Can there be a priori causal models of natural selection? *Australasian Journal of Philosophy*, *89*, 591–599.

Matthen, M., & Ariew, A. (2002). Two ways of thinking about fitness and natural selection. *Journal of Philosophy*, *99*(2), 55–83.

Matthen, M., & Ariew, A. (2009). Selection and causation. *Philosophy of science*, *76*(2), 201–224.

Mills, S., & Beatty, J. (1979). The propensity interpretation of fitness. *Philosophy of Science*, *46*, 263–286.

Millstein, R. L. (2002). Are random drift and natural selection conceptually distinct? *Biology and Philosophy*, *17*(1), 33–53.

Millstein, R. L., Skipper, R. A, Jr, & Dietrich, M. R. (2009). (Mis)interpreting mathematical models: Drift as a physical process. *Philosophy & Theory in Biology*, *1*, 1–13.

Okasha, S. (2006). *Evolution and the levels of selection* (Vol. 16). Oxford: Clarendon Press.

Plutynski, A. (2007). Drift: A historical and conceptual overview. *Biological Theory*, *2*(2), 156–167.

Reisman, K., & Forber, P. (2005). Manipulation and the causes of evolution. *Philosophy of Science*, *72*(5), 1113–1123.

Shapiro, L., & Sober, E. (2007). Epiphenomenalism: The dos and the don'ts. In G. Wolters & P. Machamer (Eds.), *Thinking about causes: From Greek philosophy to modern physics* (pp. 235–264). Pittsburgh: University of Pittsburgh Press.

Sober, E. (1984). *The nature of selection*. Cambridge, MA: MIT Press.

Sober, E. (2011). A priori causal models of natural selection. *Australasian Journal of Philosophy*, *89*(4), 571–589.

Stephens, C. (2004). Selection, drift, and the "forces" of evolution. *Philosophy of Science*, *71*(4), 550–570.

Walsh, D. M. (2000). Chasing shadows: Natural selection and adaptation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *31*(1), 135–153.

Walsh, D. M., Lewens, T., & Ariew, A. (2002). The trials of life: Natural selection and random drift. *Philosophy of Science*, *69*(3), 429–446.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, 97–158.