

Infinite Population Models and Random Drift

Marshall Abrams*

Philosophers of science sometimes seem to imply that there are evolutionary models in which a counterfactual infinite population of organisms plays a crucial role. As is sometimes noted, this idea is incoherent if “infinite population” is understood literally. This paper uses case studies of modeling in evolutionary biology to examine roles that “infinite population”, and assumptions about random drift, play in modeling practices. Sometimes various effects of the absence of drift are understood as having to do with limits as population size goes toward infinity; in other cases these effects are conceptualized as having to do with large population sizes. Some models make assumptions about population size and effects of drift that might seem inconsistent: in some cases drift is included in part of a model, but excluded in another, or excluded even though population size is treated as finite. Because of such facts, I argue that there is no fixed set of assumptions associated with drift or its absence, and that there is no clear meaning for “infinite population” and similar terms. Rather “infinite population” is figurative language that is merely associated with various assumptions about the absence of drift.

Keywords


evolution • random drift • infinite population • model • idealization

1 Introduction

Philosophers of science often write as if evolutionary biology includes models that treat evolution as occurring in infinite populations. This, of course, doesn't mean that philosophers think that there are in fact populations that are infinite in size. Rather, what some philosophers seem to say is that evolutionary biologists can and do construct models in which the number of organisms in the model is countably infinite,¹ and that it is useful to model evolution in this way. This

1. A *countably infinite* set is one in which each member can be paired one to one with the positive integers, with no elements left over (example: the even positive integers). Continuous sets such as the real numbers between 0 and 1 are *uncountable*. A simple argument (Cantor's diagonal argument) shows that if one tries to pair all of those numbers with the positive integers, an infinite number of real numbers will be left out.

*College of Arts and Sciences - Philosophy, University of Alabama at Birmingham, mabrams@uab.edu

 <https://orcid.org/0000-0003-4749-1671>

Received 15 November 2023; Revised 8 June 2024; Accepted 9 June 2024

doi:10.3998/ptpbio.5266



treats an infinite population as an idealization—an assumption that is false of actual populations, but useful for understanding their properties. The idealization is useful, it's thought, because a population that is infinite in size would be one in which there was no genetic drift. Thus an infinite population can be used to model evolutionary influences other than drift—such as natural selection—without considering the influence of drift.

It's certainly correct that models that biologists *describe* as referring to infinite populations are extremely useful—as illustrations below show—and that the main reason that they are useful is because they model evolution without the influence of drift. However, there is significant vagueness in many philosophers' characterization of “infinite population” models and the assumptions on which these models depend. Although some philosophers give the impression that “infinite population” models are unrealistic *simply* because infinite populations of organisms don't actually exist, the idea of a population with a countably infinite number of organisms is in fact so incoherent that the idea, as such, can play no role in evolutionary biology. What biologists mean by talk of infinite populations is therefore subtly different from the idea of evolution in infinite populations per se. What is meant is shown by the roles that so-called infinite populations play in practice.

The structure of the rest of the paper is as follows. First, since random drift, infinite sets, and limits have played a number of roles in a philosophy of science, I want to forestall confusion or disappointment about the the focus of the paper. The following two subsections of this introductory portion of the paper clarify what my goals are—and are not. Then section 2 illustrates the kind of statements about infinite populations that I think are in need of clarification. Section 3 provides several conceptual and theoretical clarifications concerning “infinite population” models. I explain why one central idea of an evolving population of countably infinite size is incoherent and can have no use in evolutionary biology.² I use arguments by John Norton to clarify this point, and distinguish related but different ideas about populations with infinite numbers of states. Section 4 discusses a series of case studies of models from population genetics.³ These models illustrate roles that infinite population models and closely related models play in evolutionary biology. Some of the models are a little bit complicated, but that's the nature of contemporary evolutionary biology, and important points would be difficult to illustrate with fewer details. I try to describe the features of the models that matter to my arguments as simply as possible, however. Section 5 discusses several ways of understanding “infinite population” and roles of drift in the models discussed in section 4. In the end, I argue that “infinite population” and related terms are generally used figuratively, and are associated with a variety of potential consequences of drift and arguments concerning it. This view accommodates details of practice that don't seem to be handled by other non-literal treatments of “infinite population”. I suggest that more refined analyses would be possible, but would require varieties of research on “infinite population” models that go beyond what has been undertaken previously.

1.1 Goals

One purpose of this paper is to provide a clearer characterization of what biologists' talk of infinite populations means in practice, and to explore different ideas and strategies associated with

2. This point has been discussed more briefly by Abrams (2006) and Strevens (2019).

3. The models I discuss come from both primary sources and textbooks. Some philosophers of science seem to disdain analysis of claims from textbooks. Science textbooks can be poor guides to the history of science, but this paper is about contemporary scientific understandings, and both primary sources and textbooks provide information about contemporary thinking. One might think that textbooks are poor sources because authors sometimes say things in overly simplified ways for the sake of pedagogy. However, whether discussing textbooks or primary sources, one has to go beyond that what scientists say they are doing, and look more deeply at their practices.

“infinite population” talk in evolutionary biology. Though philosophical discussions of so-called (countably) infinite population models correctly treat these models as assuming the absence of genetic drift, these discussions usually fail to make clear that an infinite population is neither a necessary nor sufficient condition for the absence of genetic drift. This follows from the claim below that infinite populations, as such, can play no role in evolutionary modeling. Given that it is not the idea of evolution in an infinite population that supports an assumption that drift is absent, it’s valuable to explore exactly what biologists do mean when they speak of infinite populations. It turns out that there are subtly different roles that assumptions about drift play in “infinite population” models. The case studies I present illustrate some of these roles. My general suggestion will be that “infinite population” talk among biologists is merely idiomatic—and that sophisticated evolutionary biologists implicitly know this. “Infinite population” assumptions have different consequences in different cases, but they at least function as ways of stating that *particular* effects of random drift are not represented in a given model. A further purpose of the paper is to illuminate and discuss ways of understanding different roles that assumptions about drift or its absence play in models of evolution. Though the idea of an infinite population as such is incompatible with even hypothetical evolution, other kinds of apparent inconsistency within models can be quite useful. For example, I’ll look at a model in which drift is described as both present and absent, though in different parts of the model. I’ll suggest that such modeling practices are fruitful, and that phrases such as “infinite population” need not have clear, unambiguous meanings in order to be useful in evolutionary biology.

1.2 *Non-goals*

Given (sometimes reasonable) misunderstandings or unfounded expectations about earlier versions of this paper, I suggest readers at least skim this section—which summarizes issues that are not part of the focus of the paper.

It’s not about the math per se While one of my goals is to elucidate ways that talk of infinite populations depends on well-known mathematical facts about models and probability, I have nothing very novel to say about the mathematics as such. I intend my discussion to clarify and illuminate subtleties of relationships between infinite population talk, modeling, and large populations in ways that I believe are illuminating for philosophers of science and some biologists.

It’s not historical The focus of this paper is on illuminating recent scientific practice; it’s not a historical study. There is a valuable body of historical and philosophical work on modeling, statistical methods, and empirical applications in evolutionary biology during the early and middle years of the 20th century. Some of this research concerns roles of large populations in one way or another, and some of it touches on talk of infinite populations. Historical work on this period wouldn’t in itself tell us what parts of the earlier biological research were relevant to contemporary biology. For example, if it’s true, as is sometimes suggested, that Fisher discussed infinite population models in early 20th-century works (such as Fisher, 1922, [1930] 2000),⁴ one can’t simply assume that meanings and uses of contemporary talk of infinite populations today are fixed by Fisher’s (cf. Kuhn 1962).

4. Despite a great deal of historical and philosophical discussion related to Fisher’s and his contemporaries ideas about very large populations (e.g., Wimsatt 1981; Provine 1986; Turner 1987; Hodge 1992; Morrison 2000, 2009, 2015; Skipper 2002; Plutynski 2004, 2005a,b, 2008; Winther et al. 2013, 2015; Ishida 2017), as well as some attributions by biologists (Kimura and Crow, 1964, 728), my own reading of Fisher suggests that he did not discuss infinite biological populations in the papers usually cited in support of this claim. That is a topic for a different paper, however.

It's not about philosophical drift/force debates This paper is not primarily a contribution to ongoing philosophical debates about what drift is, whether it exists, or exactly how it is or is not related to natural selection. In philosophy of biology, some discussions of drift are framed as focused on whether and in what sense drift is an evolutionary “force”.⁵ There are important issues here, and some of my thoughts about them have appeared elsewhere (Abrams, 2007, 2023). In this paper, I'll simply assume that there's something called drift that involves random effects, so that I can discuss ways that evolutionary biologists model it. I'll take for granted here that it's useful to discuss distinct types of causal factors influencing evolution, and I'll sometimes use “force” as an informal term for such factors, as many biologists do.

It's not an intervention in modeling debates There are two areas of philosophical research to which this paper will rightly seem especially relevant. First, some philosophers of science have developed general views about what scientific models are (see, e.g., Weisberg 2013; Gelfert 2016), for example that they are fictions of a certain kind (see, e.g., Suárez 2009a). Second, there has been a great deal of discussion of models incorporating infinite structures or limits, mostly in physical sciences (e.g., Fletcher et al. 2019). My discussion may have implications for such areas, and I'll mention a few connections to relevant literature here and there. However, given the length of this paper, I want to leave extensive discussion of work in these areas for other papers. My theorizing here will remain close to the texts of the case studies I discuss and the variety of relationships toward the absence of drift that they illustrate.

It's not a criticism of biological practices It should be clear below that nothing I say is intended to criticize the models that biologists have been using. I assume—with good reason, I believe—that these models are useful tools, are not used in incoherent or inconsistent ways, and that they can be used to gain insight or test hypotheses about real populations. This assumption is the starting point of this paper. On the other hand, I don't believe that everything scientists say about their practices is literally true, but that need not be seen as a failing. Sometimes one must closely examine ways that scientific tools are used in order to understand the truth about them. Occasionally it turns out that this truth conflicts with what some scientists seem to say about the tools. But when scientific practices are coherent and fruitful, leading to increases in knowledge, it is probably not a fault in scientific practice that relatively loose language has played a useful role in that kind of success. Thus, I hope that scientists reading this paper will bear with me if it sometimes looks as if I'm trying to undermine or misinterpret valuable ideas or practices in evolutionary biology. My goal is to try to give a clearer understanding of those valuable ideas and the practices they depend on.⁶

5. The relevant literature on evolutionary forces in philosophy of biology is ongoing and multifaceted, and overlaps with debates about other issues, including how to define fitness, what natural selection is, and what drift consists in (e.g., Bourrat 2018; Clatterbuck et al. 2013; Clatterbuck 2015; Gildenhuys 2009; Huneman 2012; Matthen 2010; McShea and Brandon 2010; McShea et al. 2019; Millstein 2002; Millstein et al. 2009; Otsuka 2016; Godfrey-Smith 2009b; Pence 2017, 2021; Plutynski 2007; Ramsey 2013; Shapiro and Sober 2007; Walsh et al. 2017).

6. One might wonder why I suggest that philosophers' statements about infinite populations are in need of clarification, while allowing that it's OK if biologists' statements about infinite populations are vague and even literally false. The difference is that biologists are engaged in modeling practices that go beyond what they sometimes say, and so some of what scientists say along the way is less important than their practices and the conclusions those practices justify. Philosophers of science, however, are engaged in understanding scientific concepts, methods, practices, theory, and how these affect justification and our understanding of the world. In that context, understanding of nuances of scientific theory and practice is a component of the main goals of research, so it's important to clarify potentially misleading language.

2 Philosophers on infinite populations

I said above that I think that philosophers have written about “infinite population” models in ways that are somewhat vague. Because certain kinds of claims about infinite populations are so routine, it will be helpful to illustrate exactly what I mean. I’ll italicize phrases that seem to imply that models themselves incorporate infinite populations—i.e., that the models treat populations represented within the model as literally infinite.

Angela Potochnik (2017, *passim*) uses infinite population models as one of her paradigmatic illustrations of an idealization—again, an assumption of the model that is known to be false of systems that are modeled (55).⁷ Potochnik writes, for example, that “*Representing a population as infinite in size, ... can be epistemically acceptable when researchers want to understand certain other features of these phenomena*” (99). Margaret Morrison wrote that the Hardy-Weinberg law “... relates allele or gene frequencies to genotype frequencies and states that *in an infinite, random mating population, ...*” (2015, 35f). Michael Weisberg writes that “the entire mathematical argument [in a paper by Karlin and Feldman] is made in terms of extremely general *properties of infinite populations*” (Weisberg, 2013, 65). Elliott Sober uses the idea of infinite population models in evolutionary biology for a similar purpose, and says that “evolutionary biologists consider models that assume that *populations are infinitely large*” (2008, 80). One kind of implication of the arguments below is that contrary to what Potochnik, Morrison, Weisberg, and Sober wrote, scientists do *not* represent “a population as infinite in size” (Potochnik), the Hardy-Weinberg law does not make a claim about would happen “in an infinite ... population” (Morrison), Karlin and Feldman did not make an argument “in terms of ... properties of infinite populations” (Weisberg), and that biologists do not “consider models that assume that populations are infinitely large” (Sober). Every one of these statements, understood literally, is false, I’ll argue. I am not saying that it is false that actual populations are infinite—it’s obvious that they are finite. All of the authors quoted know this. What I am saying is that it’s false that the models *represent* evolution using the idea of an infinite population of the kind to which these authors’ statements allude. (In section 3.3 I mention different sorts of models that one might refer to as “infinite population” models.) And it’s false that the idea of such an infinite population plays any role in a model of evolution or in details of how it is used. It’s possible that that these authors know this, and intend a non-literal reading of their words, but if so their language in these passages obscures that point.⁸ Most other philosophical discussions of “infinite population” models I’ve found, with exceptions noted below, exhibit similar problems. My hope is that this paper will show that, at the very least, additional subtlety about “infinite population” talk is useful.

It’s worth looking at a more specific claim that Potochnik made, which shows the apparent importance of the idea of an infinite population in some perspectives. Potochnik wrote that

Evolutionary game theory models hardly ever specify that the population is infinite in size. Instead, this idealization is generally left implicit. Nonetheless it is a requisite assumption whenever genetic drift is not taken into account. Any time

7. Norton’s (2012; 2014; see below in §3.2), Godfrey-Smith’s (2009a), and Appiah’s (2017) similar uses of “idealization” are related to but not equivalent to the Sober/Weisberg/Potochnik sense (Weisberg 2013, 98; Potochnik 2017, 55, *passim*; Sober 2008, 80) that I adopt here (except where noted). Morrison’s (2015, 20) definition of “idealization” is different from the Sober/Weisberg/Potochnik concept.

8. Sober plausibly intends a non-literal reading in the quotation above, since in his earlier classic *The Nature of Selection*, he wrote that “... the meaning behind the idea in population genetics that models in which selection acts alone are described presuppose infinite population size” (Sober, 1984, 44) is that, in the limit as population size increases, the values of probabilities in terms of which drift is quantified go to zero (see also 110). Sober’s claim, that talk of “infinite populations” expresses the idea the influence of drift goes to zero as population size increases without limit, is one of those I discuss below. I’m not aware that the other authors quoted gave this kind of clarification.

drift is neglected, the population is represented as if it were infinite. (Potochnik, 2017, 55)

Notice that what Potochnik claims is that *if* drift is neglected, *then* “the population is *represented* as if it were infinite” (my emphasis). I agree that when drift is left out of a model, it is sometimes said that it makes use of an infinite population. My claim will be that a biological population is never *represented as*, taken to be, treated as, etc., being (countably) infinite. Further, Potochnik says that this is a “requisite” assumption when there is no drift, suggesting that a necessary condition for the absence of drift is that a population is infinite. One of my points will be that an infinite population can’t be a necessary condition for the absence of drift, because the idea of evolution in an infinite population is incoherent. Drift can be neglected in a model even though the population is treated as finite.

3 What infinite population models are, and are not

3.1 *Why there is no evolution in (countably) infinite populations*

Let’s start with some basic terminology. An allele⁹ is a genetic variant at a particular locus, a location on a chromosome. Some organisms in a population may have one or more other alleles at the same locus. We can also focus on genotypes, combinations of alleles at one or more loci, or even phenotypes—traits—if they can in some sense be transmitted from one organism to another. I’ll take natural selection to involve probabilities of changes in frequencies of such things in a population. A standard claim is that what is selected for has a higher probability of increasing in frequency than other mutually exclusive alleles/genotypes/phenotypes *if no forces other than selection are acting*. Compare natural selection with random drift, which is inversely related to population size:¹⁰ The smaller the population, the more that drift acts to add randomness to outcomes, for essentially the same reason that smaller samples have higher variance. Pure natural selection is known as a “deterministic” force that acts alone when a population is infinite; it is adulterated by drift when the population is finite.¹¹ Here is a quotation from a population genetics textbook that illustrates some of these ideas:

Unless some of the parameters, such as the *selection* coefficients and *mutation* rates, required to specify the above evolutionary forces fluctuate at random, these forces will be *deterministic*. In a *finite population*, however, allelic frequencies will vary probabilistically because of the random sampling of genes from one generation to the next. This process is called *random genetic drift*, *random sampling drift*, or simply *random drift*. (Nagylaki, 1992, 3, emphasis mine except in the last sentence)

Even without understanding all the technical terms, one can see that Nagylaki contrasts “deterministic” forces with what happens in a finite population, seemingly implying that deterministic forces would be what one would get in an infinite population. Nagylaki’s statement suggests that in an infinite population, natural selection (whose strength can be measured by selection

9. This is one of several meanings of “gene”.

10. Or to “effective population size”, a concept that allows one to extend the use of certain mathematical models to a variety of realistic situations. Nothing will turn on this distinction below. In particular, in many—probably most—modeling contexts, population size increases without bound if and only if effective population size does.

11. Natural selection is never deterministic in a strict, physical manner, even if we understand infinite populations in terms of limits (see below). For example, as population size $N \rightarrow \infty$, the probability that the most fit trait will go extinct goes to zero, yet it never becomes impossible for the trait to go extinct: the set of scenarios in which the trait goes extinct is not empty. Deterministically produced effects, in this sense, need not occur (e.g., Sober 1984, 111).

coefficients) and mutation (whose strength is measured by mutation rates) act to change or maintain changing (relative) frequencies in a biological population in a deterministic way. When a population is finite, however, random drift introduces probabilistic variation in frequencies.

All the same, it requires only a little thought to see that the idea that of natural selection or mutation-caused changes in an infinite population is absurd. As I remarked earlier, this point has nothing to do with the fact that no actual biological population is infinite (e.g., Plutynski 2004; Giere 2009; Matthewson and Calcott 2011), nor that as a representation of a finite population, an infinite population model would attribute false properties to it (Potochnik, 2017). Since population genetic models are nearly always intended as approximate characterizations of their targets, and may involve a number of idealizations, the fact that an infinite quantity is used to model a large one need not in itself be problematic.

However, as Strevens (2019) and I (Abrams, 2006) implied earlier, the only relative frequencies definable in a (countably) infinite population are 0 and 1. For example, how many elements from an infinite set are needed to give us a relative frequency of $1/2$? If we're willing to consider arithmetic operations involving infinity at all, we should say that half of infinity is infinity, and that for any finite number n , $n/\text{infinity} = 0$. However, $\text{infinity}/\text{infinity}$ is undefined, in general. So there is no subset of elements of an infinite set whose relative frequency is $1/2$. The same point can be made about any number greater than 0 and less than 1. Yet evolution has to do with changes in frequencies of types—usually frequencies other than 0 and 1. Thus evolution in a countably infinite population makes no sense.¹² Again, the point is not simply that a model that supposedly represents evolution in an infinite population represents the world falsely. It's not even that the idea of evolution in an infinite population is internally inconsistent, although that might be enough of a problem. The point is that there is simply no way to represent an absolutely central idea of natural selection—the possibility of change in relative frequency—in an infinite population. There is no way that a model in which there is supposed to be evolution in an infinite population, *as such*, can do any work in helping us to understand evolution. (“As such”, because I’ll suggest below that there may be less central, informal roles that ideas about infinite populations do play.)

3.2 Norton on idealization

Norton’s distinction between approximation and idealization in physics provides a useful perspective on talk of infinite populations in evolutionary biology. For example, Norton writes:

An *approximation* is an inexact description of a target system

e.g., of a real-world system. On the other hand, an idealization (in a different sense than defined above)

is a real or fictitious system, distinct from the target system, ...

such that

12. There are also models known as “infinitely many sites” and “infinitely many alleles” models (e.g., Ewens 2004). In an infinite sites (alleles) model, it’s assumed that the same locus (allele) never appears more than once in the model. What makes these models work is not even that the number of loci or alleles is large, let alone infinite; it’s simply that there is always another locus (allele) available. The models don’t require a change of frequencies in an infinite set, so the fundamental problem with the idea of an infinite population model is absent. Moreover, the papers (Kimura, 1969; Kimura and Crow, 1964) that introduced these models (Tajima, 1996) focused on finite populations. A more recent treatment in an advanced textbook (Ewens, 2004) usually assumes a finite population in these models as well. See also note 38.

an *exact* description of the [real or] fictitious system for aspects of interest turn out to provide an *inexact* description of the target system. (2014, 199, emphasis in original)

Thus, an approximation is a statement or a collection of statements that describe(s) a system in a way that is not fully accurate, while an idealization, in Norton's sense, is a second system, perhaps imaginary, whose precise description would provide an approximation for a real system. Norton points out that the fact that a system has an idealization implies that it has a corresponding approximation, but that the reverse implication need not hold.

One of Norton's illustrations of this distinction uses a model of a reversible process in thermodynamics, which represents heat as being transferred without an increase in entropy (Norton, 2014).¹³ In the model, one can let the temperature difference between two objects go to zero while simultaneously letting the elapsed time go to infinity. The model is reasonable for all nonzero values of temperature difference and all finite values of time as the limits are approached, but the idea of a system in which these limits are actually achieved is "nonsense" according to Norton (2014, 202). In such a system, basic physical principles would imply that the amount of heat transferred was represented by zero times infinity, an undefined quantity. This is why Norton calls the idea of a system in which the limit is actually achieved is nonsense.

We might apply Norton's distinction to countably infinite populations in this way: An infinitely large population of organisms in which allele frequencies (for example) can change over time might be an idealization, if it could *possibly* exist. However, it can't. It would have to have an impossible combination of properties.¹⁴ We could still say that phrases like "infinitely large population" provide ways of identifying models that are approximations to large populations. These models include, for example, those in which a real population is large enough that it can be treated as if frequencies of alleles in one generation directly determine the frequencies of alleles in some later stage of a model.

3.3 Other "infinite population" models

Some models could easily be confused with the ones referenced above and discussed in the rest of the paper, so it's worth briefly characterizing these other models in order to put them aside. Given the large number of modeling strategies that I want to exclude at this point, I hope readers will allow me to make broad generalizations about existing (and sometimes merely possible) evolutionary models in this section without referencing any models explicitly. The next sections of the paper will return to particular models of the kind that are my focus, discussing some of them in detail.

Note, first, that most of the points in the previous two sections could apply to an uncountably infinite population as well, but in this case the idea of the relative frequency of organisms with a particular trait has even less of a direct connection to real populations. For example, suppose that a model were to treat the number of organisms in a population as the number of points between 0 and 1, viewing the number of organisms with a trait *A* as the number of points between 0

13. At one point Norton (2012) talks about a biological case—a model of the growth of a population of bacteria—but his discussion is only indirectly relevant to my concerns here.

14. Here I am using the broad philosophical sense of possibility according to which a system, or more commonly a universe or "world", is possible if and only if there is no inconsistency in it. More specifically, there are no correct descriptions of the possible world that require that a statement be both true and false of that world. (Section 5.1 discusses McLoone's suggestion that an impossible population could nevertheless play a role in a model through use of special semantics that allow one to incorporate inconsistent propositions. I argue in that section that even if that strategy were pursued for infinite populations, the idea of an infinite population could do no work to aid understanding of evolution.)

and 0.4. Of course $0.4/1.0$ is well-defined, but this hypothetical model doesn't represent the number of organisms with A as the number 0.4; it represents that number of organisms as uncountably infinite. Dividing the uncountably infinite number of points between 0 and 0.4 by the uncountably infinite number of points between 0 and 1 is of course undefined.

However, biologists can and do create useful models in which it's stipulated that real numbers between 0 and 1 (inclusive) represent frequencies in a population. For example, 0.4 can represent the idea that 400 out of $N = 1000$ organisms in a population have trait A . In this case the number 0.4 doesn't represent a subset of points in $[0, 1]$. The number 0.4 is also not a relative frequency per se: it's not *defined* as a ratio between an integer representing a number of organisms and a population size. It's just a real number stipulated to *represent* relative frequencies indirectly, and often approximately. Such real numbers can then be mathematically manipulated according to rules specified in the model. Though there is an (uncountable) infinity of points between 0 and 1, representing frequencies in this manner doesn't imply that one is modeling the population as an infinite set of organisms.¹⁵

Let's consider randomly sampling from a population. This is relevant to some examples described below, and it's of interest because models of random drift are often conceptualized as involving random sampling from some set of organisms. Let's start by considering the idea of sampling from either a finite or a countably infinite population. Here is one way to randomly choose an element from a finite set of size N : give each element equal probability $1/N$, and select elements with that probability. With a countably infinite set, N is infinity, and as noted above, $1/\infty = 0$, so the simple rule for random selection that works with a finite population would give each element of an infinite set zero probability of being chosen. That is, the obvious idea that we can sample organisms by giving each of them an equal probability doesn't work when N is infinity.

However, a model can also stipulate that there is a probability distribution over traits shared by many organisms, and then choose an organism with trait A with probability $P(A)$. (This stipulated distribution need not be derived from relative frequencies over a population size in the manner illustrated in the previous paragraph.) One can use these probabilities to model trait frequencies more directly. One way to do this is to view $P(A)$ as the probability that an organism is randomly sampled from the population with probability $P(A)$. This makes sense as a representation of relative frequency, because if one chooses a member of a finite population, giving each organism an *equal* probability of being sampled, the probability $P(A)$ of choosing an organism with A would be equal to the relative frequency of A in that population. By starting with the probability distribution Pr , a model can abstract from details about particular population size, however.

Notice, now, that probabilities such as $P(A)$ can be allowed to vary continuously from 0 to 1. So an uncountably infinite number of possible population states can be represented using such probabilities. However, as noted above, the infinity of possible states doesn't imply that the modeled population is infinitely large. The model simply uses a form of mathematical representation with an uncountably infinite number of possible values, to provide approximate

15. Throughout this section, most of my remarks target models in which a population has a potentially uncountable number of states. In theory, similar models could represent states in a finite population of any size using rational numbers with large finite denominators. The set of rational numbers is countable, though, not uncountable (e.g., https://simple.wikipedia.org/wiki/Cantor's_diagonal_argument). My remarks in this section would apply to such models, if there were any, because they would represent states of a population as involving an infinite number of states that don't correspond directly to absolute frequencies of organisms. I ignore this kind of possible model for the sake of keeping language simple. (In practice computer models use rational numbers with large denominators, but the point of doing that is usually to approximate real numbers.)

representations of relative frequencies in a population that in real cases will have finite size.¹⁶

The important point is that while in general (a) it can be entirely legitimate to model a population using a continuously varying quantity that represents trait frequencies, (b) doing so does not have to force the assumption that random drift is absent. This is because though such models represent changes in frequency using a countably infinite number of values, this representation has no intrinsic connection to population size. It abstracts from population size. In fact, in diffusion models of evolution, a common kind of model with continuous variation in population states, drift is required by the formalism except in degenerate cases. In diffusion models, it is, rather, directional influences such as natural selection that are optional additions.¹⁷

In the rest of the paper, except where noted, I'll use "infinite population" to refer to countably infinite populations that are infinite in size because they include an infinite number of discrete organisms. This is the idea of an infinite population that's associated with the idea that drift fails to influence the population in some respect.

4 Infinite population models and their kin: case studies

Section 3.1 explained why certain so-called infinite population models of evolution can't in any sense incorporate infinite populations. Yet evolutionary biologists do often describe models as involving infinite populations—in ways that I believe have motivated philosophical claims about infinite populations. Widespread modeling practices that are described as involving infinite populations must be meaningful and useful. What's going on? In this part of the paper I look at several "infinite population" models and some closely related models in order to work toward answers.

4.1 Terminological background

My first example is just a little bit complicated, but I believe it's important to get a sense of how concepts are used in research, and not only in contexts where pedagogical constraints might distort normal practices. Vagne et al. (2015) investigated the evolution of genes at two loci when those genes interact to produce phenotypes in a manner known as "reciprocal sign epistasis". These authors were interested in cases in which the resulting phenotypes' fitnesses exhibited a pattern known as truncation selection. I'll start by introducing two terms, "epistasis" and "linkage disequilibrium". The latter will come up repeatedly in the case studies below. Readers already familiar with these concepts should feel free to skip to the next section.

Epistasis *Epistasis* occurs when genes at different loci interact to produce an organism's phenotype, and the contributions of the two loci can't be decomposed into separate quantities of influence that can be added to produce an effect of a given size.¹⁸ For example, two loci that have an additive, non-epistatic effect on height are ones in which the alleles at one locus add

16. One *could* view a model that represents trait frequencies using numbers that vary continuously between 0 and 1 as a model in which there was an uncountably infinite number of organisms. But here again, that assumption by itself is unlikely to do any work; only the mathematical terms and manipulations do. A modeler could assume that there is no drift in the population, but that assumption is not required by of the representation of relative frequencies as probabilities with an infinite number of possible values.

17. In probability theory, the coefficients in diffusion models that represent directional influences such as natural selection are called "drift coefficients" (Grimmett and Stirzaker, 2020, §13.3). One sometimes finds this mathematical use of "drift" in discussions of evolutionary diffusion models (e.g., Gillespie 1974).

18. There is some ambiguity in "epistasis". Sometimes it refers to an effect that cannot be decomposed into influences whose values can be *multiplied* to produce the overall effect size.

or subtract the same amount from height regardless what allele is at the other locus. Suppose instead that the amount that's added to height by the change from one allele to another at locus A depends on which allele is at locus B. That means that the loci interact epistatically. *Sign epistasis* occurs when the interaction between two loci is such that the *direction* of the effect (e.g., on height) of an allele at one locus depends on alleles at the other locus.

Linkage disequilibrium *Linkage disequilibrium* is a nonzero population-wide correlation between alleles at two loci. In Vagne et al.'s (2015) model, the hypothetical organisms' genome includes two loci that the authors call A and B. Each organism has one of two alleles at each of these loci: an organism can have either allele A or allele a at locus A, and either allele B or allele b at locus B. Linkage disequilibrium between A and B occurs, for example, when they are found together in the same individuals more often than would be expected if all of the A and B alleles in the population were simply randomly shuffled together. An extreme case of linkage disequilibrium occurs when every individual has either both A and B or else has both a and b.

Linkage disequilibrium can occur for various reasons, including selection on epistatic effects, loci near each other on a chromosome (genetic linkage), and various chance processes. I discuss some of these factors below. One reason that linkage disequilibrium is important is because when there is selection for an allele that's found more often with a selectively neutral allele, this second, neutral allele can increase in frequency as a result—as if it were under selection, even though it makes no difference to survival or reproduction. In some contexts, this is called genetic “hitchhiking” (see below), and it's the basis of recent studies in which complex statistical patterns of linkage disequilibrium provide evidence for natural selection (e.g., Sabeti et al. 2002; Voight et al. 2006; Gazal et al. 2017; Goszczynski et al. 2018; Lynch and Walsh 2018).

4.2 Case 1: fixed-effect frequency

Vagne et al.'s (2015) models are of populations of sexually reproducing organisms with haploid genomes. This means that each organism has only one copy of each chromosome, and Vagne et al. assume that when two organisms mate, recombination can occur: corresponding pieces of chromosomes from each parent may be probabilistically swapped to produce each offspring's chromosomes. The researchers specified relationships between phenotypes generated by the four genotypes AB, ab, Ab, and aB, so that AB's phenotype was fitter than ab's, which in turn was fitter than either aB's or Ab's phenotype. (One could abbreviate these relationships as $AB > ab > aB, Ab$, taking “ $_ > _$ ” to mean “ $_$ has a fitter phenotype than does $_$.”) This implements a kind of sign epistasis with respect to fitness since, for example, a change from a to A would produce a more fit or less fit organism depending on whether there was a b or B allele at the B locus. It is *reciprocal sign epistasis* because the AB and ab genotypes are considered opposite extremes, and they are fitter than either of the intermediate genotypes aB and Ab.

For each (discrete, nonoverlapping) generation in Vagne et al.'s models, the researchers calculated the frequencies of the four genotypes AB, ab, Ab, aB after selection on possible parents, and subsequent random mating between those organisms they allowed to reproduce. That is, Vagne et al. modeled *truncation selection* in which a fixed percentage of organisms was allowed to reproduce; these organisms had the genotypes that produced the fittest phenotypes. Frequencies were calculated using specific values for a parameter r that measures the probability of recombination between the A and B loci.¹⁹ Calculation of frequencies in the next generation

19. Specifically, r is the probability that, during reproduction, corresponding pieces of (homologous) chromosomes will be exchanged in such a way that a new chromosome will be formed in which the A allele from one parental chromosome will be combined with the B allele from the other parental chromosome.

also involved linkage disequilibrium, D , between A and B . That is, in a given generation, it could turn out that A and B or a and b occur together more often or less often than would be the case if their combinations were determined randomly. For example, $D = 0$ when the frequency of AB is equal to the frequency of A times the frequency of B ; that is, there is no linkage disequilibrium in this case.

The researchers repeatedly iterated this process of calculating frequencies from the previous generation's frequencies, and did this with a variety of model parameters. Their goal was to determine the conditions that would allow certain outcomes, and to determine properties of the evolving population. For example, one property of interest was the number of generations until one of the four genotypes went to fixation, i.e., until all members of the population had that genotype.

Vagne et al. explicitly described their models as involving infinite populations:

A three-step process was applied to gain insight into the evolution of a population selected for a trait subject to reciprocal sign epistasis. First, we ... determined the equilibrium states. ... Secondly, times to fixation were estimated by numerical calculations in random mating in *populations of infinite size*. Finally, simulations of *finite populations* enabled us to investigate the role of genetic drift. (Vagne et al., 2015, 45, emphasis added)

That is, after analyzing mathematical models with “populations of infinite size”, Vagne et al. performed simulations using populations of finite size. I'll focus on the former.²⁰

Examination of Vagne et al.'s mathematical models shows that what makes the populations into ones “of infinite size” is that the models include terms for relative frequencies of genotypes AB , aa , Ab and aB , and these terms enter into the calculation of which organisms are selected in a very direct way, as does the parameter r : the calculation of which combinations of alleles have which frequencies in the next generation is performed without the kind of stochastic fluctuation that would be expected in a finite population. For example, in one of Vagne et al.'s models, the frequency p_{ab}^{n+1} of the ab individuals in generation $n + 1$ is (Vagne et al., 2015, 47, under “Configuration 1”):

$$p_{ab}^{n+1} = \frac{p_{ab}^n}{p} - r D^{n'} \tag{1}$$

where p_{ab}^n is the frequency of genotype ab in generation n , and p is the proportion of individuals selected to be parents of the next generation. $D^{n'}$ is the linkage disequilibrium, the (frequency-based) correlation between A and B , in generation n , after selection. The formula for $D^{n'}$ is a little bit complicated, but I'll point out what is important to notice about it:

$$D^{n'} = \frac{p_{AB}^n p_{ab}^n}{p^2} - p_{aB}^n p_{Ab}^n \left(\frac{p - p_{AB}^n - p_{ab}^n}{p (p_{aB}^n + p_{Ab}^n)} \right)^2 \tag{2}$$

The new terms here, p_{AB}^n , p_{aB}^n , and p_{Ab}^n , are frequencies of AB , aB , and Ab respectively in generation n . The thing to notice about the calculation of p_{ab}^{n+1} (the frequency of ab in generation

20. The fact that in Vagne et al.'s model, fixation can occur in finite time in an “infinite population” without any further manipulation is an unusual feature of the model resulting from the role of truncation selection in it. Fixation occurs in the model in finite time because in each generation the fraction p of the population containing the fittest individuals is chosen to reproduce. So when the fittest genotype has a proportion of the population that's greater than p , all of the chosen individuals will have that genotype in the next generation. This occurs regardless of the size of the population; cf. note 27. (I'm grateful to an anonymous reviewer for pointing this out and pressing for clarification.)

$n + 1$) from (1) and (2), is that p_{ab}^{n+1} is calculated solely from frequencies in generation n , plus a fixed proportion p of individuals who reproduce, along with a constant recombination rate r . There is nothing stochastic in such a model. In particular, it does not include anything that could count as modeling genetic drift. In a real population, or in a model with drift, the relative frequencies in the current generation would typically influence the frequencies in the next generation, but the precise frequencies in the next generation would depend on how many individuals of each genotype happened to be selected stochastically from the current generation.

Thus, an infinitely large population *as such* plays absolutely no role in Vagne et al.'s "infinite population" model. Infinity plays no role in the mathematics. What is implied by "populations of infinite size" in Vagne et al.'s models is simply that frequencies in one generation can be calculated directly from frequencies in previous generations along with constant parameters, without any stochastic effects corresponding to genetic drift.²¹ (Later, when introducing their simulations, Vagne et al. say that "Drift was taken into account by setting a finite population size ($N = 1000$ and $N = 5000$)" (Vagne et al., 2015, 48).) We don't have to worry about paradoxes concerning relative frequencies other than 0 and 1 in infinite populations, because infinity plays no role in the models' calculations. Vagne et al.'s models, as they are used, involve no inconsistencies of the kind described in section 3.1.²² (This is not to say that we have to understand models as *merely* mathematical. At the very least, the variables in the Vagne et al. use are interpreted from the start.)

4.3 Case 2: equilibria²³

Zhao and Charlesworth (2016) also discussed a model having to do with linkage disequilibrium, in which a locus A that is not under selection can appear to be under selection if, by chance, it becomes correlated with a locus B that is under selection. Again the model assumes that there are two alleles at each locus. Zhao and Charlesworth write:

In the present case, the starting point is assumed to be a population of *infinite size, at equilibrium under mutation and selection* at the B locus and *with no LD* [linkage disequilibrium] between the two loci. It is thereafter maintained at a population size of N breeding individuals each generation. (Zhao and Charlesworth, 2016, 1317, emphasis added)

In contrast to Vagne et al.'s model, "infinite population" here does not imply that frequencies will be applied directly in calculating the makeup of the next generation. The "infinite population" assumption only applies to the initial state in this model—before calculations are iterated to model subsequent generations with finite size N . One reason for describing initial population as having infinite size here is that it is only in large populations that linkage disequilibrium will

21. Of course, that biologists sometimes leave stochastic effects out of models is not a new idea. In philosophy of biology, such practices were obviously part of Sober's (1984) motivation for a stronger claim, that evolutionary theory should be understood as a "theory of forces"—as illustrated by his discussions of natural selection and Hardy-Weinberg equilibrium.

22. This makes it sound as if population genetic models must have more coherence than I think they do in fact need in practice. I come back to this point in section 4.5.

23. The examples in sections 4.3–4.5 are of a general kind discussed by Wimsatt (1980) in his criticism of Williams' (1966) arguments for genes as units of selection. Wimsatt characterizes models similar to the ones I discuss below as motivated by reductionistic research heuristics. He argues that Williams' argument depends on the fact that the population genetic models that result from these heuristics are all biased in the same ways, because of the manner in which these heuristics are applied. Wimsatt's argument need not, however, be taken as a general criticism of models of the kind described here. Rather, it is a warning not to draw inappropriate conclusions from such models without noticing their limitations.

be near zero, since linkage disequilibrium can be generated by genetic drift (e.g., Charlesworth and Charlesworth 2010, 380, 383; cf. quote from Wang below). So “infinite population” implies that LD is equal to zero. However, setting LD to zero does not required modeling evolution in an infinite population per se. If we were to take the infinite size claim literally, it would be difficult to make sense of the claim that mutation and selection have been at equilibrium in the initial population. Mutation is something that happens in individuals, and the overall rate of mutation in a population is thus a function of the number of organisms in it (e.g., Gillespie 2004, 29, 33; Charlesworth and Charlesworth 2010, 43). Selection, similarly, depends on frequencies of offspring per individual. However, as in the preceding section, there is no infinite population as such in this model. Rather, the model simply includes assumptions—such as there being no linkage disequilibrium—that would be unlikely to hold if there were drift.

4.4 Case 3: fixed-effect frequency with nonzero minimum value

The examples in the two preceding sections were from theoretical—but empirically oriented—articles. Examples from textbooks are useful as well when they reflect experience and established wisdom from modelers. In this section, I consider an example from a well-known textbook that at first glance may seem to be in tension with some of my preceding remarks. Gillespie (2004) seems to model populations without drift in roughly the same way as in the example that I described in my section 4.2, yet Gillespie describes the population as finite. It’s worth looking closely at the text to see what he means.

I consider a model from Gillespie (2004, sec. 4.2). Gillespie does not describe this model as involving an infinite population; in fact the only reference to population size that does any work occurs in a description of parameters of a computer simulation used to generate a plot illustrating the model (110, cf. 117).²⁴ Gillespie says on page 110 that in the simulation, N (population size) was set equal to 5000. By contrast, near the end of the next section of the book (4.3), which discusses models closely related to the one discussed in Gillespie’s section 4.2, he explicitly says that a new model will be “for an infinite population ($N = \infty$) in order to remove all effects of genetic drift” (Gillespie, 2004, 115f). Nevertheless, although Gillespie doesn’t describe the model in section 4.2 as involving an infinite population, it includes no randomness—in particular, no genetic drift—and in that respect is like the model near the end of Gillespie’s section 4.3.²⁵

For example, consider this equation from section 4.2 (Gillespie, 2004, 107),

$$x'_1 = \frac{x_1 \bar{w}_1 - w_{14} r D}{\bar{w}} . \quad (3)$$

This uses notation that’s a little bit different from Vagne et al.’s. It gives the next-generation frequency x'_1 of organisms with allele A_1 at the A locus and allele B_1 at the B locus. This frequency x'_1 is a function of the frequency x_1 of $A_1 B_1$ organisms in the current generation, where w_{14} is a fixed fitness parameter, and \bar{w}_1 and \bar{w} are average fitnesses calculated from (i) fitness parameters and (ii) genotype frequencies in the current generation. D is again linkage disequilibrium, although its calculation is simpler than in the Vagne example:

$$D = x_1 - p_1 p_2$$

24. On page 109, Gillespie mentions that the initial frequency of a new allele would be $1/2N$, where N is population size (see also §4.5), but this value is not used until later, in the computer simulation.

25. I believe that Gillespie (2004) is more explicit about the absence of drift in section 4.3 because there he is discussing a process, “genetic draft” that has some properties in common with genetic drift although it doesn’t depend on on finite population size (cf. Skipper 2006). Early in section 4.3, Gillespie presents models in which the population size is finite, and drift does play a role, but Gillespie wants to show that once a new mutation is introduced into a population, draft operates even when drift doesn’t.

Here p_1 and p_2 are the overall frequencies of the A_1 and B_1 alleles, respectively (102).

Gillespie's mathematical model has the same character as those that Vagne et al. use for "populations of infinite size" (quoted above): frequencies in the next generation are calculated directly from frequencies in the previous generation along with constant parameters. So it seems as if Gillespie's model can be viewed as an infinite population model. Yet the computer simulation that generates a plot illustrating this model is said (110) to use a finite population size ($N = 5000$). So in what sense can this be a drift-free model? Gillespie gives sample computer source code at the end of the chapter as a solution to an exercise in which the reader is supposed to recreate the data that produced his figure.²⁶ In this code, a variable N is set to 5000. This variable appears exactly once more in the program, in the next line, where it's used to initialize a variable eps (epsilon) with the value $1/(2 \times N)$. The variable eps is in turn used for only one purpose, which is to determine when the frequency p_1 of the A_1 allele—the allele being selected for in the model—becomes close enough to 1.0 that the computer program should stop. The idea is that if the population size is 5000 and each organism has the A locus on two chromosomes, the number of A_1 alleles in the population would be 10,000. Thus if the frequency p_1 of A_1 in the population differs from 1.0 by less than $\text{eps} = 1/10000$, that means that the number of organisms that lack the A_1 allele is less than 1, so the population has gone to fixation: all organisms have A_1 on both chromosomes. In that case, the simulation can be stopped.²⁷

Apparently, this is the only role that finite population size plays in the model described in Gillespie's section 4.2—the model that is embodied in the simulation code in his appendix. In other respects, the model is similar in character to models that are characterized as involving an infinite population, including Gillespie's section 4.3 model. So we might view Gillespie's model as both assuming both infinite population size and finite population size. That sounds contradictory. However, if we understand "infinite population size" in models as Vagne et al. seem to do—as implying *merely* that the model calculates next-generation frequencies directly from current generation frequencies—then there is no contradiction.

4.5 Case 4: drift and no drift

Gillespie's model did not allow drift, though it depended in a narrow sense on assuming that the population was finite. Charlesworth and Charlesworth (2010) describe a more complex hitchhiking model that also includes what might, initially, be thought of as a theoretical conflict. The model first assumes that there is no drift, then adds an assumption that the population is finite, and subsequently considers a particular effect of drift in a finite population. The Charlesworths' presentation is based on Barton's (2000, §2(a)) version of a model introduced by Maynard Smith and Haigh (1974).²⁸ Most of the details of this Barton/Charlesworth model are a bit too complex to summarize here, but certain aspects of the model provide a good illustration of modeling practices that I want to highlight.

In modeling the effect of the introduction by mutation of a new, beneficial allele B_2 at one locus, Charlesworth and Charlesworth (2010, 410f, box 8.7) gradually develop and simplify an expression for the change Δq_A of the frequency q_A of an allele A_2 at a different locus. For more than half of the presentation, the Charlesworths' model involves none of the stochastic effects that would usually be present in a realistic population of finite size. So this again is an "infinite population" model in the same sense that Vagne et al.'s model was. However, about

26. Gillespie's sample answer in Python (a computer language) generates but doesn't plot the data.

27. In the mathematical model on which the simulation is based, the "frequency" of the A_1 allele could get closer and closer to 1 but never reach fixation as such.

28. See Charlesworth and Charlesworth (2010, 407, 409) and Barton (2000, 1554).

two-thirds of the way through the presentation (411, middle), the Charlesworths introduce a new element to the model. They note that since there is usually only one copy of a new allele when it is introduced by mutation,²⁹ the frequency of such an allele would be $1/2N$, where N is population size—assuming that each organism has two copies of each chromosome, and hence of each locus. The Charlesworths then plug this value $1/2N$ into the equation for Δq_A that they have derived without any explicit reference to population size, and without any drift-like effects represented. As the analysis continues, the authors note that most new mutations are lost by chance—essentially, because of drift—and they give an estimate for the probability that a new mutation persists in the population—i.e., that it is not lost. This estimate uses a function of population size and other factors (sN_e/N). The estimate is used to derive a new formula for the expected value of Δq_A . (Until the introduction of the chance that a new allele would be lost, bare quantities themselves were estimated, but once stochasticity was introduced into the model, the Charlesworths switch to a focus on expectations of certain values such as Δq_A .)

Thus in estimating the change in frequency of A_2 from one generation to the next, the Charlesworths take into account certain probabilistic consequences of finite population size in one part of the derivation, even though they ignore other probabilistic consequences of finite population size in earlier parts of the derivation. Like Gillespie's model, this model can be described as treating a population as infinite in some respects (in Vagne et al.'s sense), but finite in others. We can also describe it as involving both drift and its absence. I don't think there's anything surprising here for biologists. Neither Barton (2000) nor the Charlesworths (2010) remark on what might seem, superficially, like a contradiction in the model. And given the model's pedigree from Maynard-Smith and Haigh's (1974) paper, to Barton's (2000) revised model, subsequently enshrined in the Charlesworths' (2010) textbook, there is no reason to think that the *appearance* of contradiction is the result of a mistake, or that the model represents a problematic departure from mainstream evolutionary biology.

It's valuable to see that there are no *mathematical* contradictions in the model. It's true that there are assumptions associated with the mathematics that are inconsistent with each other: some parts of the mathematics are those associated with a drift-free evolutionary process, and others are associated with a process that incorporates drift. However, the mathematics itself is consistent. When the Charlesworths assume that a new mutation can be lost through drift in an otherwise "deterministic" model, they are simply adding a new premise—concerning the probability of loss of a mutation—to a mathematical derivation that had not previously incorporated terms for mutation. At each point in the Charlesworths' derivation, it is clear what it is about possible real populations that is being approximated.

This kind of modeling "inconsistency" seems common. A look at many extended analytical treatments of population genetical models shows that, even after a model's initial simplifying assumptions are stated, various other modifications are made in the course of the analysis.³⁰ That is, it's not simply that *different* models of the *same* system may incorporate incompatible assumptions but—as in the Barton/Charlesworth model—the *same* model may incorporate incompatible assumptions. For example, a modeler might include a squared fraction in one part of a derivation, but treat it as equal to zero at another point because it will be small relative to other terms.³¹ Or a modeler might replace an expression with its mathematical expectation,

29. For the same genetic mutation to arise in two members of a population in the same generation would be extremely improbable, so population genetic models often idealize away this possibility.

30. Zhao and Charlesworth (2016) provide nice illustrations of this point, although I didn't discuss those aspects of the paper above. There are also other illustrations in Charlesworth and Charlesworth (2010) and Gillespie (2004).

31. Gillespie's well-known (1977) paper illustrates this idea with a long equation (7) on page 1013 containing many terms that Gillespie ignores in order to justify the more widely discussed equations (1) $\mu - \frac{1}{2}\sigma^2$ and (2)

and use the resulting value as if it were the original expression. This kind of piecemeal, flexible, jury-rigged, empirically based mathematical inference is not unique to population genetics, but it seems common in both applied and theoretical population genetical modeling.

5 What are infinite population models (and kin)?

In this part of the paper, I discuss several ways of understanding models described as involving an infinite population, and ways to think about (superficially) inconsistent models such as the Gillespie and Barton/Charlesworth models described above.

5.1 Impossible worlds?

McLoone (2021) notes that biologists sometimes approximate the growth of a population (e.g., of rabbits) using a logistic equation model:

$$\frac{dN}{dt} = aN \left(1 - \frac{N}{K} \right). \quad (4)$$

Here N is population size, a is a growth rate, and K is the carrying capacity of the environment, i.e., the number of organisms that it can support (cf. Roughgarden 1979, 303, 306; Hartl and Clark 1989, 518). McLoone remarks that mathematically, this treats population size as a continuously varying quantity, with an uncountably infinite number of states between 0 and the population size N . McLoone specifies, as an assumption I view as separate from the continuity assumption implied by the mathematical formalism, that “the population of rabbits is infinitely large ($N \rightarrow \infty$)” (McLoone, 2021, 12157).³² In section 3.3, I explained that a model with uncountable states needn’t be treated as embodying the idea of an infinite population. So the kind of model that McLoone discusses doesn’t seem to require an infinite population in the sense that is my focus here. Nevertheless, McLoone proposes a novel strategy for thinking about models with infinite states, and while this isn’t the place for a full discussion of his paper, it’s worth considering whether his approach could be relevant to some of the models I discussed above.

McLoone notes that while the model he discusses represents population size as a continuous quantity, no rabbit population size varies continuously. He argues that in order to understand practices involving such models we should use a semantics for scientific statements that allows impossible worlds as well as possible worlds. McLoone’s paper goes on to describe such a semantics based on the common philosophical idea of defining possible worlds as sets of propositions.³³ Treating worlds as sets of propositions makes it possible to define “impossible” worlds, ones

$\mu - \frac{1}{N}\sigma^2$ that summarize contributions of variance σ^2 and expectation μ of numbers of offspring to fitness. (The full justification for (1) and (2) from (7) comes from three of Gillespie’s earlier articles.)

32. To put this statement in relation to distinctions important to the present paper, McLoone may mean that we are to take N to be finite but allow it to increase without bound. If he meant that $N = \infty$, then equation (4) would imply that the change over time of infinity (i.e., dN/dt) would be equal to negative infinity or to an undefined quantity, depending on the order in which we applied arithmetic operations on the right. It’s possible that McLoone intended that taking the limit as N goes to infinity is what justifies equations such as (4). I’m not sure, though. If N were treated as a discrete variable with integer values representing possible population sizes, taking its limit would give us only a countable infinity. But if N were allowed to vary continuously over non-negative real numbers, we’d already be dealing with continuous variation in population states for finite N ’s.

33. If you think of a world or universe as a configuration of interrelated things and properties, then each such world can be characterized by the set of all propositions true in it, and we can understand the semantics of statements involving possibility in terms of such sets.

that contain inconsistent propositions, by including in a world (i.e., a set of propositions) some propositions that are inconsistent with each other.

McLoone's argument for using impossible world semantics is based on some authors' arguments that models should be understood as specifying counterfactual worlds. McLoone suggests that a semantics involving impossible worlds would be needed to capture the meaning of counterfactual statements such as

If a population of rabbits satisfied the assumptions of the logistic equation, then its size would eventually equal the carrying capacity. (McLoone, 2021, 12161)

However, McLoone's argument doesn't seem to depend on models for continuous quantities per se. Since many models idealize in the sense of misrepresenting what is modeled (e.g., Weisberg 2013; Potochnik 2017), any statement that conjoins assumptions of a model with truths about what is modeled is likely to involve contradiction.

McLoone may be right about such model-world conjunctive statements, and he may be right that philosophical views that take models to specify counterfactual worlds mean that such statements are implied by scientific models like the logistic rabbit model. However, I don't believe we need to take such conjunctions as central to modeling in biology. One common view is that a model based on the logistic equation uses continuous quantities to *represent* discrete numbers of rabbits (e.g., Weisberg 2013; Morrison 2015; Gelfert 2016; Potochnik 2017), rather than *identifying* discrete and continuous numbers (cf. §3.3). We needn't assume that conjunctions of model assumptions and what's true in the world must play a role in modeling practices in biology. We only have to assume that the model is used to represent the world. So I don't see this aspect of McLoone's approach as needed for cases like those discussed in section 4.

However, my argument in section 3.1 was not that (countably) infinite populations give rise to a paradox in the model-world relation. The paradox was in the model itself. If semantics involving impossible worlds was potentially valuable in some contexts, one might wonder whether they could be used to make sense of biologists' talk of infinite populations. Let us consider the idea of a non-actual world in which there was evolution in a countably infinite population. Such a world would be one, I suppose, that contained both the proposition α that the number of organisms of type T was infinite, and a proposition β that the relative frequency of organisms of type T was equal to some finite number strictly between 0 and 1. These statements are inconsistent, as I argued, but that seems OK for a semantics that allows impossible worlds. The thing to notice, however, is that the infinite population claim would not do any work in a model corresponding to this picture. For modeling purposes, we would need propositions related to β that stipulated how relative frequencies of trait types changed over time (§3.3), but propositions such as α concerning the infinite size of the population would play no role in inferences about changes in frequencies. So an impossible worlds semantics doesn't add anything of immediate use for thinking about evolution in an infinite population.

On the other hand, perhaps it could be argued that a semantics involving impossible worlds would provide a useful way to think about the Gillespie and Barton/Charlesworth models of sections 4.4 and 4.5, since in these models, there is both a finite population, or drift, in one sense, and an infinite population, or no drift, in another. Note that this would be a claim only about the model itself, and not about populations represented by the model or the mathematics of the model. I don't think we need impossible world semantics to understand such cases, and I'll describe what I see as better alternatives. However, I won't rule out the possibility that some views about modeling might have a role for such a strategy.

5.2 Limits

The examples I discussed in section 4 concern models that, I argued, incorporate *particular* assumptions that have sometimes been described as consequences of assuming an infinite population. Specifically, the models assume that frequencies at time t directly determine subsequent changes (§§4.2, 4.4, 4.5), or they assume the absence of linkage disequilibrium (§4.3). Given usual ways of thinking about evolutionary processes, both of these assumptions imply the absence of stochastic effects of random drift—effects whose probability would go to zero as population size increased. So we might suggest that statements about infinitely large populations are “justified as shorthands for claims about limits as population size is increased without bound” (Abrams, 2006, 264f), or that as Sober suggested, “the meaning behind the idea” (1984, 44) of an infinite population should be given in terms of limits (see §2). Strevens (2019) argued that reasoning using such limit statements should be viewed as “a rational reconstruction of what is going through population geneticists’ minds when they advance the infinite population idealization”.³⁴ The general idea here is that probabilities of effects of drift lessen as population size is increased, and these effects have a probability of zero in the limit.³⁵ Mathematical models that ignore drift would be literally true of what happens *in the limit*. This is consistent with the argument in section 3.1 that evolution in an infinite population per se is nonsensical. As Norton (2014) notes (§3.2), something can be true in the limit as N increases toward infinity without it being true for N equal to infinity.

The following statement from a recent scientific paper can be seen to lend some support to this limits view of infinite population claims. (It’s not necessary to understand all of the details of the quotation.)

... with [population size] $N = 64$, *linkage disequilibrium is important* [has a significant effect] and does not decrease much with genome size. Even with an infinite genome size (i.e., free recombination between loci), the correlations are still 0.95 between r_G and r_M and 0.88 between F_G and F_M . *When the population size is increased to 512, these correlations reduce to 0.92 and 0.66. It is predictable that these correlations reduce to zero for an infinite genome size in an infinite population.* (Wang, 2016, 8, emphasis added)

Here F_G , F_M , r_G , and r_M are different ways of measuring genetic relationships that would be influenced by linkage disequilibrium. The quotation shows that Wang is envisioning a progression as population size N (and genome size) increases. The last sentence shows that the terminal step of this progression is an infinite population (of organisms with infinite genomes). However, the justification for this point comes solely from a claim about limits. Thus, one might reinterpret claims about evolution in infinite populations as being about limits as $N \rightarrow \infty$, but not about populations where $N = \infty$ (since that makes no sense).

However, if the limit interpretation were correct about all references to infinite populations in evolutionary biology, the Gillespie model and the Barton/Charlesworth model discussed above show that this claim might have to be understood in a nuanced way: It may be correct that in *one part of the mathematical development of a model*, the assumption that there is no drift is justified by allowing population size N to go to infinity. However, as we’ve seen, the model can assume a finite N at another step. So the limit justification could not apply to the entire model in such cases.

34. The notion of a rational reconstruction (Carnap, [1928] 1969) is, roughly, a proposed way of providing a more systematic understanding of what scientists could or should be saying, which preserves much of what they do say and do.

35. Strictly speaking, this doesn’t mean that effects of drift are impossible; see note 11.

5.3 Large numbers

Another problem with understanding infinite population talk solely in terms of limits is that though papers such as those of Vagne et al. (2015) and Zhao and Charlesworth (2016) are quite theoretical, they seem to be oriented toward possible empirical applications—where population sizes do not increase toward infinity. Zhao and Charlesworth’s interest in empirical applications is shown by their remark that their “results shed light on experiments on the loss of variability at marker loci in laboratory populations” (Zhao and Charlesworth, 2016, 3). Vagne et al.’s empirical orientation is apparent from the article’s title itself: “When is recombination favorable in a pre-breeding program with a selfing species?” Here the “term ‘pre-breeding’ refers to the transfer of genes from related wild ancestors or from ancient varieties to breeding material” (Vagne et al., 2015, 45). That is, the models in the paper are intended to illuminate what happens to genetic patterns when animal or plant breeders practice artificial selection on a population of animals, at least some of whom were formerly wild. Similar methods are sometimes used for experiments on evolution (e.g., Alexander et al. 2014). Gillespie’s and the Charlesworths’ textbooks also repeatedly highlight connections to empirical work.

Why would it matter for empirical studies what happens as N increases without bound? Why discuss “infinite population” models at all, if they concern what would be the case in the limit? One reason is that if we understand infinite population size in terms of limits, that implies that if a population is sufficiently large, its behavior will be close to the behavior of a population as size goes to the limit.

So perhaps all talk of infinite populations is *simply* a way of talking about large, finite populations. Infinite population models are often intended to give us insights about the approximate behavior of large populations. In fact, many textbooks and articles fail to talk about infinite populations where they might have been expected to do so. Earlier, in section 3, I quoted a passage from the introduction to a textbook by Nagylaki that seemed to allude to infinite populations. However, when Nagylaki gets down to work and discusses actual models, he avoids talk of infinite populations, writing, for example:

The total number of offspring,

$$N = \sum_i n_i \quad (2.1)$$

must be sufficiently large to allow us to neglect random drift. (Nagylaki, 1992, 5)

Thus one can ignore random drift, which involves changes whose probability goes to zero as N goes to infinity, as long as N is large enough that those probabilities are small. Another illustration can be found in an article by Waxman and Loewe:

Within the life cycle we assume that a very large number zygotes of all genotypes are produced, so that viability selection is essentially deterministic in character. (Waxman and Loewe, 2010, 246)

That is, a “large number” of organisms (in the form of fertilized eggs) are produced, so selection is “essentially deterministic”, i.e., it occurs without the interference of drift. In the Charlesworths’ textbook, there is an explicit statement of this equation of the “infinite” with the merely large:

... we assume that ... the population size is so large that random fluctuations in genotype frequencies can be disregarded (for convenience, this is called an “infinite” population). (Charlesworth and Charlesworth, 2010, 50)

In the previous section, we considered the possibility that “infinite population” meant “in the limit as N goes to infinity.” The examples in this section, by contrast, suggest that “infinite population” sometimes means “large population”—i.e., large enough that the influence of drift can be ignored.

5.4 *Beyond limits and large numbers*

It’s somewhat reasonable to think that “infinite population” always means “large population” or “in the limit as population size increases”. However, if the point of large populations or limiting size claims is that they are what justify leaving various effects of drift out of a model, how are we to understand the Gillespie and Barton/Charlesworth models? The first ignores effects of drift without assuming the population is large,³⁶ and the second would seem to need its assumptions justified both by assuming a population is large and by assuming it is not large. (A population’s size might be large enough to ignore for one purpose, but not for the other. However, it doesn’t seem as if these two models are supposed to be restricted to carefully tailored ranges of population sizes in this way.) Similar remarks can be formulated in terms of limits as $N \rightarrow \infty$.

I think the most reasonable view is that in such cases, the modeler(s) simply idealize away effects of finite population size in one part of the model, but not in another. Or more specifically, they assume the absence of drift in one part of the model, but allow drift or other effects of finite population size in another part of the model. Recall that in neither the Gillespie model nor the Barton/Charlesworth model was there any mathematical inconsistency. There is no need to think of population size as both large and small, or as taking N to the limit while keeping N bounded.

Now, neither Gillespie nor the Charlesworths used the term “infinite population” to describe the models I’m currently discussing. However, either there are drift-free models that involve infinite populations in some sense, and also drift-free models that don’t; or it’s incorrect to identify claims about the “infinite size” of a population with the absence of drift within a model. The former alternative seems incorrect, because there appears to be no meaningful difference between models in which drift is absent that are associated with phrases like “infinite population”, and those that aren’t. Describing a model as involving an infinite population can aid communication, but it’s inessential. My guess is that Gillespie and the Charlesworths avoided using “infinite population” for these two models simply because they recognized that using “infinite population” would have confused readers. They did use the phrase elsewhere, in places where it would not have been confusing (§§4.4, 5.3).

Thus it seems most plausible to view “infinite population” talk as an informal way to convey that some effects of drift that might have been thought relevant to a model are excluded from it. This means that an infinite population model does not incorporate an idealization in the form of an infinite population assumption. That is, the model does not make a false assumption that a represented population is infinite. On the other hand, such models do incorporate idealizations in the sense that they falsely represent real populations as if they were free from one or more effects of drift.³⁷ But it’s wrong to view “infinite population” talk as always meaning (Sober, 1984), as

36. Even if $N = 5000$ were large, this assumption plays no role in justifying leaving drift out of Gillespie’s model. Assuming that $N = 50$, for example, would simply end the computer simulation sooner.

37. Godfrey-Smith (2009a) and Potochnik (2017), following Jones (2016), distinguish between idealization, in which a model incorporates assumptions that are false of systems that are modeled, and “abstraction”, in which features of some real-world system are not misrepresented, but are simply not represented. So one might think that a model in which certain effects of drift are absent is one that incorporates an abstraction. This may seem plausible given views that treat drift as a variety of physical process that’s distinct from natural selection (e.g., Millstein 2002; Gildenhuys 2009). A model that ignores drift could be viewed as simply failing to represent drift processes in natural populations. Other authors treat natural selection and drift as intimately involved aspects of

justified by (Abrams, 2006), or as rationally reconstructible as (Strevens, 2019) statements about limits.³⁸ Further, contrary to what Potochnik wrote, it seems clear that an infinite population is not a “requisite assumption whenever genetic drift is not taken into account” (Potochnik, 2017, 55).

5.5 *Virtues of loose language*

The case studies and quotations above (§§4, 5.2–5.4) suggest that sometimes biologists think about infinite population assumptions as involving limits, while at other times they think of these assumptions as merely concerned with large populations. I suggest that at a minimum, “infinite population” means nothing more than “we’re leaving out such and such particular effects of drift”. Some readers may want an analysis of “infinite population” that gives it a more specific, determinate meaning. But in each of the examples I discussed in section 4, it was always clear from context what assumptions about drift or its absence were to be used. When authors used phrases like “infinite population”, it was clear what the implications of the phrasing was. Shouldn’t that be enough? I see no reason for assuming that biological language must express meanings with more precision than is needed for successful research. There is little reason to think that “infinite population” has a single, precisely specifiable meaning (even putting aside models that use uncountable numbers of population states, such as those I discussed in section 3.3).

Consider Keller’s response to the idea that scientists should adhere to the “harsh mandate” of always working to replace figurative language with precisely defined terms:

The difficulty is obvious: scientific research is typically directed at the elucidation of entities and processes about which no clear understanding exists, and to proceed, scientists must find ways of talking about what they do not know—about that which they as yet have only glimpses, speculations. To make sense of their day-to-day efforts, they need to invent words, expressions, forms of speech that can indicate or point to phenomena for which they have no literal descriptors. ... Making sense of what is not yet known is thus necessarily an ongoing and provisional activity, a groping in the dark; and for this, the imprecision and flexibility of figurative language is indispensable. (Keller, 2002, 118)

On my view, the term “infinite population” is optional figurative language. Rather than a being used to grope toward an understanding of new phenomena, as in Keller’s quotation, the flexible associations of “infinite population” may aid communication about similar but different modeling

the same process (e.g., Abrams 2007, 2023; Clatterbuck et al. 2013). On this view, leaving drift out of a model is an idealization, since it distorts the character of a single selection-and-drift process. Note, however, that even on a distinct-process view such as Millstein’s, drift is always present in real populations, and typically influences its evolution. That means that inferences drawn from a model that does not incorporate all relevant effects of drift would distort implications concerning probable changes in a real population. In that sense, even on the distinct-process view, a model that leaves out effects of drift involves an idealization.

38. Strevens (2019) gives a complex argument for a complex non-literal interpretation of “infinite population” talk, motivated by generalizations based on simple models that abstract from details like those I discuss here. Although Strevens’ conclusions don’t seem consistent with all of the cases I discuss above, his view might provide useful perspectives on some modeling contexts. A detailed discussion of his proposal belongs in another paper, though. (Strevens also argues that infinite alleles and infinite sites models should always be understood as rational reconstructions of statements about limits, as the number of alleles or loci increases without bound. As I remarked in note 12, those models don’t seem problematic in the way that infinite population models are. Strevens’ account of infinite alleles/sites models might be valuable even though it’s based on abstracting from details of practice. I think it’s more plausible that “infinite” in “infinite sites” and “infinite alleles” is merely a loose idiom, but I won’t try to argue for that view here.)

patterns in related contexts. I suggest that the variation in uses of “infinite population”, and the deployment of different assumptions about drift and population size that I’ve described are consequences of this flexibility.

This is not to say that all important terms and concepts in evolutionary biology have the same flexibility. Some terms may come to have single, precise meanings, or may have a series of more or less precisely defined variants. For example, Ereshefsky (1992) may be right that there are only a few alternative species concepts, and the few gene concepts that Griffith and Stotz (2014) described may capture many common uses of “gene”. On the other hand Waters argues that

Sometimes it is useful to be vague, and in such contexts biologists invoke a blunt concept akin to the gene concept of classical genetics In other contexts it is important to be precise. When precision is important, biologists employ what I have called the *molecular gene concept* (Waters, 2017, 94, emphasis in original)

Waters characterizes the molecular gene concept as a parameterized family of ways of narrowing its application:

The molecular gene concept can be specified as follows:

A gene g for linear sequence l in product p synthesized in cellular context c is a potentially replicating nucleotide sequence, n , usually contained in DNA, that determines the linear sequence l in product p at some stage of DNA expression.

The reference of any gene, g is a specific sequence of nucleotides. The exact sequence to which a g refers depends on how the placeholders l , p , and c are filled out. (Waters, 2017, 95)

So according to Waters, while some gene concepts are vague, the molecular gene concept is not. It is a parameterized family of more precise applications of the concept (see also Waters 2019).

Somewhat like the molecular gene concept, “infinite population” language is associated with a family of closely related ways of thinking about and applying ideas about population size and drift. “Infinite population” language per se is so vague, however, that it is *just* these associations that constitute what the term amounts to in practice. Although the associated ideas about population size (size in the limit, large size, etc.) and about effects of drift (frequency change, loss of mutations, linkage equilibrium, etc.) are logically and mathematically related, they can’t be organized in terms of a few parameters as in the case of the molecular gene concept. The applications of ideas concerning drift in modeling remain very flexible.

6 Conclusion

Some mentions of (countably) infinite population models by philosophers vaguely suggest, without necessarily requiring, that the idea of an infinite population per se plays a role in models (e.g., Sober 2008; Weisberg 2013; Morrison 2015; Potochnik 2017). It is clear that infinite populations per se can’t play that role. Where philosophers have clearly expressed a less problematic view (Sober, 1984; Abrams, 2006; Strevens, 2019), they’ve usually argued for the idea that infinite population claims should be understood in terms of limits. The arguments above imply that at the very least, the view that infinite population talk should be understood in terms of limits gives a misleading perspective on scientific practices. It’s valuable to look closely

at a wider variety of cases than the sorts of models that seem to have motivated claims like those made by Sober (1984), Abrams (2006), and Strevens (2019). More generally, I've argued that it's a mistake to view "infinite population" and the like as having a precise meaning, justification, or rational reconstruction that fits all of its uses. The variety of models that are associated with infinite populations seems best understood by viewing "infinite population" as figurative language, in a scientific tradition of practices that involve closely related ways of representing the absence of effects of drift.

I suspect that some readers will find this view unsatisfying. Am I just giving up on the philosophical ideal of providing a systematic account of an area of science, or on the scientific ideal of describing the world with precision? I share the desire for a relatively simple, systematic account of scientific terminology such as "infinite population", and of practices involving modeling. The facts, it seems, don't support such an account in this case: the variation in useful practices is too great. Does that mean that biologists' practices involving models and drift are ad hoc and irrational? No, of course not. There is systematicity to models involving drift or its absence, but it's just the systematicity that results from biologists' reasoning, and from social practices concerning models and the natural world. Varied and often novel, appropriate practices can arise from reasoning, social interactions, modifications of the work of others, and so on, without there being a *simple* systematic account of the resulting practices. It might be possible to give a more specific, systematic account of uses of "infinite population" and modeling practices concerning drift, or a set of heuristic descriptions of these practices. However, an account that reflected examples like those I've discussed would not be simple, I believe. Its development might require social scientific research on contemporary practices across different labs and research contexts.

I end with remarks about possible questions for future research. First, to what extent are such extremely flexible uses of language and models common in science, reducing the value of attempts to systematize scientific language? Elsewhere (Abrams, 2023), I argue that "population" terminology used in empirical research is flexible in ways that might not be specifiable outside of particular research contexts,³⁹ though the flexibility characteristic of population talk is quite different in character from that of infinite population talk. Some treatments of modeling in the physical sciences seem to consider models as if they could not include the kind of piecemeal, superficial inconsistency of the Gillespie and Barton/Charlesworth models (e.g., Suárez 2009b; Callender 2001). That makes some sense, because physical scientists study a physical world that seems to be subject to systematic laws, dispositions, etc. By contrast, biological sciences study systems that exhibit subtle or large variations, with messiness involving complex differences in many respects (Wimsatt, 2007; Mitchell, (2009) 2012; Waters, 2017, 2019; Abrams, 2023). It would perhaps be natural, then, if modeling in biological sciences turned out to require more ad hoc flexibility. However, one can find illustrations of practices in physical sciences that seem to exhibit similar sorts of flexibility (McComb 2004; Wimsatt 2007, ch. 13; Wilson 2006, ch. 4). Moreover, some authors such as Bill Wimsatt and Jim Griesemer (e.g., Wimsatt and Griesemer 2007; Wimsatt 2007) have provided analyses of scientific practices in some contexts that capture patterns of reasoning and norms in science without providing a fully systematic account, which might be impossible for the cases they discuss. This kind of strategy may be an option for drift and "infinite population".

Second, the fact that some evolutionary models explicitly assume that a population is finite while also representing it as if it were free of drift, or treat a population as subject to drift in some respects but not in others, may present special challenges for some philosophical theories of

39. Others such as Millstein (2010) argue that evolutionary biology and ecology do or should depend on a few well-defined population concepts.

scientific models. Advocates of the view that models are fictions⁴⁰ may need to address the fact that models such as the Gillespie model and the Barton/Charlesworth model seem to require inconsistent fictional worlds—worlds perhaps more like dreams than most novels, or perhaps like the impossible worlds of McLoone’s formalization. Formal characterizations of similarity between models and targets (e.g., Weisberg 2013) might also need to be refined to deal with cases like these. If it turned out to be difficult for some theories of modeling to account for cases like the Gillespie and Barton/Charlesworth models, because those philosophical theories attribute structure to models that fits poorly with such cases, perhaps that’s a reason to favor the kind of the deflationary model semantics advocated by Callender and Cohen (2006) and Odenbaugh (2021). Their approaches allow mathematical expressions to represent in a manner that is primarily dependent on the semantic capabilities of scientists, and scientists’ intentions about what is to be represented by various parts of a model. Biologists already seem to have consistent ways of thinking about what is represented in models like the Gillespie and Barton/Charlesworth models, so theories of representation that make it parasitic on scientists’ existing thought processes might need no refinement for such cases.

Acknowledgments

I’m immensely grateful for a diversity of feedback from anonymous reviewers. This paper evolved from a presentation at the Munich Center for Mathematical Philosophy, where Samuel Fletcher, Patricia Palacios, Paul Teller, and Michael Weisberg were among those who provided very helpful comments.⁴¹ Conversations with Yoichi Ishida, Samuel Scheiner, Glenn Shafer, and Elay Shech were also very helpful.

Literature cited

- Abrams, Marshall. 2006. “Infinite Populations and Counterfactual Frequencies in Evolutionary Theory.” *Studies in History and Philosophy of Biological and Biomedical Sciences* 37 (2): 256–68.
- Abrams, Marshall. 2007. “How Do Natural Selection and Random Drift Interact?” *Philosophy of Science* 74 (5): 666–79.
- Abrams, Marshall. 2023. *Evolution and the Machinery of Chance*. Chicago: University of Chicago Press.
- Alexander, H. J., J. M. L. Richardson, and B. R. Anholt. 2014. “Multigenerational Response to Artificial Selection for Biased Clutch Sex Ratios in *Tigriopus californicus* Populations.” *Journal of Evolutionary Biology* 27 (9): 1921–29.
- Appiah, Kwame Anthony. 2017. *As If: Idealization and Ideals*. Cambridge, MA: Harvard University Press.
- Barberousse, Anouk, and Pascal Ludwig. 2009. “Models as Fictions.” In *Fictions in Science: Philosophical Essays on Modeling and Idealization*, edited by Mauricio Suárez, 56–73. Routledge.
- Barton, N. H. 2000. “Genetic Hitchhiking.” *Philosophical Transactions: Biological Sciences* 355 (1403): 1553–62.
- Bokulich, Alisa. 2009. “Explanatory Fictions.” In *Fictions in Science: Philosophical Essays on Modeling and Idealization*, edited by Mauricio Suárez, 91–109. Routledge.

40. See, e.g., (Fine, 1993; Godfrey-Smith, 2009c; Frigg and Nguyen, 2016; Barberousse and Ludwig, 2009; Bokulich, 2009; Suárez, 2009b).

41. I am always grateful for the support of family members for my travel and research. In the case of the Munich workshop, the support of my mother, Lois Barnett, to whom this paper is dedicated, and the support of my sister Julie Abrams were particularly meaningful and important.

- Bourrat, Pierrick. 2018. "Natural Selection and Drift as Individual-Level Causes of Evolution." *Acta Biotheoretica* 66 (3): 159–76.
- Callender, Craig. 2001. "Taking Thermodynamics Too Seriously." *Studies in History and Philosophy of Modern Physics* 32 (4): 539–53.
- Callender, Craig, and Jonathan Cohen. 2006. "There Is No Special Problem About Scientific Representation." *Theoria* 55: 67–85.
- Carnap, Rudolf. [1928] 1969. *The Logical Structure of the World*. University of California Press.
- Charlesworth, Brian, and Deborah Charlesworth. 2010. *Elements of Evolutionary Genetics*. Roberts and Company.
- Clatterbuck, Hayley. 2015. "Drift Beyond Wright–Fisher." *Synthese* 192 (11): 3487–507.
- Clatterbuck, Hayley, Elliott Sober, and Richard C. Lewontin. 2013. "Selection Never Dominates Drift (nor vice versa)." *Biology & Philosophy* 28 (4): 577–92.
- Ereshefsky, Marc. 1992. "Eliminative Pluralism." *Philosophy of Science* 59 (4): 671–90.
- Ewens, Warren J. 2004. *Mathematical Population Genetics, I. Theoretical Introduction*. 2nd ed. Springer.
- Fine, Arthur. 1993. "Fictionalism." In *Midwest Studies in Philosophy*, vol. XVIII, edited by Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, 1–18. Minneapolis: University of Minnesota Press.
- Fisher, Ronald A. 1922. "On the Dominance Ratio." *Proceedings of the Royal Society of Edinburgh* 42: 321–41.
- Fisher, Ronald A. [1930] 2000. *The Genetical Theory of Natural Selection*. Oxford University Press.
- Fletcher, Samuel C., Patricia Palacios, Laura Ruetsche, and Elay Shech. 2019. "Infinite Idealizations in Science: An Introduction." *Synthese* 196 (5): 1657–69.
- Frigg, Roman, and James Nguyen. 2016. "The Fiction View of Models Reloaded." *Monist* 99 (3): 225–42.
- Gazal, Steven, Hilary K. Finucane, Nicholas A. Furlotte, Po-Ru Loh, Pier Francesco Palamara, Xuanyao Liu, Armin Schoech, Brendan Bulik-Sullivan, Benjamin M. Neale, Alexander Gusev, and Alkes L. Price. 2017. "Linkage Disequilibrium-Dependent Architecture of Human Complex Traits Shows Action of Negative Selection." *Nature Genetics* 49: 1421–27.
- Gelfert, Axel. 2016. *How to Do Science with Models: A Philosophical Primer*. Springer.
- Giere, Ronald N. 2009. "Why Scientific Models Should Not Be Regarded as Fictions." In *Fictions in Science: Philosophical Essays on Modeling and Idealization*, edited by Mauricio Suárez, 248–58. Routledge.
- Gildenhuys, Peter. 2009. "An Explication of the Causal Dimension of Drift." *British Journal for the Philosophy of Science* 60: 521–55.
- Gillespie, John H. 1974. "Natural Selection for Within-Generation Variance in Offspring Number." *Genetics* 76: 601–6.
- Gillespie, John H. 1977. "Natural Selection for Variances in Offspring Numbers: A New Evolutionary Principle." *American Naturalist* 111: 1010–14.
- Gillespie, John H. 2004. *Population Genetics: A Concise Guide*. 2nd ed. Baltimore: Johns Hopkins University Press.
- Godfrey-Smith, Peter. 2009a. "Abstractions, Idealizations, and Evolutionary Biology." In *Mapping the Future of Biology: Evolving Concepts and Theories*, edited by Anouk Barberousse, Michel Morange, and Thomas Pradeu, 47–56. Springer.


- Godfrey-Smith, Peter. 2009b. *Darwinian Populations and Natural Selection*. Oxford: Oxford University Press.
- Godfrey-Smith, Peter. 2009c. "Models and Fictions in Science." *Philosophical Studies* 143 (1): 101–16.
- Goszczynski, D. E., C. M. Corbi-Botto, H. M. Durand, A. Rogberg-Muñoz, S. Munilla, P. Peral-Garcia, R. J. C. Cantet, and G. Giovambattista. 2018. "Evidence of Positive Selection Towards Zebuine Haplotypes in the BoLA Region of Brangus Cattle." *Animal* 12 (2): 215–223.
- Griffiths, Paul E., and Karola Stotz. 2014. "Conceptual Barriers to Interdisciplinary Communication." In *Enhancing Communication & Collaboration in Interdisciplinary Research*, edited by Michael O'Rourke, Stephen Crowley, Sanford D. Eigenbrode, and J. D. Wulfhorst, 195–215. Sage Publications.
- Grimmett, G. R., and D. R. Stirzaker. 2020. *Probability and Random Processes*. 4th ed. Oxford University Press.
- Hartl, Daniel L., and Andrew G. Clark. 1989. *Principles of Population Genetics*. 2nd ed. Sinauer.
- Hodge, M. J. S. 1992. "Biology and Philosophy (Including Ideology): A Study of Fisher and Wright." In *The Founders of Evolutionary Genetics*, edited by Sahotra Sarkar, 231–93. Springer.
- Huneman, Philippe. 2012. "Natural Selection: A Case for the Counterfactual Approach." *Erkenntnis* 76 (2): 171–94.
- Ishida, Yoichi. 2017. "Sewall Wright, Shifting Balance Theory, and the Hardening of the Modern Synthesis." *Studies in History and Philosophy of Biological and Biomedical Sciences* 61: 1–10.
- Jones, Martin R. 2016. "Idealization and Abstraction: A Framework." In *Idealization XII: Correcting the Model: Idealization and Abstraction in the Sciences*, edited by Martin R. Jones and Nancy Cartwright, 173–217. Brill.
- Keller, Evelyn Fox. 2002. *Making Sense of Life: Explaining Biological Development with Models, Metaphors, and Machines*. Cambridge, MA: Harvard University Press.
- Kimura, Motoo. 1969. "The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population Due to Steady Flux of Mutations." *Genetics* 61 (4): 893–903.
- Kimura, Motoo, and James F. Crow. 1964. "The Number of Alleles That Can Be Maintained in a Finite Population." *Genetics* 49 (4): 725–38.
- Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. 1st ed. Chicago: University of Chicago Press.
- Lynch, Michael, and Bruce Walsh. 2018. *Evolution and Selection of Quantitative Traits*. Oxford: Oxford University Press.
- Matthen, Mohan. 2010. "What Is Drift? A Response to Millstein, Skipper, and Dietrich." *Philosophy & Theory in Biology* 2: 2.
- Matthewson, John, and Brett Calcott. 2011. "Mechanistic Models of Population-Level Phenomena." *Biology & Philosophy* 26 (5): 737–56.
- Maynard Smith, John, and John Haigh. 1974. "The Hitch-Hiking Effect of a Favourable Gene." *Genetical Research* 23 (1): 23–35.
- McComb, W. D. 2004. *Renormalization Methods: A Guide for Beginners*. Oxford: Oxford University Press.
- McLoone, Brian. 2021. "Calculus and Counterpossibles in Science." *Synthese* 198 (12): 12153–74.
- McShea, Daniel W., and Robert N. Brandon. 2010. *Biology's First Law: The Tendency for Diversity and Complexity to Increase in Evolutionary Systems*. Chicago: University of Chicago Press.
- McShea, Daniel W., Steve C. Wang, and Robert N. Brandon. 2019. "A Quantitative Formulation of

- Biology's First Law." *Evolution* 73 (6): 1101–15.
- Millstein, Roberta L. 2002. "Are Random Drift and Natural Selection Conceptually Distinct?" *Biology & Philosophy* 17: 35–53.
- Millstein, Roberta L. 2010. "The Concepts of Population and Metapopulation in Evolutionary Biology and Ecology." In *Evolution Since Darwin: The First 150 Years*, edited by M. A. Bell; D. J. Futuyma; W. F. Eanes; and J. S. Levinton, 61–86. Sinauer.
- Millstein, Roberta L., Robert A. Skipper, and Michael A. Dietrich. 2009. "(Mis)interpreting Mathematical Models: Drift as a Physical Process." *Philosophy & Theory in Biology* 1: 2.
- Mitchell, Sandra D. [2009] 2012. *Unsimple Truths: Science, Complexity, and Policy*. Chicago: University of Chicago Press.
- Morrison, Margaret. 2000. *Unifying Scientific Theories: Physical Concepts and Mathematical Structures*. Cambridge: Cambridge University Press.
- Morrison, Margaret. 2009. "Fictions, Representation, and Reality." In *Fictions in Science: Philosophical Essays on Modeling and Idealization*, edited by Mauricio Suárez, 110–35. Routledge.
- Morrison, Margaret. 2015. *Reconstructing Reality: Models, Mathematics, and Simulations*. Oxford: Oxford University Press.
- Nagylaki, Thomas. 1992. *Introduction to Theoretical Population Genetics*. Springer-Verlag.
- Norton, John D. 2012. "Approximation and Idealization: Why the Difference Matters." *Philosophy of Science* 79: 207–32.
- Norton, John D. 2014. "Infinite Idealizations." In *European Philosophy of Science—Philosophy of Science in Europe and the Viennese Heritage: Vienna Circle Institute Yearbook*, vol. 17, 197–210. Springer.
- Odenbaugh, Jay. 2021. "Models, Models, Models: A Deflationary View." *Synthese* 198: 1–16.
- Otsuka, Jun. 2016. "A Critical Review of the Statisticalist Debate." *Biology & Philosophy* 31 (4): 459–82.
- Pence, Charles H. 2017. "Is Genetic Drift a Force?" *Synthese* 194 (6): 1967–88.
- Pence, Charles H. 2021. *The Causal Structure of Natural Selection*. Cambridge: Cambridge University Press.
- Plutynski, Anya. 2004. "Explanation in Classical Population Genetics." *Philosophy of Science* 71 (5): 1201–15.
- Plutynski, Anya. 2005a. "Explanatory Unification and the Early Synthesis." *British Journal for the Philosophy of Science* 56(3):595–609.
- Plutynski, Anya. 2005b. "Parsimony and the Fisher-Wright Debate." *Biology & Philosophy* 20 (4): 697–713.
- Plutynski, Anya. 2007. "Drift: A Historical and Conceptual Overview." *Biological Theory* 2 (2): 156–67.
- Plutynski, Anya. 2008. "Explaining How and Explaining Why: Developmental and Evolutionary Explanations of Dominance." *Biology & Philosophy* 23 (3): 363–81.
- Potochnik, Angela. 2017. *Idealization and the Aims of Science*. Chicago: University of Chicago Press.
- Provine, William B. 1986. *Sewall Wright and Evolutionary Biology*. Chicago: University of Chicago Press.
- Ramsey, Grant. 2013. "Driftability." *Synthese* 190 (17): 3909–28.
- Roughgarden, J. 1979. *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. Macmillan.
- Sabeti, Pardis C., David E. Reich, John M. Higgins, Haninah Z. P. Levine, Daniel J. Richter, Stephen F. Schaffner, Stacey B. Gabriel, Jill V. Platko, Nick J. Patterson, Gavin J. McDonald, Hans C. Ackerman, Sarah J. Campbell, David Altshuler, Richard Cooperk, Dominic Kwiatkowski, Ryk Ward, and

- Eric S. Lander. 2002. "Detecting Recent Positive Selection in the Human Genome from Haplotype Structure." *Nature* 419: 832–37.
- Shapiro, Larry, and Elliott Sober. 2007. "Epiphenomenalism—the Do's and the Don'ts." In *Thinking about Causes: From Greek Philosophy to Modern Physics*, edited by Gereon Wolters and Peter Machamer, 235–64. Pittsburgh: University of Pittsburgh Press.
- Skipper, Robert A. 2002. "The Persistence of the R.A. Fisher–Sewall Wright Controversy." *Biology & Philosophy* 17 (3): 341–67.
- Skipper, Robert A. 2006. "Stochastic Evolutionary Dynamics: Drift versus Draft." *Philosophy of Science* 73 (5): 655–65.
- Sober, Elliott. 1984. *The Nature of Selection*. Cambridge, MA: MIT Press.
- Sober, Elliott. 2008. *Evidence and Evolution: The Logic Behind the Science*. Cambridge: Cambridge University Press.
- Strevens, Michael. 2019. "The Structure of Asymptotic Idealization." *Synthese* 196 (5): 1713–31.
- Suárez, Mauricio, ed. 2009a. *Fictions in Science: Philosophical Essays on Modeling and Idealization*. Routledge.
- Suárez, Mauricio. 2009b. "Scientific Fictions as Rules of Inference." In *Fictions in Science: Philosophical Essays on Modeling and Idealization*, edited by Mauricio Suárez, 158–78. Routledge.
- Tajima, Fumio. 1996. "Infinite-Allele Model and Infinite-Site Model in Population Genetics." *Journal of Genetics* 75 (1): 27–31.
- Turner, John R. G. 1987. "Random Genetic Drift, R. A. Fisher, and the Oxford School of Ecological Genetics." In *The Probabilistic Revolution*, vol. 2, edited by Lorenz Krüger, Gerd Gigerenzer, and Mary S. Morgan, 313–54. Cambridge, MA: MIT Press.
- Vagne, Constance, Jacques David, Muriel Tavaud, and Bénédicte Fontez. 2015. "Reciprocal Sign Epistasis and Truncation Selection: When Is Recombination Favorable in a Pre-Breeding Program With a Selfing Species?" *Journal of Theoretical Biology* 386: 44–54.
- Voight, Benjamin F., Sridhar Kudaravalli, Xiaoquan Wen, and Jonathan K. Pritchard. 2006. "A Map of Recent Positive Selection in the Human Genome." *PLOS Biology* 4 (3): 446–58.
- Walsh, Denis M., André Ariew, and Mohan Matthen. 2017. "Four Pillars of Statisticalism." *Philosophy, Theory, and Practice in Biology* 9: 1.
- Wang, Jinliang. 2016. "Pedigrees or Markers: Which Are Better in Estimating Relatedness and Inbreeding Coefficient?" *Theoretical Population Biology* 107: 4–13.
- Waters, C. Kenneth. 2017. "No General Structure." In *Metaphysics and the Philosophy of Science: New Essays*, edited by Matthew H. Slater and Zanja Yudell, 81–107. Oxford: Oxford University Press.
- Waters, C. Kenneth. 2019. "Presidential Address, PSA 2016: An Epistemology of Scientific Practice." *Philosophy of Science* 86 (4): 585–611.
- Waxman, David, and Laurence Loewe. 2010. "A Stochastic Model for a Single Click of Muller's Ratchet." *Journal of Theoretical Biology* 264: 1120–32.
- Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.
- Williams, George C. 1966. *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.
- Wilson, Mark. 2006. *Wandering Significance: An Essay on Conceptual Behavior*. Oxford: Oxford University Press.
- Wimsatt, William C. 1980. "Reductionistic Research Strategies and Their Biases in the Units of

- Selection Controversy.” In *Scientific Discovery*, vol. 2, edited by Thomas Nickles, 213–59. Reidel.
- Wimsatt, William C. 1981. “The Units of Selection and the Structure of the Multi-Level Genome.” In *PSA 1980*, vol. 2, 122–83. Chicago: University of Chicago Press.
- Wimsatt, William C. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.
- Wimsatt, William C., and James R. Griesemer. 2007. “Reproducing Entrenchments to Scaffold Culture: The Central Role of Development in Cultural Evolution.” In *Integrating Evolution and Development*, edited by Roger Sansom and Robert N. Brandon, 227–323. Cambridge, MA: MIT Press.
- Winther, Rasmus Grønfeldt, Michael J. Wade, and Christopher C. Dimond. 2013. “Pluralism in Evolutionary Controversies: Styles and Averaging Strategies in Hierarchical Selection Theories.” *Biology & Philosophy* 28 (6): 957–79.
- Winther, Rasmus Grønfeldt, Ryan Giordano, Michael D. Edge, and Rasmus Nielsen. 2015. “The Mind, the Lab, and the Field: Three Kinds of Populations in Scientific Practice.” *Studies in History and Philosophy of Biological and Biomedical Sciences* 52: 12–21.
- Zhao, Lei, and Brian Charlesworth. 2016. “Resolving the Conflict Between Associative Overdominance and Background Selection.” *Genetics* 203 (3): 1315–34.

© 2024 Author(s)

This is an open-access article, licensed under
Creative Commons Attribution 4.0 International 

This license requires that reusers give credit to the creator(s). It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format.

ISSN 2475-3025